HUMAN RETROVIRUSES

and AIDS

1997

*A COMPILATION AND ANALYSIS OF
NUCLEIC ACID AND AMINO ACID SEQUENCES*

Editors

Bette Korber
Theoretical Division
Los Alamos National Laboratory

Beatrice Hahn
Department of Medicine & Microbiology
University of Alabama

Brian Foley
Theoretical Division
Los Alamos National Laboratory

John W. Mellors
Graduate School of Public Health
University of Pittsburgh

Thomas Leitner
Theoretical Division
Los Alamos National Laboratory

Gerald Myers
Theoretical Division
Los Alamos National Laboratory

Francine McCutchan
Henry M. Jackson foundation

Carla Kuiken
Theoretical Division
Los Alamos National Laboratory

Project Officer

James Bradac
Division of AIDS
National Institute of Allergy and Infectious Diseases

Database and Analysis Staff

Kersti MacInnes, Rama Thakallapally, Charles Calef,
Priya Dighe, Satish Pillai, Vijaya Doddi
Los Alamos National Laboratory

This publication is being funded by the Division of AIDS, National Institute of Allergy and Infectious Diseases, through an interagency agreement with the U.S. Department of Energy.

*Published by Theoretical Biology and Biophysics
Group T-10, Mail Stop K710
Los Alamos National Laboratory, Los Alamos, New Mexico 87545 U.S.A.*

MASTER

LA-UR 98-1268

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.

CONTENTS

More detailed tables of contents are to be found within the various parts of the compendium. The following provides merely an overview.

ACKNOWLEDGMENTS	ii
INTRODUCTION	iii
GLOSSARY and LANDMARKS	v
PART I. NUCLEIC ACID ALIGNMENTS	
Contents	I-1
Introduction	I-2
A. Nucleotide Alignment of HIV-1 Complete Genomes	I-18
B. Nucleotide Alignment of HIV-2/SIV Complete Genomes	I-92
PART II. AMINO ACID ALIGNMENTS	
Introduction	II-1
Contents	II-4
A. HIV-1 Amino Acid Alignments	II-A-1
B. HIV-2/SIV Amino Acid Alignments	II-B-1
C. SIV AGM and SIV SYK Amino Acid Alignments	II-C-1
PART III. ANALYSES	
Contents	
HIV-1 Coreceptor Use: A Molecular Window into Viral Tropism	III-1
The History of HIV-1 Biological Phenotypes Past, Present and Future	III-13
Updated Proposal of Reference Sequences of HIV-1 Genetic Subtypes	III-19
Intersubtype Recombinant HIV-1 Sequences	III-25
Fine Structure of HIV and SIV	III-31
Reagents for HIV/SIV Vaccine Studies	III-45
gp41, A Multifunctional Protein Involved in HIV Entry and Pathogenesis	III-55
Global Variation in the HIV-1 V3 Region*	III-74
Mutations in Retroviral Genes Associated with Drug Resistance	III-206
Drug Resistance Mutations Superimposed on the Structures of HIV-1 Protease and Reverse Transcriptase	III-250
PART IV. RELATED SEQUENCES	
Table of HIV-related Cellular Proteins	IV-1

*See pages 91-95 for a listing of two-letter country codes.

Acknowledgments

ACKNOWLEDGMENTS

The HIV Sequence Database and Analysis Project is funded by the Vaccine and Prevention Research Program of the AIDS Division of the National Institute of Allergy and Infectious Diseases (Dr. James Bradac, Project Officer) through an interagency agreement with the U.S. Department of Energy.

We thank the many researchers who have made their sequences available prior to publication.

We also thank editors from previous editions, Simon Wain-Hobson, Kuan-Teh Jeang, Lou Henderson and George Pavlakis, who were not a part of this year's publication but who may, at one time or another, return in this role.

Introduction

This compendium is the result of an effort to compile, organize, and rapidly publish as much relevant molecular data concerning the human immunodeficiency viruses (HIV) and related retroviruses as possible. The scope of the compendium and database is best summarized by the four parts that it comprises: (I) Nucleic Acid Alignments, (II) Amino Acid Alignments, (III) Reviews and Analyses, and (IV) Related Sequences. Information within all the parts is updated throughout the year on the Web site, <http://hiv-web.lanl.gov>. This year we are not including floppy diskettes as the entire compendium is available both at our Web site and at our ftp site. If you need floppy diskettes please contact either Bette Korber (btb@t10.lanl.gov) or Kersti Rock (karm@t10.lanl.gov) by email or fax ((505) 665-4453).

While this publication could take the form of a review or sequence monograph, it is not so conceived. Instead, the literature from which the database is derived has simply been summarized and some elementary computational analyses have been performed upon the data. Interpretation and commentary have been avoided insofar as possible so that the reader can form his or her own judgments concerning the complex information. The exception to this are reviews submitted by experts in areas deemed of particular and basic importance to research involving AIDS viral sequence information. These are included in Part III, and are contributed by scientists with particular expertise in the area of interest. In addition to the general descriptions below of the parts of the compendium, the user should read the individual introductions for each part.

Part I. Nucleic Acid Alignments. Annotated nucleic acid sequence alignments of complete genomes of HIV-1/CPZ (Part A) and HIV-2/SIV (Part B) are presented. The hard-copy annotation includes coding regions, regulatory structures, splice sites, and other features of functional significance. The authority for this annotation is often invariance, the recurrence of patterns such as TATAA and AATAAA, although a brief listing of some of the references used for the annotation is included. Although our practice has been to conservatively annotate, we caution the user against docility: sequence information regarding transcripts, for example, is neither certain or complete at this time. Brief descriptions of the sequences, and the isolates from which they were derived, are provided, as well as additional references. Beginning in 1997, only full-length sequences are presented in the hard copy edition, to save space. Other HIV and SIV sequences are catalogued, and long fragments of coding regions for particular genes are aligned; these coding region alignments are available at our Web or ftp sites. The full formatted GenBank entries of these sequences are located on the Web site (<http://hiv-web.lanl.gov>) and the database FTP server.

Part II. Amino Acid Alignments. This section contains alignments of the amino acid sequences (mostly full-length) of all known coding regions, and open reading frames of HIV-1/CPZ (Part A), HIV-2/SIV (Part B), and SIV AGM/SYK (Part C). Sequences representing the range of global variation, including commonly used reference strains, were selected for inclusion in the printed copy. Other alignments with more complete sets of sequences are available on the Web site. Protein processing sites are annotated when known as well as key functional domains and overlapping reading frames. The reader should consult the introduction to Part II for further explanation of the presentation and annotation of the amino acid sequences.

Part III. Analyses. This section is open-ended with the constraint that the sequence analyses and compilations be basic and of interest to the diverse community of database users. In 1997, analyses and curatorial contributions include: updating of tables of mutations relating to drug resistance, a description of full-length recombinant genomes, a section on gp41, a reference set of sequences for use in HIV-1 subtyping, background information on the second receptor usage of different isolates, a description of viral phenotyping nomenclature, etc.

Part IV. Related Sequences. In the past this section contained related sequences, but this year the emphasis has shifted. This section now contains a summary of cellular proteins known or hypothesized to interact with or be affected by HIV and related viruses.

Introduction

A comprehensive compilation of the nucleic acid and protein sequences published in the Human Retroviruses and AIDS Database since 1987 is available through our Web site, <http://hiv-web.lanl.gov> and on our FTP Server.

We are prepared to quickly enter both protein and nucleotide sequences into the Human Retroviruses and AIDS database, and in the case of nucleotide sequences oversee their entry into the large gene libraries. Submission of unpublished sequences is invited and encouraged. Sequence data or inquiries regarding the database should be addressed to

Bette Korber Theoretical Division T-
10, MS K710
LANL Los Alamos, NM 87545

(505)-665-4453; fax (505)-665-3493
e-mail:
btik@t10.lanl.gov

A short glossary follows.

HIV-1

HIV-2

HIV/SIV PROTEINS			
NAME	SIZE	FUNCTION	LOCALIZATION
Gag MA	p17	membrane anchoring; env interaction; nuclear transport of viral core. (myristylated protein)	virion
CA	p24	core capsid	virion
NC	p7	nucleocapsid, binds RNA	virion
	p6	binds Vpr	virion
Protease (PR)	p15	gag/pol cleavage and maturation	virion
Reverse transcriptase (RT), RNase H	p66 p51 (heterodimer)	reverse transcription, RNase H activity	virion
Integrase (IN)		DNA provirus integration	virion
Env	gp120/gp41	external viral glycoproteins bind to CD4 receptor	plasma membrane, virion envelope
Tat	p16/p14	viral transcriptional transactivator	primarily in nucleolus/nucleus
Rev	p19	RNA transport, stability and utilization factor (phosphoprotein)	primarily in nucleolus/nucleus shuttling between nucleolus and cytoplasm
Vif	p23	promotes virion maturation and infectivity	cytoplasm (cytosol, membranes) virion
Vpr	p10-15	promotes nuclear localization of preintegration complex, inhibits cell division, arrests infected cells at G2/M	virion, nucleus (nuclear membrane?)
Vpu	p16	promotes extracellular release of viral particles; degrades CD4 in the ER; (phosphoprotein only in HIV-1 and SIVcpz)	integral membrane protein
Nef	p27-p25	CD4 downregulation (myristylated protein)	plasma membrane, cytoplasm (virion?)
Vpx	p12-16	vpr homolog? (not in HIV-1, only in HIV-2 and SIV)	virion (nucleus?)
Tev	p28	tripartite tat-env-rev protein (also named Tnv)	primarily in nucleolus/nucleus

LANDMARKS ON THE HIV-1 GENOMIC RNA

PROTEINS

RNA SITES

SPLICE SITES AND EXONS

LANDMARKS:

HIV GENOMIC STRUCTURAL ELEMENTS

- LTR** Long terminal repeat, the DNA sequence flanking the genome of integrated proviruses. It contains important regulatory regions, especially those for transcription initiation and polyadenylation.
- TAR** Target sequence for viral transactivation, the binding site for Tat protein and for cellular proteins; consists of approximately the first 45 nucleotides of the viral mRNAs in HIV-1 (or the first 100 nucleotides in HIV-2 and SIV.) TAR RNA forms a hairpin stem-loop structure with a side bulge; the bulge is necessary for Tat binding and function.
- RRE** Rev responsive element, an RNA element encoded within the env region of HIV-1. It consists of approximately 200 nucleotides (positions 7327 to 7530 from the start of transcription in HIV-1.) The RRE is necessary for Rev function; it contains a high affinity site for Rev; in all, approximately seven binding sites for Rev exist within the RRE RNA. Other lentiviruses (HIV-2, SIV, visna, CAEV) have similar RRE elements in similar locations within env, while HTLVs have an analogous RNA element (RXRE) serving the same purpose within their LTR; RRE is the binding site for Rev protein, while RXRE is the binding site for Rex protein. RRE (and RXRE) form complex secondary structures, necessary for specific protein binding.
- CRS** cis-acting repressive sequences postulated to inhibit structural protein expression in the absence of Rev. One such site was mapped within the pol region of HIV-1. The exact function has not been defined; splice sites have been postulated to act as CRS sequences.
- INS** Inhibitory/Instability RNA sequences found within the structural genes of HIV-1 and of other complex retroviruses. Multiple INS elements exist within the genome and can act independently; one of the best characterized elements spans nucleotides 414 to 631 in the gag region of HIV-1. The INS elements have been defined by functional assays as elements that inhibit expression posttranscriptionally. Mutation of the RNA elements was shown to lead to INS inactivation and up regulation of gene expression.

GENES AND GENE PRODUCTS

- GAG** genomic region encoding the capsid proteins (group specific antigens). The precursor is the p55 myristylated protein, which is processed to p17 (MAtrix), p24 (CApsid), p7 (NucleoCApsid), and p6 proteins, by the viral protease. Gag associates with the plasma membrane where the virus assembly takes place. The 55 kDa Gag precursor is called assemblin to indicate its role in viral assembly.
- POL** the genomic region encoding the viral enzymes protease, reverse transcriptase and integrase. These enzymes are produced as a Gag-pol precursor polyprotein, which is processed by the viral protease; the Gag-pol precursor is produced by ribosome frameshifting at the C-terminus of gag.
- ENV** viral glycoproteins produced as a precursor (gp160) and processed to the external glycoprotein gp120 and the transmembrane glycoprotein gp41. The mature proteins are held together by non-covalent interactions; as a result, a substantial amount of gp120 is released in the medium. gp120 contains the binding site for the CD4 receptor.
- TAT** Transactivator of HIV gene expression. One of the two necessary viral regulatory factors (Tat and Rev) for HIV gene expression. Two forms are known, Tat-1exon (minor form) of 72 amino acids and Tat-2exon (major form) of 86 amino acids. The electrophoretic mobility of these two forms in SDS gels is anomalous, with apparent sizes of approximately 16 kD and 14 kD for Tat- 2exon and Tat-1exon, respectively. Low levels of both proteins are found in persistently infected cells. Tat has been localized primarily in the nucleolus/nucleus by immunofluorescence. It acts by binding to the TAR RNA element and activating transcription initiation and/or elongation from the LTR promoter. It is the first eukaryotic transcription factor known to interact with RNA rather

than DNA and may have similarities with prokaryotic anti-termination factors. Extracellular Tat can be found and can be taken up by cells in culture.

REV The second necessary regulatory factor for HIV expression. A 19 kD phosphoprotein, localized primarily in the nucleolus/nucleus, Rev acts by binding to RRE and promoting the nuclear export, stabilization and utilization of the viral mRNAs containing RRE. Rev is considered the most functionally conserved regulatory protein of lentiviruses. Rev cycles rapidly between the nucleus and the cytoplasm.

VIF Viral infectivity factor, typically 23 kD. Promotes the infectivity but not the production of viral particles. In the absence of Vif the produced viral particles are defective, while the cell-to-cell transmission of virus is not affected significantly. Found in almost all lentiviruses, Vif is a cytoplasmic protein, existing in both a soluble cytosolic form and a membrane-associated form. The latter form of Vif is a peripheral membrane protein that is tightly associated with the cytoplasmic side of cellular membranes. Some recent observations suggest that Vif is incorporated in the virion.

VPR Vpr (viral protein R) is a 96-amino acid (14 kD) protein, which is incorporated into the virion. It interacts with the p6gag part of the Pr55gag precursor. Vpr detected in the cell is localized to the nucleus. Proposed functions for Vpr include the nuclear import of preintegration complexes, cell growth arrest, transactivation of cellular genes, and induction of cellular differentiation. Found in HIV-1, HIV-2, SIVmac and SIVmdr. It is homologous to VPX of SIVagm.

VPU Vpu (viral protein U) is unique to HIV-1 and SIVcpz, a close relative of HIV-1. There is no similar gene in HIV-2 or SIV. Vpu is a 16-kD (81-amino acid) type I integral membrane protein with at least two different biological functions: (a) degradation of CD4 in the endoplasmic reticulum, and (b) enhancement of virion release from the plasma membrane of HIV-1-infected cells. Vpu probably possesses an N-terminal hydrophobic membrane anchor and a hydrophilic moiety. It is phosphorylated by casein kinase II at positions Ser52 and Ser56. Vpu is involved in env maturation; not found in the virion.

NEF (previously named 3' ORF) is an approximately 27-kD myristylated protein produced by an ORF located at the 3' end of the primate lentiviruses. Other forms of Nef are known, including nonmyristylated variants. Nef is predominantly cytoplasmic and associated with the plasma membrane via the myristyl residue linked to the conserved second amino acid (Gly). Nef has also been identified in the nucleus and found associated with the cytoskeleton in some experiments. Its association with the virion is suspected but not proven. One of the first HIV proteins to be produced in infected cells, it is the most immunogenic of the accessory proteins. Initially thought to be a negative factor, Nef was found to be important for viral replication in vivo. The nef genes of HIV and SIV are dispensable in vitro, but are essential for efficient viral spread and disease progression in vivo. Nef is necessary for the maintenance of high virus loads and for the development of AIDS in macaques. Nef downregulates CD4, the primary viral receptor, and is also proposed to increase viral infectivity. Nef contains PxxP motifs that bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of HIV but not for the downregulation of CD4.

VPX Virion protein of 12 kD found only in HIV-2/SIVagm and not in HIV-1 or SIVmdr. Vpx function in relation to Vpr is not fully elucidated. Vpx is necessary for efficient replication of SIV in PBMCs. Some studies indicate that Vpx and Vpr proteins may be functionally distinct. Progression to AIDS and death in SIV-infected animals can occur in the absence of Vpr or Vpx. Double mutant virus lacking both vpr and vpx was severely attenuated, whereas the single mutants were not, suggesting a redundancy in the function of Vpr and Vpx related to virus pathogenicity.

TEV (also named *tnv*) tripartite 28 kD viral phosphoprotein produced by some HIV-1 strains. Found primarily in the nucleolus/nucleus. Tev contains the first exon of Tat, a small part of Env and the second exon of Rev. It exhibits both Tat and Rev functions and can functionally replace both essential regulatory proteins of HIV-1. It is produced very early in infection.

Landmarks

STRUCTURAL PROTEINS/VIRAL ENZYMES The products of gag, pol and env genes, which are essential components of the retroviral particle.

REGULATORY PROTEINS Tat and Rev proteins of HIV/SIV and Tax and Rex proteins of HTLVs. They modulate transcriptional and posttranscriptional steps of virus gene expression and are essential for virus propagation.

ACCESSORY OR AUXILIARY PROTEINS additional virion and non-virion- associated proteins produced by HIV/SIV retroviruses: Vif, Vpr, Vpu, Vpx, Nef. Although the accessory proteins are in general not necessary for viral propagation in tissue culture, they have been conserved in the different isolates; this conservation and experimental observations suggest that their role in vivo is very important.

COMPLEX RETROVIRUSES Retroviruses regulating their expression via viral factors and expressing additional proteins (regulatory and accessory) essential for their life cycle.

I

HIV-1/CPZ and HIV-2/SIV Nucleotide Alignments of Complete Genomes

Contents Part I 1997

PART I HIV-1/CPZ and HIV-2/SIV Complete Genome Nucleotide Alignments

Contents	I-1
Introduction	I-2
Summary of HIV-1 Sequences in the Nucleotide Alignment	I-5
Notes on full-length HIV-1 Sequences in the Nucleotide Alignment	I-7
Summary Table of HIV-2/SIV Sequences in the Nucleotide Alignment	I-17
 Nucleotide Alignment of HIV-1 Complete Genomes	 I-18
Nucleotide Alignment of HIV-2/SIV Complete Genomes	I-92

HIV-1

Introduction This year many new full length viral sequences have become available, originating from diverse geographic origins and representing the spectrum of known HIV variation. Thus we have decided to publish only full length HIV-1/CPZ and HIV-2/SIV sequences in our printed nucleotide alignment section, as this set is now becoming an adequate representation of the overall diversity of the virus.

As of November 1997 there were 63 complete (or nearly complete) HIV-1 genomes in the database (all are listed in Table 1). Of these, some were not included in the printed alignment, as they are very closely related, but the complete alignment including all sequences is available at our WEB (<http://hiv-web.lanl.gov/>) and ftp sites. Ten sequences were excluded because they were related to the IIB/LAI lab strain of HIV-1, which is represented by the two sequences LAI and HXB2 in the alignment. Three more sequences have regions which contain IIB/LAI, spliced onto regions of new isolate sequence (F12 Z11530, and WC001 AF003888, AF003887). A fourteenth sequence was omitted because it was a re-sequencing of the Z2Z6 sequence (Z2Z6 K03458) already represented by Z2Z6 M22639. This left 49 HIV-1 sequences in the alignment. Sequences from viral strains isolated from chimpanzees, CPZANT and CPZGAB, were also included. In phylogenetic analyses, the CPZ sequences are the closest simian-derived viruses to HIV-1; in fact HIV-1 M group and HIV-1 O group sequences are roughly as distant from one another as they are from the CPZ sequences.

All 63 of the complete genomes for HIV-1 have been updated with annotation of the major gene start and end sites. All are available from the HIV database WWW site (<http://hiv-web.lanl.gov/>) by using the sequence search interface (http://retro.lanl.gov/hivDB_search/index.html) to search for HIV-1 sequences with length ranging from 5,000 to 12,000 bases (this will select for only the full length or near complete genome sequences when using in the search tool).

The sequences are identified by their common name preceded by the HIV subtype designations appropriate for the sequence. The primary sequence reference, country of origin, database accession number, and brief notes describing the isolate and sequence, with some additional relevant references, can be found in Table 2 for the set of sequences included in the printed alignment. The sequences that have been found to be recombinants with portions of the genetic sequences associated with different subtypes are indicated by listing the all of the subtypes in the prefix to the name. For example, the prefix AG simply indicates that some regions of the sequence are subtype A-like, others G-like. The subtypes are organized alphabetically and not meant to reflect the proportion of either subtype in the mosaic genome. The "T" subtype is still in the process of being characterized. (See Robertson, D., et al., part III pages 25-30, of this compendium, (1997), for further details concerning the I subtype; the patterns of the inter-subtype recombination breakpoints for the all of the recombinant full length genomes are also can be found in Robertson et al.)

Alignment This alignment was generated by using the HMMER Hidden Markov Model sequence alignment software developed by Sean Eddy (available at <http://genome.westl.edu/eddy/hmmer.html> and <http://predict.sanger.ac.uk/mirrors/hmm/hmm.html>), using an iterative process involving alignment of the genomes using HMMER, followed by hand-editing (using an in-house revised version of the MASE alignment editing program (Faulkner, D., and Jurka, A., Trends in Biochem. Science, 13:321-322 (1988)), and SE-AL V1.d1, 1995, a beta test version of an alignment program developed by Andrew Rambaut at Oxford), followed by rebuilding a new HMMER model and realigning the sequences, and then more hand editing. The resulting final alignment is not suggested to be an "optimal alignment" with the absolute minimum number of gaps and mismatches. It is a compromise between optimal alignment, readability, and an attempt to keep insertions and deletions from altering the protein reading frame presentation. Most gaps have been introduced in multiples of 3 bases to maintain open reading frames when translated directly from the alignment.

After the final alignment was generated, a HMMER model was built with the hmmb program, using this alignment as the input or training set. The same set of complete genomes were then realigned using this model, to test the performance of HMMER with this model. The hmmb -R option (for ragged-end alignment due to not all LTRs being complete) was used. The resulting alignment was very

comparable to the alignment presented here, except that gaps were not preferentially placed to preserve codons. The average nucleotide identity between B_WEAU and the other 60 sequences was 0.88572 in the automated alignment and 0.88520 in the alignment presented here. Thus the automated alignment is closer to optimal than this alignment which has been adjusted by hand to be more presentable (*i.e.*, gaps moved from within codons to between codons).

The final HMMER model based on the full length genomes has been tested here with partial genomes as well. Using the HMMER -R option for ragged ends (gaps inserted at the ends of sequences are given very low weight) the HMMER program did a reasonable job of aligning the complete and partial env genes to each other. The model was used again to align the complete genomes plus the env gene sequences, and in this case all sequences were reasonably aligned to each other. We are in the process of making these models available at our web site.

The annotation. The annotation for the precursor peptide cleavage sites in Gag and Gag-Pol is based on the information published in [Tozser et al.(1991), Le Grice et al.(1989)]. The annotation of the Gag-Pol ribosomal slip site is based on information published in [Reil et al.(1993), Kollmus et al.(1994), Le et al.(1989)]. The annotation for the cis-acting transcriptional activation domains in the LTR section is based on information published in [Zhang et al.(1997), Estable et al.(1996), Montano et al.(1997), Gao et al.(1996)]. There are a varying number of NF-kappaB binding sites in C subtype sequences, with some sequences carrying an additional site [Gao et al.(1996), Carr et al.(1996), Montano et al.(1997)]. The annotation for the Rev responsive element (the RRE) is based on [Charpentier et al.(1997)].

The WEAU reference nucleotide reference sequence is translated into all three reading frames at the top of the alignment using the single character amino acid designation. At the bottom of the alignment, protein sequences, based on the WEAU sequence are indicated; the HIV genome has many overlapping coding regions, and all are shown. For more complete annotation of functional domains see the protein sequence alignments in Part II.

HIV-2/SIV

Alignments of the 26 HIV-2/SIV full length genomes are included as a separate nucleotide alignment in Part I, following a similar annotation strategy as the HIV-1/CPZ alignment. Table 3 includes a list of all sequences included here, the subtype associated with each sequence, the GenBank accession number, and the primary reference. Sometimes a second reference is included if it contains basic information about the sequence or the isolate.

This alignment is also available at our WEB (<http://hiv-web.lanl.gov/>) and ftp sites. All corresponding sequence entries are available from the HIV database WWW site (<http://hiv-web.lanl.gov/>) by using the sequence search interface (http://retro.lanl.gov/hivDB_search/index.html) to search for HIV-2 and SIV sequences with length ranging from 9,000 to 12,000 bases.

The alignment was prepared by adding new sequences to an existing alignment using an in-house revised version of MASE, Multiple Alignment Sequence Editor (Faulkner, D., and Jurka, A., Trends in Biochem. Science, 13:321-322 (1988)). Phylogenetic analysis using neighbor-joining was used to check subtype designations. The STM (stump-tailed macaque) sequence is not included in the SIV-SD subtype since it is phylogenetically distinct from the other simian sequences.

The HIV-2 isolate BEN was selected as the reference sequence for this alignment since it is representative of the larger HIV-2 subtype, A, and it is one of the longer HIV-2 sequences available.

SIV sequences from other simian species such as African Green monkeys (AGM), sykes (SYK), and mandrill (MND) are not presented in this alignment. They are phylogenetically more distant from the HIV-2s than the macaques and mangabeys. The complete genome nucleotide sequence alignment of the sequences from African Green monkeys will be presented on the web and ftp sites when it is complete.

Introduction

References

- [Carr et al.(1996)] J. K. Carr, M. O. Salmineen, C. Koch, D. Gotte, A. W. Artenstein, P. A. Hegerich, L. S. D., D. S. Burke, & F. B. McCutchan. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. LA - Eng. *J Virol* 70:5935-43, 1996. OTE: (Medline: 96323109), (Genbank: U54771).
- [Charpentier et al.(1997)] B. Charpentier, F. Schultz, & M. Rosbash. A dynamic *in vivo* view of the HIV-1 Rev-RRE interaction. *J Mol Biol* 266:950-962, 1997.
- [Estable et al.(1996)] M. C. Estable, B. Bell, A. Merzouki, J. S. Montaner, M. V. O'Shaughnessy, & I. J. Sadowski. Human immunodeficiency virus type 1 long terminal repeat variants from 42 patients representing all stages of infection display a wide range of sequence polymorphism and transcription activity. *J Virol* 70:4053-62, 1996.
- [Gao et al.(1996)] F. Gao, D. L. Robertson, S. G. Morrison, H. Hui, S. Craig, J. Decker, P. N. Fultz, M. Gerard, G. M. Shaw, B. H. Hahn, & P. M. Sharp. The heterosexual human immunodeficiency virus type 1 epidemic in thailand is caused by an intersubtype (A/E) recombinant of african origin. *J Virol* 70:7013-29, 1996.
- [Kollmus et al.(1994)] H. Kollmus, A. Honigman, A. Panet, & H. Hauser. The sequences of and distance between two cis-acting signals determine the efficiency of ribosomal frameshifting in human immunodeficiency virus type 1 and human t-cell leukemia virus type ii in vivo. *J Virol* 68:6087-91, 1994.
- [Le et al.(1989)] S. Y. Le, J. H. Chen, & J. V. Maizel. Thermodynamic stability and statistical significance of potential stem-loop structures situated at the frameshift sites of retroviruses. *Nucleic Acids Res* 17:6143-52, 1989.
- [Le Grice et al.(1989)] S. F. Le Grice, R. Ete, J. Mills, & J. Mous. Comparison of the human immunodeficiency virus type 1 and 2 proteases by hybrid gene construction and trans-complementation. *J Biol Chem* 264:14902-8, 1989.
- [Montano et al.(1997)] M. A. Montano, V. A. Novitsky, J. T. Blackard, N. L. Cho, D. A. Katzenstein, & M. Essex. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. *J Virol* 71:8657-65, 1997.
- [Reil et al.(1993)] H. Reil, H. Kollmus, U. H. Weidle, & H. Hauser. A heptanucleotide sequence mediates ribosomal frameshifting in mammalian cells. *J Virol* 67:5579-84, 1993.
- [Tozser et al.(1991)] J. Tozser, I. Blaha, T. D. Copeland, E. M. Wondrak, & S. Oroszlan. Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in gag and gag-pol polyproteins. *FEBS Lett* 281:77-80, 1991.
- [Zhang et al.(1997)] L. Zhang, Y. Huang, H. Yuan, B. K. Chen, J. Ip, & D. D. Ho. Genotypic and phenotypic characterization of long terminal repeat sequences from long-term survivors of human immunodeficiency virus type 1 infection. *J Virol* 71:5608-13, 1997.

Table 1 Table of HIV-1 full length genomes. Only one of each related set of sequences is included in the alignment displayed in the hard copy of the compendium, but all are available at the HIV database web site: <http://hiv-web.lanl.gov/>

No.	Name	Accession	Country of Origin	Comment
1	B-WEAU	U21135	USA	Immunology Reference Strain
2	B-MN	M17449	USA	1984 6 yr-old AIDS patient, died in 1987
3	B-AD8	AF004394	USA	Macrophage-tropic, derived from T-cell tropic ADA
4	B-WR27	U26546	USA	
5	B-RF	M17451	Haiti/USA	
6	B-89.6	U39362	USA	
7	B-BCSG3	L02317	USA	Chimpanzee adapted, highly T-cell cytopathic
8	B-JRFL	U63632	USA	
9	B-JRCSE	M38429	USA	
10	B-OYI	M26727	Gabon	
11	B-CAM1	D10112	United Kingdom	
12	B-NY5	M38431	USA	
13	B-LAI	K02013	France	IIIB/LAI (formerly BRU)
14	B-LAI	A04321	France	LAI/IIIB, the first HIV-1 isolate (formerly called BRU)
15	B-pNL43.420	U26942	France/USA	LAI/NY5 lab recombinant
16	B-pNL43	M19921	France/USA	LAI/NY5 lab recombinant
17	B-LW123	U12055	France	1987 sample from lab worker infected with LAI
18	B-BH10	M15654	France	Derived from IIIB/LAI
19	B-IIIBW	X01762	France	Derived from IIIB/LAI
20	B-PV22	K02083	France	Derived from IIIB/LAI
21	B-MCK1	D86068	France	Derived from IIIB/LAI
22	B-HXB2R	K03455	France	Derived from IIIB/LAI
23	B-PM213	D86069	France	Derived from IIIB/LAI
24	B-TH475	L31963	France	Derived from IIIB/LAI
25	B-F12	Z11530	France	Supposed to be from an Italian AIDS patient, but highly similar to IIIB/LAI in Env, not in other regions
26	B-YU2	M93258	USA	
27	B-YU10	M93259	USA	
28	B-ACH320A	U34604	The Netherlands	M64489-M64492
29	B-ACH320B	U34603	The Netherlands	M64489-M64492
30	B-SF2	K02007	USA	
31	B-D31	U43096	Germany	
32	B-MANC	U23487	United Kingdom	
33	B-HAN	U43141	Germany	
34	B-C18MBC	U37270	Australia	Nef deletion, long-term survivor
35	B-WC001	AF003887	USA	IIIB/Patient WC001 Chimeric
36	B-WC001	AF003888	USA	

Introduction

Table 1 (cont.)

No.	Name	Accession	Country of Origin	Comment
35	BF-93BR029	AF005495	Brazil	
36	F-93BR020	AF005494	Brazil	
37	H-90CR056	AF005496	Central African Republic	
38	AC-ZAM184	U86780	Zambia	
39	AC-92RW009	U88823	Rwanda	
40	C-ETH2220	U46016	Ethiopia	
41	C-92BR025	U52953	Brazil	
42	AG-92NG083	U88826	Nigeria	
43	AG-92NG003	U88825	Nigeria	
44	AG-Z321	U76035	Zaire	
45	AG-IBNG	L39106	Nigeria	
46	A-U455	M62320	Uganda	
47	A-92UG037	U51190	Uganda	
48	A-Q23	AF004885	Kenya	
49	AE-93TH253	U51189	Thailand	
50	AE-CM240	U54771	Thailand	
51	AE-90CF402	U51188	Central African Republic	
52	AD-MAL	X04415	Zaire	
53	D-84ZR085	U26942	Zaire	
54	D-NDK	M27323	Zaire	
55	D-Z226	M22639	Zaire	
56	D-ELI	K03454	Zaire	
57	D-94UG114	U88824	Uganda	
58	O-MVP5180	L20571	Cameroon	
59	O-ANT70	L20587	Cameroon	
60	SIVCPZANT	U42720	Zaire	
61	SIVCPZGAB	X52154	Gabon	

Table 2 Notes on selected full-length nucleotide sequences that appear in the alignments.

Sequence	Accession	Origin	Reference
B-WEAU	U21135	USA	Ghosh, S. K., Unpublished (1995)
<p>WEAU is the reference strain for this alignment as well as for the Los Alamos Molecular Immunology Compendium, the companion volume for this HIV sequence compendium. The sequence is from a cytopathic HIV-1 virus clone from an isolate from an acutely infected patient, from Birmingham, Alabama. The lambda phage clone was obtained from a co-culture of this patient's PBMCs, first with normal donor PHA-stimulated lymphocytes for 14 days, and then with the H9 T-cell line for another 14 days. The blood specimen was obtained 15 days after the onset of clinical symptoms of acute (primary) infection, and 35 days after a single sexual encounter (receptive anal intercourse) with a partner whose virus was proven phylogenetically to be responsible for the transmission event. The clone that was sequenced has an SI phenotype, and a frameshift in <i>nef</i>. The single nucleotide deletion in <i>nef</i> found in the full length WEAU 1.60 clone is not present in 10/10 PCR amplified sequences from the patient's uncultured PBMCs (instead there is a "T"). The full-length WEAU 1.60 provirus was sequenced in its entirety by two different laboratories (George Shaw and Leroy Hood) with 100% concordance. George Shaw provided detailed information concerning this sequence. The patient WEAU is "Patient #1" in Clark, S. J. et al., <i>N Engl J Med</i> 324:954-960 (1991); "WEAU 0575" in Piatak, M., et al., <i>Science</i> 259:1749-1754 (1993) and is also discussed in Borrow, P., et al., <i>Nat Med</i>, 3:205-11 (1997).</p>			
A-U455	M62320	Uganda	Ozom, J. et al., ARHR 6:1073-1078 (1990)
<p>This sequence is from the 1985 Ugandan isolate U455. It was cloned in phage, and is defective <i>env</i>, <i>vpr</i> and <i>vpu</i>. The <i>env</i> ORF in this sequence is interrupted by an in-frame stop codon beyond the COOH end of the V5 region. This sequence clusters with subtype A HIV-1.</p>			
A-92UG037.1	U51190	Uganda	Gao, F. et al., submitted (1997)
<p>Sample 92UG037 is part of a set of isolates obtained through the WHO Global Programme on AIDS (WHO Network, ARHR 10:1327-1344 (1994)) and comes from an asymptomatic 31-year-old female from Entebbe, Uganda; she had not taken any anti-retroviral therapy prior to sampling. The risk factor for infection was heterosexual contact. The isolate 92UG037 was established and propagated by short term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified and sequenced. 92UG037 is subtype A. An LTR sequence is available under accession numbers U51287, and an additional <i>env/nef</i> sequence with accession number U09127. There is an inframe stop codon in <i>pol</i> at position 3144 in this clone. The isolate from which this sequence was derived is NSI and uses CCR5 or CCR8 (Bjornstal, A. et al., <i>J Virol</i> 71:7478 (1997) and Rucker, J., et al., <i>J Virol</i> 71:8999-9007 (1997). See also: Gao, F. et al., <i>J Virol</i> 70:7013-7029 (1996) This sequence was kindly made available prior to publication, and the manuscript is currently submitted (Gao F., et al., (1997)).</p>			
A-Q23.17	AF004885	Kenya	Poss, M. Unpublished (1997)
<p>This subtype A sequence was from a woman from Mombasa, Kenya, who had been recently infected with HIV-1. The blood sample was drawn in 1993. An <i>env</i> gene fragment from a PCR amplification from this same blood sample was published in Poss M., et al. ARHR 13:493-499 (1997). The full length sequence was kindly released prior to publication by M. Poss and colleagues, U. Washington.</p>			

Table 2 (cont.)

Sequence	Accession	Origin	Reference
B-HIVMN	M17449	USA	Gurgo C., et al., <i>Virology</i> 164:531-536 (1988)
MN is from one of the earliest available isolates, and is a commonly used reference and vaccine strain. The MN isolate was taken from a 6 year old male pediatric AIDS patient in 1984. The MN sequence was cloned from the isolate in lambda phage. The coding sequences for pol, nef and vpu are prematurely truncated. A set of V3 sequences from this isolate are available L48364-L48379, Lukashov, V. and Goudsmit J., <i>AIDS</i> 9:1307-1311 (1995)			
B-HIV1AD8	AF004394	USA	Theodore, T. S., et al., <i>ARHR</i> 12:191-194 (1996)
AD8 is a molecular clone that was replication competent after reconstruction. It has a defective vpu due to the loss of the start codon. It was derived from the macrophage tropic isolate AD87. AD87 was derived from isolate ADA, described in Westervelt, P., et al., 88:3097-3101 <i>PNAS</i> (1991)			
B-WR27	U26546	USA	Salminen, M. O., et al., <i>Virology</i> 213:80-86 (1995)
WR27 was a patient with clinical progression to Walter Reed Stage 5 when blood was drawn for viral isolation in 1988. This sequence is from a PCR clone from a primary isolate that was expanded in donor PBMC. The virus sequence had a V3 loop predictive of an SI phenotype. Rev has an inframe stop codon in both exons.			
B-RF	M17451	Haiti	Starcich, B. R., et al., <i>Cell</i> 45:637-648 (1986)
RF is among the first isolates, and is commonly used reference and vaccine study strain. The sequence is from the full-length lambda phage clone HAT-3, from isolate RF, cultured in HUT-78 cells. RF is from a 28 year old symptomatic Haitian male, who moved to the U.S. in 1980 and was sampled in 1983, shortly before his death. RF has defective gag and vpu genes. Several env genes are available from this isolate, U30778-U30781. See also: Reitz M., et al., <i>ARHR</i> 8:1950 (1992)			
B-89.6	U39362	USA	Collman, R., et al., <i>J Virol</i> 66:7517-7521 (1992)
89.6 is cloned from a highly cytopathic primary macrophage-tropic isolate from an AIDS patient living in Philadelphia, although the patient was originally from Jamaica. 89.6 was cloned in phage and was replication competent after reconstruction. Also see Kim, F., et al., <i>J Virol</i> 69:1755-1761 (1995)			
B-BCSG3	L02317	USA	Ghosh, S. K., et al., <i>Virology</i> 194:858-864 (1993)
The isolate BC was obtained in 1987, from an individual from the U.S. with terminal AIDS. The clone was obtained after short term co-culture of HUT-78 cells with a PHA-stimulated PBMCs. BCSG3 is the sequence from the provirus SG3; the full length provirus was cloned intact in lambda phage. It is very cytopathic in human T-cell lines, and did not replicate in monocyte-macrophages. BC and BCSG3 replicate more efficiently in chimpanzee than in human lymphocytes. All typical HIV-1 genes are intact except vpu, which is normally 243 bp in length; vpu was disrupted after 70 nucleotides by a 23 bp deletion that results in a frameshift and truncation of the open reading frame.			
B-JRFL	U63632	USA	O'Brien, W. A., et al., <i>Nature</i> 348:69-73 (1990)
This sequence is from an infectious lambda phage clone of the 1986 isolate JRFL, derived from from the brain of a patient who died with Kaposi's sarcoma and severe AIDS encephalopathy. The infectious clone JRCSF was isolated from the CSF of the same patient. Both are NSI. Also see: Pang, J., et al., <i>JAIDS</i> , 4:1082-92 (1991) and Klasse, P. J., et al., <i>ARHR</i> 12:347-350 (1996).			

Table 2 (cont.)

Sequence	Accession	Origin	Reference
B-JRCSF	M38429	USA	O'Brien, W. A., et al., <i>Nature</i> 348: 69-73 (1990)
This sequence is from an infectious lambda phage clone of the 1986 isolate JRCSF, derived from from the CSF of a patient who died with Kaposi's sarcoma and severe AIDS encephalopathy. The infectious clone JRFL was isolated from the brain of the same patient. Both are NSI. A SI mutant clone of this isolate has also been sequenced in env U45960. Also see: Pang, J., et al., <i>JAIDS</i> , 4:1082-92 (1991) and Klasse, P. J., et al., <i>ARHR</i> 12:347-350 (1996).			
B-OYI	M26727	Gabon	Wain-Hobson, S., et al., <i>AIDS</i> 3:707 (1989)
This sequence is derived from the Gabonese isolate OYI, isolated from a healthy HIV-1 infected individual. GA.OYI was the first viral sequence from Africa that was found to phylogenetically cluster with North American viruses. This sequence is from a lambda phage clone, and the cloned provirus was functionally defective. The vpu gene does not have a start codon.			
B-CAM1	D10112	UK	McIntosh, A., Unpublished (1991)
This sequence is from the British isolate CAM1. It has a defective vpu gene. McIntosh A, and Karpas A, Thesis (1991), Cambridge University, England.			
B-NY5	M38431	USA	Wiley, R., et al., <i>PNAS</i> 83:5038-5042 (1986)
This sequence is from the 1984 New York T-cell tropic isolate NY5. It was cloned in lambda phage and is not replication competent. It has a defective vpu gene due to the loss of the start codon. See also GenBank accession number K03346, for an env gene sequence from this isolate.			
B-pNL43	M19921	USA/France	Adachi, A., et al., <i>J Virol</i> 59, 284-291 (1986)
A NY5/LAI laboratory generated chimera that has been extensively studied, with envelope from LAI(BRU) and gag-pol from NY5.			
B-HXB2	K03455	France	Wong-Staal F., et al., <i>Nature</i> 313:277-284 (1985)
This sequence was from provirus cloned in lambda phage, and is derived from the IIIB isolate related to LAI (see LAI, below). This clone has been extensively studied.			
B-LAI	K02013	France	Wain-Hobson, S., et al., <i>Cell</i> 40:9-17 (1985)
This sequence is from the French isolate LAI (formerly called BRU), and was obtained from a complete provirus cloned in lambda phage. This is an infectious molecular clone after reconstruction. Also see: Alizon, M. et al., <i>Cell</i> 46:63-74 (1986), and Wain-Hobson, S., <i>Science</i> 252:961-965 (1991), for clarification on the origins of early French isolates. The IIIB strain is also very closely related to LAI. There are several other LAI-IIIB related clones that have been fully sequenced (LAI/BRU, BH10, PV22, PM213, MCK1, LW12-3, HXB2, and TH475), but in our printed compendium we include only the two common reference strains HXB2 and LAI; our ftp site alignment includes the full set of full length LAI-related HIV sequences. These were the first HIV isolates, are T-cell tropic, and are the most extensively studied variants in terms of mutational analysis of functional domains, phenotype, vaccine design, and antigenicity. The LW12-3 sequence is a complete viral genome derived from a IIIB/LAI infected lab worker who was first infected in 1985, sampled in 1987, and who was still asymptomatic in 1994 (Reitz, M., <i>ARHR</i> 10:1143-1155 (1994)). As the use of LAI-related cultures, reagents and clones is ubiquitous in AIDS research, these sequences are occasionally found as contaminants in sequencing studies (although LAI/IIIB sequences are certainly not the only possible source of contamination, they are the most common). A very unusual QR insertion in the V3 loop of the Envelope protein has been used as red flag to signal the possibility of LAI/IIIB contaminations.			

Table 2 (cont.)

Sequence	Accession	Origin	Reference
B-YU2	M93258	USA	Li, Y., et al., J Virol 66:6587-6600 (1992)
YU-2 is a lambda phage clone that is replication competent. It was from the uncultured brain tissue of a patient with AIDS dementia complex. YU2 and YU10 differ by 0.26% in nucleotide sequence, YU-2 was fully replication competent after reconstruction, in both primary T lymphocytes and monocyte-macrophages. YU-2 a defective vpu gene due to the loss of the start codon. Also see Li, Y., et al., J Virol 65: 3973-3985 (1991).			
B-YU10	M93259	USA	Li, Y., et al., J Virol 66:6587-6600 (1992)
YU10 is a sequenced lambda phage clone from the same source as YU2, the uncultured brain tissue of a patient with AIDS dementia complex. There is a frame shift mutation in the pol gene in YU10. YU2 and YU10 differ by 0.26% in nucleotide sequence. Also see Li, Y., et al., J Virol 65:3973-3985 (1991).			
B-ACH320.2A.2.1	U34604	The Netherlands	Guillon, C., et al., ARHR 11:1537-1538 (1995)
This is the sequence of the complete genome of an NSI, macrophage tropic lambda phage clone from an isolate taken from the PBMC of a patient who was making a shift from NSI to SI in bulk phenotype. A syncytium inducing clone from the same isolate has also been completely sequenced, ACH320.2A.1.2. The patient, isolates and phenotype of the molecular clones are described in Groenink, M., et al., J Virol 65:1968-1975 (1991).			
B-ACH320.2A.1.2	U34603	The Netherlands	Guillon, C., et al., ARHR 11:1537-1538 (1995)
This is the sequence of the complete genome of an SI, syncytium-inducing lambda phage clone from an isolate taken from the PBMC a patient who was making a shift from NSI to SI in bulk phenotype. An NSI clone from the same isolate has also been completely sequenced, ACH320.2A.2.1. The patient, isolates and phenotype of the molecular clones are described in Groenink, M., et al., J Virol 65:1968-1975 (1991)			
B-SF2	K02007	USA	Sanchez-Pescador, R., et al., Science 227:484-492 (1985)
This sequence is from an infectious phage clone from the U.S. isolate ARV-2. ARV-2/SF2 was isolated from the PBMC of a patient with oral candidiasis after co-culture with mitogen-stimulated PBMCs, (Levy, J., et al., Science 225:840-842, (1984)). It is a standard reference strain, and has been used for vaccine studies.			
B-D31	U43096	Germany	Kreutz, R., et al., ARHR 8:1619-1629 (1992)
This sequence is from isolate D31, about which little information is available.			
B-MANC	U23487	UK	Zhu, T., and Ho, D. Nature 374:503-504 (1995)
This sequence was PCR amplified from the 1959 sample "Manchester sailor" kidney tissue (see: Corbett, G., et al., Lancet 336:51 (1990)). The sequence of the complete genome is available, and it is indistinguishable from contemporary subtype B HIV-1 sequences in phylogenetic analysis. This information, together with the observation that additional tissue samples were HIV PCR negative, suggests that the HIV clone that came from the MANC kidney sample was very likely to be a contemporary clinical contaminant. The sequence was assembled from multiple PCR amplified fragments. All reading frames in this sequence are intact.			

Table 2 (cont.)

Sequence	Accession	Origin	Reference
B-HAN	U43141	Germany	Sauermann, U., et al., ARHR 6:813-823 (1990)
HAN2 was isolated from a 39 year old homosexual German patient with AIDS related complex in 1986. This patient died from complications of AIDS in 1987. HAN2 was highly cytopathic in MT-2 cells. Genomic DNA from infected MT-2 cells was used to prepare a lambda phage genomic library. Two full-length clones, HAN2/2 and HAN2/3 were purified. Clone HAN2/3 was used for DNA sequencing, and has a defective env gene.			
B-C18MBC	U37270	Australia	Deacon, N., et al., Science 270:988-991 (1995)
A sequence from one of the six Australian long term survivors (subject C18) that were all infected via blood transfusions with blood from the same HIV+ donor, D36, a homosexual male who was infected some time between Dec. 1980 and April 1981. All were healthy, and symptom-free 10 to 14 years after infection, in 1995, at the time of sampling. C18 was born in 1912, and was infected by transfusion in August of 1983. C18 PMBC DNA was PCR amplified in overlapping fragments which were cloned and sequenced. The virus is attenuated due to extensive deletions the nef coding region, and the U3 region of the LTR. Different members of the cohort have different patterns of nef deletions. The 18MBC sequence was assembled from multiple PCR amplified fragments. The authors point out the importance of Nef in determining the pathogenicity of HIV-1 and suggest this strain of HIV-1 as a possible basis for a live attenuated vaccine, or for further study of this approach.			
C-ETH2220	U46016	Ethiopia	Salminen, M. et al., ARHR 12:1329-1339 (1996)
ETH2220 is the first reported near full length subtype C sequence from Ethiopia. The patient sample from which this clone was obtained was taken in 1986. In its genomic organization, this clone closely resembles subtype A, B, and D isolates except that the core promoter contains three potential binding sites for the transcription factor NF-kB instead of containing two, a feature which was preserved in other Ethiopian C subtype samples, as well as C viruses from Zambia. This sequence was cloned as a PCR amplified near full length genome, and has a defective tat gene.			
C-92BR025.8	U52953	Brazil	Gao, F. et al., submitted (1997)
This sequence is from a PCR clone from a primary isolate that is part of a set of isolates obtained through the WHO Global Programme on AIDS (WHO Network, ARHR 10:1327-1344 (1994)). 92BR025 was from a 23 year old male hemophiliac from Porto Alegre, Brazil. He had recently seroconverted, though sometime more than 1.2 months prior to the date this blood sample was collected in 1992. He was asymptomatic, and had not taken any anti-retroviral therapy prior to sampling. The isolate 92BR025 was established and propagated by short-term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified and sequenced. The HIV isolate BR025 exhibited an NSI phenotype when assayed by the WHO. The clone has two inframe stop codons in pol at positions 2141, and 3115, and a frame shift mutation at position 4131. This sequence was kindly made available prior to publication, and the manuscript is currently submitted (Gao F., et al., submitted (1997)). Additional env, nef, and ltr region sequences are available from this isolate: U09126, U09132, U51282, and U15121.			
D-84ZR085	U88822	Zaire	Gao, F. et al., submitted (1997)
Sample 84ZR085 was obtained from an AIDS patient from Zaire. The near full length genome was cloned in phage and sequenced. D-84ZR085 is subtype D. There was a frame shift mutation in gag/pol, position 1692. This isolate was obtained from Thomas Jefferson University, and isolate phenotyping information was not available. This sequence was kindly made available prior to publication, and the manuscript is currently submitted (Gao, F., et al., submitted (1997)).			

Table 2 (cont.)

Sequence	Accession	Origin	Reference
D-NDK	M27323	Zaire	Spire, B. et al., <i>Gene</i> 81:275-284 (1989)
This is an infectious molecular clone derived from a very cytopathic isolate. It was cloned in phage and is replication competent. All reading frames in this sequence are intact. Spire et al. reported that only minor sequence differences appear to be responsible for the "acute biological effect". This sequences clusters with HIV-1 subtype D in phylogenetic analysis.			
D-Z226	M22639	Zaire	Theodore, T. and A. Buckler-White, Unpublished (1988)
An infectious molecular clone of this virus was created by reconstruction. It was cloned in phage and sequenced from the isolate Z2, also called CDC-Z34. All reading frames in this sequence are intact.			
D-ELI	K03454	Zaire	Alizon, M. et al., <i>Cell</i> 46:63-74 (1986)
This sequence is of a phage clone derived from the Zairean isolate ELI. ELI was recovered in 1983 from a 24 year old woman with AIDS. All reading frames are intact. A reconstructed infectious clone is available. <i>Gag</i> (M27954) and <i>env</i> (M27949) sequences from the same isolate are also available.			
D-94UG114.1	U88824	Uganda	Gao, F. et al., (1997)
Sample 94UG114 was obtained from an asymptomatic 31-year-old man from Butuku, Uganda, as part of the WHO/UNAIDS study. He had not taken any anti-retroviral therapy prior to sampling. His risk factor for infection was heterosexual contact. The near full length genome was PCR amplified from a short term culture of a PBMC sample and sequenced. 92UG114.1 is subtype D. There were no defective gene products, and an infectious molecular clone of 94UG114.1 is available. The isolate from which this sequence was derived is NSI by an MT-2 assay. This sequence was kindly made available prior to publication, and the manuscript is currently (Gao F., et al., submitted (1997)).			
F-93BR020.1	AF005494	Brazil	Gao, F. et al., submitted, (1997)
This isolate is part of a part of a set obtained through the WHO Global Programme on AIDS (WHO Network, ARHR 10:1327-1344 (1994)) and came from a asymptomatic HIV seropositive 52 year old man from Rio de Janeiro, Brazil, sampled in 1993. The risk factor for infection was bisexual contact. The isolate 92BR020 was established and propagated by short term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified and sequenced. The isolate 93BR020 was described as syncytium inducing (SI) using an MT-2 assay. An envelope gene sequence from this isolate is described in Gao, F., et al., <i>J Virol</i> 70, 1651-1657 (1996). This sequence was kindly made available prior to publication, and the manuscript is currently submitted Gao, F., et al., (1997)).			
H-90CF056.1	AF005496	CAR	Gao, F. et al., (1997)
This sequence clusters with available HIV-1 subtype H sequences in phylogenetic analysis, and is the first available full length H subtype sequence. The isolate comes from Bangui, in the Central African Republic, and was sampled in 1990, from an asymptomatic individual, who had no anti-retroviral therapy. The isolate had an NSI phenotype by an MT-2 assay, and the sample was obtained from the Pasteur Institute, Bangui. The isolate 90CF056 was established and propagated by short term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified, cloned and sequenced. The isolate was at one point designated 90CR056, but was changed to 90CF056 as CR stands for Costa Rica, and CF for Central African Republic. The first genetic characterization of this virus isolate (an <i>env</i> V3 sequence designated 4056, GenBank accession number L11497, Murphy, E., et al., ARHR 9:997-1006 (1993)) left the subtype designation as unclassified, but a second study of this <i>env</i> region sequence classified it as subtype H (Janssens, W., ARHR 10:877-879 (1994)). This sequence was kindly made available prior to publication, and the manuscript is currently submitted (Gao, F., et al., (1997)). An additional <i>ltr</i> (U51290) and <i>env</i> and <i>nef</i> genes are available from this isolate (U08797 and U27401).			

Table 2 (cont.) Inter-subtype recombinants

Sequence	Accession	Origin	Reference
AC-ZAM184	U86780	Zambia	Salminen, M., et al., <i>J Virol</i> 71:2647-2655 (1997)
<p>The envelope gene from this sample taken from a Zambian woman in 1990 (Louwagie, J., et al., <i>J Virol</i> 69:263-271, 1995) proved to be an A/C recombinant in subsequent phylogenetic analyses (D. Robertson et al., <i>Nature</i> 374:124-126 (1995)). The full length provirus was recovered using PCR, and sequenced, according to the method described in Salminen, M., et al., <i>Virology</i> 213:80-86 (1995). Additional clones from serial samples from the index case ZAM184, and from her spouse who was also HIV-1 positive ((GenBank Accession # U86768-U86781, inclusive) represent gag, env, or additional full sequences from this couple. The A/C mosaic pattern of the full length genomic sequence is presented in Salminen, M., et al., <i>J Virol</i> 71:2647-2655 (1997), and also in Robertson, D., et al., part III pages 25-30, of this compendium, 1997. All coding sequences are intact. L22955.</p>			
AC-92RW009.6	U88823	Rwanda	Gao, F., et al., submitted, (1997)
<p>The isolate 92RW009 was part of the set generated through the WHO Global Programme on AIDS (see: WHO Network, ARHR 10:1327-1343 (1994)). The virus was derived from an asymptomatic 24 year old woman from Kigali, Rwanda, whose route of infection is thought to be heterosexual contact, and who had no anti-retroviral therapy. The clone has a frameshift mutation in gag at position 213. The blood sample was taken in 1992. The original env and gag sequence from this isolate clustered with HIV-1 subtype A sequences (Gao, F., et al., ARHR 10: 1359-1368 (1994)), however subsequent in-depth analysis of the full length genome sequence from this isolate suggests it is an AC mosaic sequence with multiple crossover points. The full length sequence and the information concerning its mosaic nature was kindly made available prior to publication, and the manuscript is currently submitted (Gao, F. et al., (1997)). The breakpoints are mapped in Robertson, D., et al., part III pages 25-30, of this compendium, 1997. The cytopathic effect of both the primary isolate from this patient, and from early passage cultured virus was NSI on MT-2 cells when analyzed by two different WHO labs in 1993-1994 (WHO Network, ARHR 10:1327-1343 (1994), and De Wolf, F., et al., ARHR 10: 1387-1400 (1994)). The NIAID 1997 Reference Reagent Catalog classifies it as NSI. However, more recent papers (Zhang L., et al., <i>Nature</i> 383:768 (1996) and Zhang L., et al., ARHR 13(16), 1357-1366 (1997), classify it as SI. This sequence is from PCR-amplified proviral DNA harvested from PBMCs from the NIH NIAID reagent repository. The NSI phenotyping was determined using earlier passages, and SI was associated with later passages, (J. Bradac, personal communication). Both the full length clone 92RW009.6 and the partial gp120 clone 92RW009.14 (accession U08793) were amplified from the same DNA extracted from a short term primary PBMC culture prepared by Ogden Bioservices, stored in the NIAID Research Reagent Repository, lot 1. Other env sequences from this isolate are: U16221, U08793, U13441, U16220, and U16222, and a gag sequence is U86545.</p>			
ADI-MAL	K03456	Zaire	Alizon, M. et al., <i>Cell</i> 46:63-74 (1986)
<p>This sequence is from a lambda phage clone derived from the Zairean isolate MAL. MAL was recovered in 1985 from a 7 year old boy with ARC, probably infected by a blood transfusion in 1981, as his parents were seronegative. All reading frames are intact except for vpu, due to the loss of a start codon. A reconstructed infectious clone is available. MAL was one of the first African sequences characterized, and soon after the initial characterization it was determined to be a mosaic (Li, W.-H., et al., <i>Genetics</i> 116:s44 (1987)). Recent analysis suggests that it has three different distinct subtype associations, A, D and I, and some regions that are difficult to characterize as associated with any known subtype (see Robertson, D., et al., part III pages 25-30, of this compendium, 1997). The I association is based on phylogenetic associations with a clone from Cyprus isolate that was the first I subtype characterized (94CY032), the 94CY032 clone appears to be an A-G-I mosaic (Robertson, D., et al., part III pages 25-30, of this compendium, 1997).</p>			

Table 2 (cont.)

Sequence	Accession	Origin	Reference
AE-93TH253.3	U51189	Thailand	Gao, F. et al., J Virol 70:7013-7029 (1996)
<p>This virus was isolated in 1993 from a 21 year old man from Chiang Mai, Thailand, who had end-stage AIDS. The isolate was previously designated CMU010, or 302053. The isolate was expanded in donor PBMCs, then in H9 cells, then a lambda phage clone was generated and sequenced. The sequence has a defective vpu due to the lack of a start codon, and a defective env gene. Like other "E" subtype viruses from both Asia and Africa, large stretches of the genome are associated with the A subtype, and all share a common mosaic pattern of A/E breakpoints, suggesting that the currently identified A-E recombinants all share a common ancestor. The breakpoints are mapped in Robertson, D., et al., part III pages 25-30, of this compendium, 1997.</p>			
AE-CM240	U54771	Thailand	Carr, J. et al., J Virol 70:5935-5943 (1996)
<p>Blood from an asymptomatic heterosexual 21-year-old Thai man was transported from Thailand to the USA where PBMCs were separated and co-cultivated with PHA-stimulated donor PBMCs. DNA from p24 antigen-positive culture was used to PCR amplify the proviral DNA. The complete genomic sequence of the provirus was determined by the compilation of three clones containing different parts of the viral genome. CM240 is an example of a Thai subtype E virus, which is a mosaic of a clade A virus and clade E virus, with the gag gene (and other regions) of subtype E viral genome falling within clade A in phylogenetic analysis. This is the pattern of A-E sequences found throughout Asia and Africa, and no full length E subtype reference strain has been identified (as for 93TH253). Carr et al. provide detailed analysis of the breakpoints, and point out that the A/E mosaic genomes have a natural pseudotype structure where the external envelope protein spikes on the virion essentially are contributed by the E subtype, and the rest of the viral proteins have a subtype A origin. The breakpoints are also mapped in Robertson, D., et al., part III pages 25-30, of this compendium, 1997. See also the env sequence from the same isolate (L14572), Mascola, J., et al., JID 169:48-54 (1993)</p>			
AE-90CF402.1	U51188	CAR	Gao, F. et al., J Virol 70:7013-7029 (1996)
<p>The isolate 90CF402 was sampled from a man from Bangui, the Central African Republic, who was suffering from AIDS-related conditions. His risk factor for infection was heterosexual contact. The isolate was first adapted to growth in chimpanzee cells, then re-expanded in human PBMC before a lambda phage clone was generated and sequenced. The sequence has a defective vpu due to the lack of a start codon, and a defective vif gene. A reconstructed infectious molecular clone is available. The pattern of subtype A-E recombination breakpoints is shared between A-E subtype sequences from Thailand and from the Central African Republic, suggesting a shared ancestral recombined virus that arose prior to the subsequent epidemics in the two areas. The breakpoints are mapped in Robertson, D., et al., part III pages 25-30, of this compendium, 1997.</p>			
AGI-Z321	U76035	Zaire	Choi, D. et al., ARHR 13:357-361 (1997)
<p>Z321B is a later passage of isolate Z321 (see GenBank Accession Numbers M15896, U50208, and U50207); Z321 was isolated from a 1976 Zairean serum sample. Z321 was grown to industrial scale as a chronically infected T-cell line to manufacture an inactive, therapeutic HIV-1 immunogen. Z321B was then established from this industrial scale stock. The genomic sequence was derived from multiple PCR clones amplified from Z321B. Z321B contains a mutation in the termination codon of the tat gene, (bases 2294-2296) so that the tat gene of HIVZ321 extends further downstream (bases 2342-2344), and has a defective vpr and vpu. Recent analysis suggests that it has three different distinct subtype associations, A, G and I, and some regions that are difficult to characterize as part of any known subtype (see Robertson, D., et al., part III page 25-30 of this compendium, 1997). The I association is based on phylogenetic associations with a clone from Cyprus isolate that was the first I subtype characterized (94CY032). The 94CY032 clone also appears to be an A-G-I mosaic (Robertson, D., et al., part III pages 25-30, of this compendium, 1997).</p>			

Table 2 (cont.)

Sequence	Accession	Origin	Reference
AG-92NG083.2	U88826	Nigeria	Gao, F. et al., submitted (1997)
<p>This sequence is from a PCR clone from a primary culture from the NSI isolate 92NG083; the sample was taken in 1992 from an AIDS patient from Jos, Nigeria. The isolate was originally called JVI083, but was renamed 92NG083 to be consistent with WHO nomenclature. While originally described as subtype G in env (Abimiku, A., et al., ARHR 10:1581-1583 (1994)), analysis of the full length genome revealed mosaic A/G regions. The breakpoints are mapped in Robertson, D., et al., part III pages 25-30, of this compendium, 1997. The full length clone has an altered initiation codon at position 157 and an inframe stop codon at position 360 in gag, and a vpu frameshift mutation at position 5462. This sequence was kindly made available prior to publication, and the manuscript is currently submitted (Gao F., et al., submitted (1997)).</p>			
AG-92NG003.1	U88825	Nigeria	Gao, F. et al., submitted (1997)
<p>This sequence is from a PCR clone from an NSI primary culture from isolate G3, renamed 92NG003 to be consistent with WHO nomenclature. The sample was taken in 1992 from a 27 year old, asymptomatic HIV seropositive female prostitute from Jos, Nigeria. (Abimiku, A., ARHR 10:1581-1583 (1994), env sequence accession number U13208). The isolate came from the Institute of Human Virology, Baltimore, MD. This sequence was kindly made available prior to publication, and the manuscript is currently submitted (Gao F., et al., submitted (1997)). While originally described as subtype G in env, this genome has mosaic A/G regions. The breakpoints are mapped in Robertson, D., et al., part III pages 25-30, of this compendium, 1997. There are frameshift mutations associated with 10-16 bp deletions in vpr and vpu, at positions 5024 and 5485, as well as deletions totaling 33 bp near the 3' end of the V3 loop. Nef has an altered initiation codon at position 8113.</p>			
AG-IbNg	L39106	Nigeria	Howard, T. et al., ARHR 12:1413-1425 (1996)
<p>HIV-1 sample IbNg was isolated from the PBMCs of an asymptomatic 23 year old man from Nigeria. DNA from this isolate was PCR amplified and cloned, with the complete genome sequence derived from multiple PCR amplification products. The partial env gene sequence (U48628) was originally designated subtype A (Howard, T. et al., ARHR 10:1755-1757 (1994)), as was the full length genome. The full length sequence was eventually shown to be an A-G recombinant with multiple cross-over points. The breakpoints are mapped in Robertson, D., et al., part III pages 25-30, of this compendium, 1997. The IbNg sequence has a 16 bp insertion within the Lys-tRNA primer binding site, just 3' of the 5' LTR. It also has a single nucleotide deletion in tat cds at position 5449.</p>			
BF-93BR029.4	AF005495	Brazil	Gao, F. et al., submitted, (1997)
<p>The isolate 93BR029 is part of a set of isolates obtained through the WHO Global Programme on AIDS (WHO Network, ARHR 10:1327-1344 (1994)), and came from a asymptomatic HIV seropositive 17 year old male, with unknown risk factor, from Sao Paulo, Brazil, sampled in 1993. The isolate had an NSI phenotype by an MT-2 assay. The isolate 93BR029 was established and propagated by short term cocultivation with normal donor lymphocytes, and then the near full length genome was PCR amplified and sequenced. An envelope gene sequence from this isolate is described in Gao, F., et al., J Virol 70, 1651-1657 (1996), U27413, and an LTR sequence is also available U51291. The envelope gene was first described as an F subtype, however subsequent phylogenetic analysis of the full length genome indicated that the clone is a mosaic with regions of B and F subtype. The breakpoints are mapped in Robertson, D., et al., part III pages 25-30, of this compendium, 1997. There are two frame shift mutations in gag, in positions 269 and 472. This sequence was kindly made available prior to publication, and the manuscript is currently submitted (Gao, F. et al., (1997)).</p>			

Table 2 (cont.)

Sequence	Accession	Origin	Reference
O-MVP5180	L20571	Cameroon	Gurtler, L. et al., J Virol 68:1581-1585 (1994)
The isolate MVP5180 was derived from a Cameroonian woman who had AIDS in 1991; she died of AIDS in 1992. The isolate from which the clone was derived was grown in several T cell lines, and could also grow in the monocytic U937 cells.			
O-ANT70	L20587	Cameroon	Vanden Haesevelde, M. et al., J Virol 68:1586-1596 (1994)
ANT70 was isolated from the first O group infection discovered, and the very divergent LTR sequence was first published in 1990 (de Leys, R., et al., J Virol 64:1207-1216, 1990). The isolate came from a CDC stage III infected individual with unusual serological reactivity. O group viruses have the same genetic organization as M group viruses, which dominate the epidemic, but are quite distinct in terms of their genetic sequences. For a review, See Korber B., et al., Human Retroviruses and AIDS Database, Part III, 41-56, 1996			
SIV-CPZANT	U42720	Zaire	Vanden Haesevelde, M. et al., Virology 221:346-350 (1996)
The CPZANT was isolated from a wild caught chimpanzee from Zaire. Two additional SIV CPZ (chimp) viral isolates have been generated from chimps caught in Gabon (U11495, X52154). The chimpanzee viral sequences are genetically more closely related to the HIV-1 sequences derived from infected humans than are HIV-2 strains or other SIVs.			
SIV-CPZGAB	X52154	Gabon	Huet, T. et al., Nature 345:356-359 (1990)
The CPZGAB virus was isolated from a chimpanzee caught in Gabon. The genome is more closely related to HIV-1 than are HIV-2 or other SIV viral sequences. Also see CPZGAB2, U11495 for a sequence fragment from an additional chimpanzee caught in Gabon.			

Table 3 Table of HIV-2/SIV full length genomes.

No.	Name	Accession	Origin	Reference
1	A-BEN	M30502	Germany	Kirchhoff, <i>Virology</i> 177:305 (1990)
2	A-GH1	M30895	Ghana	Hasegawa, <i>ARHR</i> 5:593 (1989)
3	A-D194	J04542	Ghana	Kuehnel, <i>PNAS USA</i> 86:2383 (1989)
4	A-UC2	U38293		Barnett, (Unpublished)
5	A-ROD	M15390	Senegal	Guyader, <i>Nature</i> 326:662 (1987)
6	A-NHZ	J03654		Zagury, <i>PNAS USA</i> 85:5941 (1988)
7	A-ISY	J04498	Gambia	Franchini, <i>PNAS USA</i> 86:2433 (1989)
8	A-ST	M31113	Senegal	Kumar, <i>J. Virol.</i> 64:890 (1990)
9	A-CAM2	D00835	Guinea-Bissau	Tristem, <i>J. Gen. Virol.</i> 72:721 (1991)
10	A-MDS	Z48731	Guinea-Bissau	Becker, (Unpublished)
11	A-KR	U22047		Kraus, (Unpublished)
12	B-UC1	L07625	Ivory Coast	Barnett, <i>J. Virol.</i> 67:1006 (1993)
13	B-D205	X61240	Ghana	Dietrich, <i>Nature</i> 342:948 (1989)
14	B-EHOA	U27200	Ivory Coast	Galabru, <i>ARHR</i> 11:873 (1995)
15	SD-MM251	M19499	macaque	Hirsch, <i>Cell</i> 49:307 (1987)
16	SD-MM32H	D01065	macaque	Rud, <i>J. Gen. Virol.</i> 75:529 (1994)
17	SD-MM1A11AA	M76764	macaque	Planellies, <i>ARHR</i> 7:889 (1991)
18	SD-MM142	M16403	macaque	Chakrabarti, <i>Nature</i> 328:543 (1987)
19	SD-MM239	M33262	macaque	Kestler, <i>Science</i> 248:1109 (1990)
20	SD-MNE	M32741	macaque	Benveniste, (Unpublished)
21	SD-SMMPBJ14.441	M31325	sooty mangabey	Dewhurst, <i>Nature</i> 345:636 (1990)
22	SD-SMMPBJ14.15	L03295	sooty mangabey	Dewhurst, <i>Nature</i> 345:636 (1990)
23	SD-SMMPBJ6.6	L09212	sooty mangabey	Hirsch, <i>J. Virol.</i> 67:2466 (1993)
24	SD-SMMH4	X14307	sooty mangabey	Hirsch, (Unpublished)
25	SD-SMMH9	M80194	sooty mangabey	Courgnaud, <i>J. Virol.</i> 66:414 (1992)
26	STM-STM	M83293	stump-tailed macaque	Novembre, <i>Virology</i> 186:783 (1992)

[illegible]

HIV1 Complete Genomes

B-WEAU	CCAGTGGAAACCAAGAGAGGTAGAAGA...AGCTATGGAAGGAGAGAACACACAGCTTGTACACCTATGAGCCTGCATGGGATGGATGACCATGAGAAAGAAGTATTATGTGTGGAACTTTGACAGCAAACTAGCATTTTCATCACGTAGC	295
B-WEAU	P_V_E_P_E_E_G_R_R_S_S_S_R_R_E_O_O_L_V_T_P_V_R_P_A_M_D_G_S_P_S_R_R_S_I_N_V_R_V_S_O_O_T_S_I_S_S_A_S	92
B-WEAU	O_W_N_O_R_V_E_I_I_A_N_R_Q_R_M_N_S_L_L_R_P_M_S_L_R_Q_M_D_D_H_E_R_R_V_L_M_W_K_F_D_S_K_L_A_P_H_H_V_A	96
B-WEAU	T_S_G_T_P_R_R_S_R_F_I_L_M_K_E_R_T_T_A_C_V_T_L_S_A_C_M_G_W_M_T_M_R_K_K_Y_S_C_G_S_L_T_A_M_S_H_F_I_T_S	90
A-0455		0
A-92UG037.1		0
A-Q23		0
B-MN	---T-G---A---G-C-A---T---A---CG---G---G---G---C---C-C---G---TTAT---	294
B-AD6	---T-G---C-A---G-C-A---T---A---ACG---G---G---GCA---C---CCC---A-G---	295
B-WR27		0
B-RF		0
B-B96	---T-G---TG-A-G---G.....CA-CA---G---C---GCT-A--A---AG-A-A---TCS---G-C---G---G---G---CCC---C---G---	289
B-BC593		0
B-JRFL		0
B-JRCSP	---T-T---G-C---T---A---A---C---CA---G---G---G---G---G---G---G---G---	296
B-OYT		0
B-CAM1	---T-G---G-C---A-A---GC---A---A---C---CA---C---G---G---G---G---G---G---G---G---	296
B-MYS	---T-G---G-C---A-A---GC---A---A---C---CA---G---G---G---G---G---G---G---G---	295
B-pNL43	---T-G---C-A---G-C-A---G---A---G---G---G---G---G---G---G---G---G---G---G---	295
B-EXB2	---T-G---T---C-CA---C---G---A---A---C---CG---G---G---G---G---G---G---G---	295
B-LA1		0
B-YU2	---T-G---R---G-C---C---T---A---A---C---CG---G---G---G---G---G---G---G---	292
B-YU10	---T-G---A---G-C---C---T---A---A---C---CG---G---G---G---G---G---G---G---	295
B_ACH320.2A.2.1	---T-G---A---G-C---C---T---A---A---C---CG---G---G---G---G---G---G---G---	295
B_ACH320.2A.1.2	---T-G---A---G-C---C---T---A---A---C---CG---G---G---G---G---G---G---G---	295
B-SF2	---T-G---G-C---G---G---G---G---G---G---G---G---G---G---G---G---G---G---	296
B-D31		0
B-MANC	---T-G---C---G-C---A---C---CA---G---G---G---G---G---G---G---G---	240
B-HAN		0
B-C18MPC	...--T---A---A---G-C---A---A---A---A---T---CTCA-C---GGG-C-TTCC---GCGCGGOC---AGT-AC-A-G.....C-	137
C-ETH2220		0
C-92BR025.6		0
D-842R085		0
D-NDK		0
D-22Z6	---T-T---C-G-G---G-C-C---C---T---G-AT---A---A---G---ACG---G---G---A---GA---A---G---A---AAG---	295
D-EL3		0
D-94UG114.1		0
F-93BR020.1		0
H-90CR056.1		0
AC-2AM184		0
AC-92RN009		0
ADI-MAL		0
AE-93TH253	---T-C---A.GAG-A---G...G-AC-CA---A---T---C---C---A---A---G---G---A---A---G---GC-G---A---TACC---CGAAGA---A---	295
AE-CM240		0
AE-90CF402	---T-T---A.GAG-A---G...G-AC-CA---A---T---C---C---A---A---G---G---A---A---G---GC-G---A---TACC---CGAAGA---A---	295
AGI-Z321		0
AG-92NG081.2		0
AG-92NG003.1		0
AG-IBNG		0
BF-93BR029.4		0
O-WVP5180	---TC-G---AG---C---GAGACTG-G---AC-ANT---AG-TGCT---TC-TC---T---AGCTT-TAAT---AGCT-G---TGCAC-CGGG---GA-C---AA---C---T---ATC-T---GC---AACA---TA---	299
O-ANP70	---TC-GA---AG---C---AGACT-GAG-AC-T-T---GGCT-ATC-CC-G---T---AGCAT-TGCC---AT-T---TACAC-T---A-C-G---T---ATCT---GCACACCC---T---T---	299
SIVCP2AMT		0
SIVCP20AB	--CC--AC-QAG---C---C---C---T---T---C---G---T---C---AT-T-A---A---TG-G-C---G-C---GG-C---CGC---GG-G-C---AQA---TA-T---	295

HIV1 Complete Genomes

	TCF-1alpha	NF-kappa-B-II	NF-kappa-B-I	Spi-III	Spi-II	
B-WEAU	CCGAGAGCTGCATCCGAGTACTTTAAAGACTGCT.	GACATCGA.GTTTCT.	ACAAAGGACTTTCCAC.	TGGGGACTTTCCGG.GGAGGCGTGATC.	TGGGGCGGG	395
B-WEAU	P_R_A_A_S_G_V_L_S_R_L_I	H_R_F_S	R_G_L_S	G_D_F_P	E_A_S	Q_G
B-WEAU	R_E_L_H_P_E_Y_F_K_D_C_F	D_I_I_F	T_R_D_F_P	B_G_T_F_R	R_R_D	N_A_G
B-WEAU	P_E_S_C_I_R_S_T_L_K_T_A	T_B_V_F	Q_G_T_F_H	M_G_L_S	G_G_V_I	G_R
A-U855						0
A-92UG037.1						0
A-Q33						0
B-NM	AC-GA	C-A	G	A	T	QC
B-AD6	AC		Q	A		QC
B-WR27						0
B-RF						0
B-896	C-GA	AACTGCT	CT	C-A	G	
B-BC9G3						0
B-JRFL						0
B-JRCSF	AC-G	C	C			QC
B-OYT						0
B-CAM1	AA	T	AC			QC
B-NY5	AC		C			QC
B-PML43	AC		C			QC
B-HXB2	C-GA	C	Q			QC
B-LA1						0
B-Y02	AC-GA	A	C			QC
B-YU10	AC-GA	A	C			QC
B-ACH320.2A.2.1	AC-G	G	C			QC
B-ACH320.2A.1.2	AC-G	G	C			QC
B-SF2	AC		C			QC
B-D31						0
B-HAMC	AC		G			QC
B-HAN		G	C			QC
B-C18MBC						0
C-ETK2220						0
C-92BR025.8						0
D-842R085						0
D-NDK						0
D-Z226	T	AC	C	A		QC
D-ELI						0
D-94UG114.1						0
F-93BR020.1						0
H-90CR056.1						0
AC-ZAM184						0
AC-92RW009						0
AD1-MAL						0
AE-93TH253	A	G	A	T	A	QC
AE-CN240						0
AE-90CF402	T	C	A	AC	A	QC
AG1-2321						0
AG-92NG083.2						0
AG-92NG003.1						0
AG-IBNG						0
BF-93BR029.4						0
O-MVP5180	TGC-AAA	C	A	CT	CCCCA-G-AACT	CTGC
O-AM770	TATGATAACT	C	A	CT	CC-GA-GGA-AAA	AACTGCT
SIVCPZANT						0
SIVCPZGNB	A	A	AA	AC	ACTTTAGAC	TGGCCCATCCGC
B-WEAU	P_R_A_A_S_G_V_L_S_R_L_I	H_R_F_S	R_G_L_S	G_D_F_P	E_A_S	Q_G
B-WEAU	R_E_L_H_P_E_Y_F_K_D_C_F	D_I_I_F	T_R_D_F_P	B_G_T_F_R	R_R_D	N_A_G
B-WEAU	P_E_S_C_I_R_S_T_L_K_T_A	T_B_V_F	Q_G_T_F_H	M_G_L_S	G_G_V_I	G_R

HIV1 Complete Genomes

	Sp1-I	TATA box	5' LTR U3 end	5' LTR R repeat begin	1-> +1 mRNA start site	Poly-A signal	
B-NEAU	ACAGGGGAGTGGGGA	GCCTCAGATGCTGCATATAGCAGCTGCTTTTGGCTG	TACTGGCTCTCTCTGCTT	AGACCAATCTGAGCCCTGGAGCTCTCTGCTAGCTAGGGAAACCCACTGCTT	AAGCTCAATAAAGCTTGGCTTGA	543	
					 TAR element stem	 loop	
B-NEAU	T G E V R I P O M L K I S S C F L P V L G L S G S T R S E P O S S L A S S G T H C I K P Q S E L F S						162
B-NEAU	Q G S G I A L R C C I S A A A F C L Y W V S L V R P D L S I G A L W L A R R E P T A I S L N K A C L E						172
B-NEAU	D R G V A I P S D A A Y R O L I F A C T Q S I W I D O I S A W E L S G S L G N P L I I A S I X I A I						167
A-U455							13
A-92UG037.1							0
A-Q23							13
B-MW	--C-----,-----CG-----C-----A-----						542
B-AD8	--T-----,-----A-----						543
B-WR27							0
B-RF					-A-CT--		57
B-896	--T-----,-----						542
B-SC8G3							90
B-JRFL							0
B-JRCSE	--T-----,-----						544
B-OVI					-A-C-----C-----A-----		89
B-CAN1	--T-----,-----				GA-----		545
B-MY5	--T-----,-----						543
B-pNL43	--T-----,-----A-----				A-----		543
B-MXB2	--T-----,-----C-----				A-----		543
B-LA1					T-----A-----		89
B-YU2	--T-----,-----				A-----A-----		539
B-YU10	--T-----,-----				A-----		542
B-ACH320.2A.2.1	--T-----,-----				A-----GA-----		544
B-ACH320.2A.1.2	--T-----,-----				A-----GA-----		544
B-SF2	--T-----T-----				A-----A-----		543
B-D31					G-----		89
B-MANC	--C-----,-----				A-----		489
B-HAN							13
B-C18MBC	--T-----,-----C-----						347
C-ETK2220							0
C-92BR025.8							0
D-942R085					-A-CT--CG-----AT-----		57
D-MDK	--T-----T-----A-----				T-----A-----		87
D-2226	--T-----T-----A-----				T-----A-----		543
D-BLI					T-----		89
D-94UG114.1							0
F-93BR020.1							0
H-90CR056.1							0
AC-ZAM184							0
AC-92RW009							0
ADI-MAL					-T-----G-----C-----A-----		87
AB-93TH253	GTT-----TT-----A-----A-----C-----C-----T-----T-----G-----G-----C-----AG-----A-----						543
AB-CH240					-T-----G-----C-----A-----		90
AE-90CF402	GTT-----T-----A-----A-----C-----C-----T-----T-----G-----C-----AG-----A-----						566
AGI-2321							0
AG-92NG003.2							0
AG-92NG003.1							0
AG-IBNG						GG-----	72
BF-93BR029.4							0
O-MVP5180	TTC-----T-----A-----A-----CC-----T-----C-----TAG-TAGAG-----G-----C-----C-----CTCTAGCT-----G-----CG-----						551
O-ANT70	TTC-----T-----A-----A-----C-----C-----T-----C-----GG-TAGAG-----G-----C-----C-----CTCTAGCT-----G-----CG-----						576
SIVCPZANT							0
SIVCP2GAB	GTT-----T-----TTT-----C-----T-----CAC-C-----T-----A-----C-----TG-A-----						563

[illegible]

HIV1 Complete Genomes

--> Gag and Gag-Pol CDS start (frame 2)			
B-WEAU	GAGAGAGATGGGTGCGAGAGCCTCGGTATTAGCGGGGGGAGAAATTAGATAGGTGGGAAAAATTCGGTTAAGGCCA...GGGGGAATATAAAATATAAATAAATAAATAGTATGGGCAGCAG	904	
B-WEAU	P_R_D_G_C_E_S_V_Q_I_K_R_Q_R_Y_R_S_V_G_K_N_S_V_K_A_S_I_G_K_S_K_I_S_T_K_T_V_S_M_G_K_Q	267	
B-WEAU	P_R_D_G_C_E_S_V_Q_I_K_R_Q_R_Y_R_S_V_G_K_N_S_V_K_A_S_I_G_K_S_K_I_S_T_K_T_V_S_M_G_K_Q	273	
B-WEAU	P_R_D_G_C_E_S_V_Q_I_K_R_Q_R_Y_R_S_V_G_K_N_S_V_K_A_S_I_G_K_S_K_I_S_T_K_T_V_S_M_G_K_Q	267	
A-U455	A.....A.....AA-A.....TCA-G.....G.....C.....G.....G.....T.....	372	
A-92UG037.1	A.....T.....GCA-G.....G.....G.....G.....C.....C.....	275	
A-Q23	A.....A.....C.....OCA-G.....G.....G.....G.....G.....A-G-G-C-A.....	369	
B-MN	C-A.....C-A.....C.....G.....G.....G.....G.....G.....T.....	902	
B-A06	A.....C-A.....A.....A.....G.....G.....G.....G.....G.....T.....	905	
B-WR27	AA-G.....A.....AA.....C.....G.....G.....G.....G.....G.....T.....	284	
B-KF	A.....C-A.....C-AA.....A.....G.....G.....G.....G.....G.....T.....	419	
B-896	A.....A.....G.....A.....G.....G.....G.....G.....G.....T.....	904	
B-BCS03	A.....G.....A-G-A.....G.....A.....G.....G.....G.....T.....	452	
B-JRFL	A.....A.....G.....AA.....G.....G.....G.....G.....G.....T.....	228	
B-JRCSF	A.....G.....G.....A.....A.....G.....G.....G.....G.....G.....T.....	905	
B-OYT	A.....AA.....A.....G.....C.....C.....C.....C.....C.....	451	
B-CAM1	A.....AA.....AA.....A.....G.....G.....G.....G.....G.....T.....	907	
B-WY5	A.....AA.....AA.....A.....G.....G.....G.....G.....G.....T.....	905	
B-pNL43	A.....AA.....AA.....A.....G.....G.....G.....G.....G.....T.....	905	
B-KXB2	A.....C-A.....C-A.....C.....G.....G.....G.....G.....G.....T.....	905	
B-LAI	A.....C-A.....C-A.....C.....G.....G.....G.....G.....G.....T.....	451	
B-YU2	A.....T-C-G.....A.....A.....G.....G.....G.....G.....G.....T.....	903	
B-YU10	A.....T-C-G.....A.....A.....G.....G.....G.....G.....G.....T.....	906	
B-ACH320.2A.2.1	A.....T-C-G.....AA-G.....G.....G.....G.....G.....T.....	907	
B-ACH320.2A.1.2	A.....T-C-G.....A.....A.....G.....G.....G.....G.....G.....T.....	905	
B-572	A.....AA.....AA.....A.....G.....G.....G.....G.....G.....T.....	905	
B-D31	A.....A.....A.....A.....G.....G.....G.....G.....G.....T.....	451	
B-MAMC	A.....A.....A.....A.....G.....G.....G.....G.....G.....T.....	851	
B-BAN	A.....A.....C-AA.....A.....G.....G.....G.....G.....G.....T.....	374	
B-C16MBC	AA.....A.....A.....A.....G.....G.....G.....G.....G.....T.....	709	
C-BTH2220	AA.....A-C-A-A.....GCC.....AA.....A.....G.....G.....G.....G.....G.....T.....	285	
C-92BR025.8	AA.....A-C-A-A.....GCC.....AA.....A.....G.....G.....G.....G.....G.....T.....	256	
D-842R085	A.....A.....A.....G.....GCT.....C.....G.....G.....G.....G.....G.....T.....	422	
D-NDK	A.....A.....A.....G.....CA.....G.....C.....A.....B.....G.....G.....G.....G.....G.....T.....	449	
D-2226	A.....A.....A.....G.....GCA.....C.....C.....A.....B.....G.....G.....G.....G.....G.....T.....	905	
D-ELI	A.....A.....A.....AA.....C.....C.....A.....A.....G.....G.....G.....G.....G.....T.....	451	
D-94UG114.1	A.....A.....A.....GAA.....C.....C.....A.....A.....G.....G.....G.....G.....G.....T.....	254	
F-93BR020.1	A.....A.....A.....GCT.....G.....G.....G.....G.....G.....G.....G.....G.....G.....T.....	248	
H-90CR056.1	A.....C.....A.....GCT.....E.....C.....G.....G.....G.....G.....G.....T.....	252	
AC-2AM194	A.....A.....A.....GCT.....G.....G.....G.....G.....G.....G.....G.....G.....G.....T.....	284	
AC-92RN009	AA.....A-C-A.....GCC.....AA.....A.....G.....G.....G.....G.....G.....T.....	255	
frameshift			
ADI-MAL	A.....A.....GCA.....G.....G.....G.....G.....G.....G.....G.....G.....G.....T.....	445	
AE-937H253	A.....A.....T.....A.....GCA.....G.....C.....G.....G.....G.....G.....G.....T.....	922	
AE-CM240	A.....T.....A.....GCA.....G.....G.....G.....G.....G.....G.....G.....G.....G.....T.....	470	
AE-90CF402	AA.....T.....A.....GCA.....G.....C.....G.....G.....G.....G.....G.....T.....	947	
AGI-2321	A.....A.....A.....AA.....G.....G.....G.....G.....G.....G.....G.....G.....G.....T.....	116	
AG-92MG003.2	G.....A.....A.....TCT.....G.....G.....G.....G.....G.....G.....G.....G.....G.....T.....	271	
AG-92MG003.1	A.....A.....A.....GCA.....G.....G.....G.....G.....G.....G.....G.....G.....G.....T.....	272	
AG-IBW3	A.....A.....T.....A.....GCA.....G.....G.....G.....G.....G.....G.....G.....G.....G.....T.....	445	
BF-93BR029.4	A.....A.....A.....AA.....A.....A.....A.....A.....A.....A.....A.....A.....T.....	257	
O-MVP5180	AGGG-GA.....T-G-G-CA-A-PA-G-GCA-OG-A-A.....A.....ATCT-A-GGC-OG-T.....	932	
O-AMT70	AGGG-GA.....T-G-G-CA-AA-CA-G-GCA-C-A-A.....A.....ATCT-A-G-G-CT-T.....	957	
SIVCP2ANP	GGGGAAGGTGGAAGTCTCTAGGAAC-GA-A-G-G-T-T-G-G-A-AGA-SC-CA-GT-CA-C-TC-C-T-C-G-G-C-TGA-C-G-T-ATC	269	
SIVCP2GAB	...GAAGGGGGAAGTCTCTAGGGG-A-TC-CA-A-C-C-G-C-T-C-C-GA-G-TGA-G-T-T.....	942	
Gag and Gag-Pol p17 Matrix start M_G_A_R_A_S_V_L_S_Q_G_E_L_D_R_W_E_R_I_R_L_R_F...G_G_N_K_K_Y_K_L_K_H_I_V_W_A_S_R Gag, Gag-Pol			

1082
 A V L Y C V GAG, GAG-Pol

**I-25
DEC 97**

HIV1 Complete Genomes

B-WEAU	GCATCAAAAGNTAGAGGTAAAGATACCAAGGAAGCTTTAGATAAGATAGAG.....GAAGGCAAAACAA...AGCAAG...AAAAGGSCACAGCAA...GCAGCA...GCT.....GA	1150
B-WEAU	A S K D R G K R Y O G S P R S D R I R A K O I R O T A S I	1150
B-WEAU	M Q X I E V K D T R E A L D K I E E E Q N X S K K K A O Q A A A D	1150
B-WEAU	C I K R S R S K I P R K L S I R S I I K S K T A A K R H B I Q O I I I	1150
A-U455	A-----G-----T-----C-----A-----A-----A-----AT-----G-----A-----C-----G-----A-----G-----A-----	618
A-92UG037.1	A-----CG-----C-----C-----A-----A-----GATA-----G-----C-----A-----A-----G-----A-----	521
A-Q23	A-----CG-----T-----C-----A-----A-----ATAA-----T-----G-----C-----A-----A-----G-----A-----	615
B-MN	A-----A-----C-----C-----G-----A-----A-----T-----C-----A-----A-----G-----GACACA-----GGAA-----	1157
B-AD8	A-----C-----T-----G-----C-----C-----A-----A-----A-----C-----A-----A-----G-----A-----	1151
B-WR27	A-----G-----G-----C-----C-----G-----A-----A-----T-----A-----A-----G-----A-----	530
B-RF	A-----T-----G-----C-----C-----C-----A-----A-----T-----A-----A-----G-----A-----	665
B-B96	A-----G-----C-----C-----C-----A-----A-----T-----A-----A-----G-----A-----	1150
B-BC9G3	A-----T-----T-----C-----C-----G-----A-----A-----T-----A-----A-----G-----A-----	498
B-JRFL	A-----G-----A-----C-----C-----A-----A-----GC-----T-----A-----A-----A-----G-----A-----	468
B-JRCSP	A-----G-----A-----C-----C-----A-----A-----C-----T-----T-----A-----A-----G-----A-----	1151
B-OYI	A-----G-----C-----C-----C-----A-----A-----T-----A-----A-----G-----A-----	697
B-CAM1	A-----T-----C-----C-----G-----A-----A-----T-----A-----A-----G-----A-----	1153
B-MY5	A-----G-----T-----C-----C-----C-----A-----A-----T-----A-----A-----G-----A-----	1151
B-pNL43	A-----G-----T-----C-----C-----C-----A-----A-----T-----A-----A-----G-----A-----	1151
B-HXB2	A-----G-----A-----C-----C-----C-----A-----A-----T-----A-----A-----G-----A-----	1153
B-LAI	A-----G-----A-----C-----C-----C-----A-----A-----T-----A-----A-----G-----A-----	597
B-YU2	A-----G-----C-----C-----G-----A-----A-----T-----A-----A-----G-----A-----	1149
B-YU10	A-----G-----C-----C-----G-----A-----A-----T-----A-----A-----G-----A-----	1152
B-ACH320.2A.2.1A	A-----G-----C-----C-----G-----A-----A-----T-----A-----A-----G-----A-----	1153
B-ACH320.2A.1.2A	A-----G-----C-----C-----G-----A-----A-----T-----A-----A-----G-----A-----	1151
B-SF2	A-----G-----T-----C-----G-----A-----A-----T-----A-----A-----G-----A-----	1158
B-D31	A-----G-----C-----C-----C-----A-----A-----T-----A-----A-----G-----A-----	697
B-NAMC	A-----G-----C-----C-----C-----A-----A-----T-----A-----A-----G-----A-----	1097
B-NAN	A-----G-----C-----C-----G-----A-----A-----T-----A-----A-----G-----A-----	620
B-C19HBC	A-----C-----CA-----G-----C-----A-----A-----T-----A-----A-----G-----A-----	970
C-87H2220	A-----A-----A-----C-----C-----C-----A-----A-----G-----A-----A-----G-----A-----	534
C-92BR025.8	A-----A-----T-----CG-----C-----C-----A-----A-----A-----A-----A-----G-----A-----	505
D-84ZR085	A-----A-----G-----A-----C-----C-----A-----A-----A-----A-----A-----G-----A-----	668
D-NDK	A-----G-----G-----C-----A-----G-----A-----G-----A-----A-----A-----A-----G-----A-----	695
D-2256	A-----G-----G-----C-----A-----G-----A-----G-----A-----A-----A-----A-----G-----A-----	1154
D-ELI	A-----A-----GGA-----T-----C-----C-----A-----A-----G-----A-----A-----A-----A-----G-----A-----	597
D-94UG114.1	A-----G-----G-----A-----GC-----AGC-----C-----A-----G-----A-----G-----A-----	500
P-93BR020.1	A-----G-----G-----C-----C-----G-----G-----C-----A-----A-----G-----A-----G-----A-----	494
H-90CR056.2	A-----G-----A-----T-----G-----C-----G-----A-----A-----A-----A-----A-----G-----A-----	498
AC-2AM184	A-----G-----G-----C-----C-----C-----A-----A-----A-----A-----A-----G-----A-----	530
AC-92RW069	A-----G-----T-----C-----G-----C-----C-----A-----A-----A-----A-----A-----G-----A-----	504
ADI-MAL	A-----G-----T-----C-----GC-----A-----A-----A-----A-----A-----G-----A-----	729
AE-93TH253	A-----C-----G-----C-----C-----A-----A-----A-----A-----A-----G-----A-----	1168
AE-CN240	A-----C-----G-----C-----C-----A-----A-----A-----A-----A-----G-----A-----	716
AE-90CF402	A-----G-----T-----C-----C-----G-----A-----A-----A-----A-----A-----G-----A-----	1193
AGI-2321	A-----T-----C-----C-----A-----C-----G-----AG-----G-----A-----A-----G-----A-----A-----G-----A-----	365
AG-92NGD83.2	A-----A-----C-----A-----CC-----GG-----AG-----G-----A-----A-----A-----A-----A-----G-----A-----	517
AG-92NGD03.1	A-----G-----C-----A-----C-----GG-----AG-----G-----A-----A-----A-----A-----A-----G-----A-----	518
AG-IBNG	A-----G-----CA-----C-----C-----A-----G-----A-----A-----A-----A-----A-----G-----A-----	691
BF-93BR029.4	A-----T-----C-----C-----A-----A-----A-----A-----A-----A-----A-----G-----A-----	503
O-MVP5180	T-CA-C-GAT-T-CA-CCG-----AC-C-G-AA-C-A-T-A-----TA-----TGGC-----G-----GTC-----AG-C-----AAG-----	1172
O-ANT70	T-CA-C-GATATA-AA-TGG-----GC-C-G-AA-C-A-T-A-----TA-----TGGG-----G-----GTC-----AG-C-----AAG-----	1200
SIVCPZAMT	A--A--G-GA-A--A--C--AG-AC--CG-----T-----GTGA-AA-T-----GTAATG-----G--A--CA-----AGAACAGAGTAACCAACCGCAGCAGA-G	524
SIVCPZGAB	A--AGTG-C-CACT--G--C-AC-A--C--AC-C-A--CGCATCAT-G--A--C-G-CCAA-AACT-----GTAA-T-----G--A--C--TGAAGGGGAG...CACTCAAGGGCTAGTGGC	1222
p17 Matrix	H Q K I E V K D T K E A L D K I E E E Q N X S K K K A O Q A A A D	Gag-Pol

HIV1 Complete Genomes

B-MEAU	CACAGAA.....	..ACAGAAACCAAGGTCAGCCAAATTAACCTATAGTCCAGAACCTACAGGGACAAATGGTACATCAGGCGCTATCACTAGAACTTTAAATGCATGGGTAAAGT	1261
B-MEAU	H_R_I.....	..I_Q_K_F_G_D_P_K_L_P_V_S_A_E_P_T_G_T_U_O_T_S_G_P_I_T_S_N_F_K_C_M_G_K_S	373
B-MEAU	T_S.....	..H_R_N_Q_V_S_Q_M_Y_P_I_V_Q_N_L_Q_G_Q_M_V_H_O_A_L_S_P_R_T_L_N_A_W_V_K_V	392
B-MEAU	T_O_S.....	..T_E_T_R_S_A_X_I_T_L_S_C_R_T_Y_R_D_R_W_V_I_R_P_V_H_L_E_L_S_M_R_G_S_X	368
A-U455	---G---	---G---	717
A-92UG037.1	---G---	---G---	632
A-Q23	---G---	---G---	726
B-MN	---G---	---G---	1268
B-AD8	---G---	---G---	1362
B-MR27	---G---	---G---	641
B-RF	---G---	---G---	776
B-896	---G---	---G---	1261
B-BC9G3	---G---	---G---	809
B-JRFL	---G---	---G---	579
B-JRCSP	---G---	---G---	1262
B-QYI	---G---	---G---	808
B-CAH1	---G---	---G---	1264
B-NY5	---G---	---G---	1262
B-PNL43	---G---	---G---	1262
B-HX82	---G---	---G---	1262
B-LAI	---G---	---G---	808
B-YU2	---G---	---G---	1260
B-YU10	---G---	---G---	1263
B-ACH320.2A.2.1	---G---	---G---	1264
B-ACH320.2A.1.2	---G---	---G---	1262
B-SF2	---G---	---G---	1269
B-D31	---G---	---G---	808
B-MANC	---G---	---G---	1208
B-HAN	---G---	---G---	731
B-C18MBC	---G---	---G---	1081
C-ETH2220	---G---	---G---	636
C-92BR025.8	---G---	---G---	607
D-848R085	---G---	---G---	779
D-MDK	---G---	---G---	797
D-Z226	---G---	---G---	1265
D-BLI	---G---	---G---	808
D-94UG114.1	---G---	---G---	511
F-93BR030.1	---G---	---G---	593
H-90CR056.1	---G---	---G---	609
AC-2AH184	---G---	---G---	535
AC-92RW009	---G---	---G---	506
ADI-MAL	---G---	---G---	840
AE-93TH253	---G---	---G---	1279
AE-CN240	---G---	---G---	827
AE-90CF102	---G---	---G---	1304
AGI-Z321	---G---	---G---	476
AG-92NG083.2	---G---	---G---	528
AG-92NG003.1	---G---	---G---	529
AG-IBNG	---G---	---G---	793
BF-93BR029.4	---G---	---G---	514
O-MVP5180	---G---	---G---	1283
O-ANT70	---G---	---G---	1308
SIVCPZANT	---G---	---G---	574
SIVCPZGAB	---G---	---G---	1323
p17 Matrix	T_E.....	..N_R_N_O_V_F_Q_M_Y_P_I_V_Q_N_L_Q_G_Q_M_V_H_O_A_L_S_P_R_T_L_N_A_W_V_K_V	Gag, Gag-Pol
		shift back	
		Gag and Gag-Pol p17 Matrix end / \ Gag and Gag-Pol p24 Capsid start	

B-WEAO	AATAGAAGAAAAGGCTTTCAGGCCAGAACTAATACCOCATGTGTTTCAGCATTATTCAGAAGGGGCCACCCCAAGATTTAAACACCATGCTAAACACAGTGGGGGACATCAAGGAGOCATGC AAAATGTTAAAGAGACCATCAATGAGGA	1411
B-WEAO	N R R K G F Q P R S N T H V F S Y I R R S K P T R P K H A K N S G G T S S S S H A N V K R D H O S G	422
B-WEAO	I E E K A F S P E V I F H F G A L S E G A T P O D L N T M L N T V G G H O A A M Q M L K E T I N E B	442
B-WEAO	S S A A L A L S A Q N S Y F C P O H Y D K S P F M K I S T P C S T Q N G D I K O P C A C S X R P S M R	412
A-U455	-G-----C-----G-----T-TG--G--TG--G-----C-G--T-----	867
A-92U0037.1	-----C-----C-----T-TG--G--T-----C-G--T-----T-----	782
A-Q23	-----C-----T-TG--G--T-----C-G--T-----T-----	876
B-HN	-G-----G-----	1418
B-ADB	-G-----G-----	1412
B-WR27	-G-G--G--T-----C-----T-----T-----T-----T-----	791
B-RF	-G--G--T-----T-----T-----T-----T-----	926
B-696	-G-----G-----G-----	1411
B-BCSG3	-G-----G-----C-----T-----	959
B-JREFL	-G-----G-----C-----C-----A-----	739
B-JRCBP	-G-----G-----C-----C-----A-----	1412
B-OVI	-G-----G-----G-----C-----	958
B-CAM1	-G-----G-----G-----	1414
B-NY5	-G-----G-----T-----	1412
B-PNL43	-G-----G-----T-----	1412
B-HXB2	-G-----G-----G-----	1412
B-LAI	-G-----G-----G-----	958
B-YU2	-G-G--G--G--G-----C-----	1410
B-YUL0	-G-G--G--G--G-----C-----	1413
B_ACH320.2A.2.1	-G-----G-----	1414
B_ACH320.2A.1.2	-G-G--G--G-----G-----	1412
B-SF2	-G-----G-----T-----	1419
B-D31	-G-----G-----	958
B-NANC	-G-----G-----T-----T-----T-----A-----	1358
B-HAN	-G-G--G--G-----A-----A-----	881
B-C18MB	-G-----G-----G-----T-----	1231
C-ETH2220	-G--G--G--A--T-----T-----C-----	786
C-92BR025.B	-G--G--G--G-----A-----T-----T-----C-----T-----	757
D-842R085	-----C-----A-----	929
D-MCK	-----C-----G-----T-----C-----C-----	947
D-Z226	-----T-----C-----G-----	1415
D-BLI	-----T-----C-----A-----	958
D-94UG114.1	-----G--G--A-----T-----T-----G--T-----	761
F-93BR020.1	G-----G-----T--T-----G-----T-----T-----T-----G--C--T-----	743
H-9DCR056.1	-G-----T-----C-----T-G--T-----T-----G-----T--A-----	759
AC-ZAM184	-G-----G-----C-----C-----T--T-----T-----T-----T-----T-----	785
AC-92RW009	-G--G--G--T--A--G-----A-----T-----T-----C-----T--A-----	756
ADI-NAL	-----G-----C-----G--G-----T-TG--G--T--T--A--C--G--T-----	990
AE-93TH253	-G--T--A--C-----G-----T-TG--T--T-----C--G--A-----	1429
AE-CH240	-O-----G--T--A--C-----G-----T-TG--T--T-----C--G--A-----	977
AE-90CF402	-G-----G--T--A--C-----T-TG--C-----T-----C--G--A-----T-----	1454
RGI-Z321	-G-G--AAC--T-----T-----T-----C-----G--T--T-----	626
AG-92MG083.2	-O-----C-----T-----G-----T-----T-----G-----C-----G--T--T-----T-----	778
AG-92MG003.1	-----AAC--T-----A-----G-----T-----C-----G-----T--T--T-----A-----	779
AG-IBNG	-G-----G-----G-----T-TG--T-----T-----C--G--A-----G-----T-----	943
BF-93BR029.4	-G-----G-----T-----T-----T-----T-----T-----T-----	764
O-MVP5180	-G-----G--C--T--A--T--A--T--T--T-----ATG-----G--TGT--CT--T--A--C--T-----G--TG--CA--A-----C--GQ--TT--A--G--G--G--AGTA-----	1433
O-ANT70	-G-----A--C--T--A--T--A--C--T--T-----CATG-----G-----G--A--TTT--CT--T--A--T--T-----TG--CA--A--A-----G--TT--A--G--C--G--AGTA-----	1459
SIVCPZANT	TG-----G--AAA--AT-----C--C--T-----T--T-----G--A--T--T--T-----T--TG--T-----AC-----G-----CG--C-----AGTA-----	824
SIVCPZGAB	GO-----A--T--T--G--V--A-----T-----T-----T-----T-----T-----G-----G-----G--AGT--T-----	1473
p24 Casaid	I E E K A F S P E V I F H F S A L S E G A T P O D L N T M L N T V G G H O A A M Q M L K E T I N E B	940, Gag-Pol

HIV1 Complete Genomes

Accession	Sequence	Position
B-HEAU	AGCTGCGAATGGGATAGATTACATCCAGTGCATGACAGGCGCTATTGACACCGCCAGATAGAGAACCCAGGCGAAGTACATAGCAGGACTACTAGTACGCTTCAGGAACAAATAGCATGATGACA	1552
B-HEAU	SCRMQSITPSACRAVCTRPDEETFRGNSENRYSPSGTNSHDD	464
B-HEAU	AARWDLNHPVHAGPIAPGGQNRPRGSDIAOTCTSTLOBQIANMT	489
B-HEAU	RLOMGIDYIQCNQQLLEQARSSENOGEVTSQELLVFPFAMKSMGS	454
A-U455	GCG	1008
A-92UG037.1	CGC	921
A-Q23	CGC	1017
B-MN	G	1559
B-AD8	G	1553
B-WR27	G	932
B-RF	G	1067
B-896	G	1853
B-BC9G3	G	1100
B-JRFL	G	870
B-JRCSF	G	1553
B-OYI	G	1099
B-CAM1	G	1555
B-NV5	G	1553
B-pNL43	G	1553
B-HXB2	G	1553
B-LAI	G	1099
B-YU2	G	1551
B-YU10	G	1554
B_ACH320.2A.2.1	G	1555
B_ACH320.2A.1.20	G	1553
B-SF2	G	1560
B-D31	G	1099
B-NANC	G	1499
B-HAN	G	1022
B-C19MBC	G	1372
C-ETH2220	G	927
C-92BR025.8	G	898
C-842R085	G	1070
D-NDK	G	1088
D-2226	G	1556
D-BLI	G	1099
D-94UG114.1	G	902
F-93BR020.1	G	684
H-90CR056.1	G	900
AC-3AM184	G	926
AC-92RN009	G	897
ADI-MAL	G	1131
AB-93TH253	G	1570
AB-CH240	G	1110
AE-90CF402	G	1595
AGI-2321	G	767
AG-92M3083.2	G	919
AG-92NG003.1	G	920
AG-IBMG	G	1084
BF-93BR029.4	G	905
O-MV5180	G	1577
O-ANT70	G	1602
SIVCPZANT	G	974
SIVCPZGAB	G	1614
g20 Capsid	G	

Accession	Sequence	Position
B-WEAU	CATCCAGTAGGAGAAATCTATAAAGATGAGTAACTAGGATTAAATAAATAGTAAGAAATGTATAGCCCTGTCTGACATTCTGGACRTAAGACAAAGGACCAAAAGAACCCCTTTAGAGATTATGTAGACCCGTTCTATAAAACCTCTAG	1702
B-WEAU	H P S R R N L S K M D N P R I K S S S K M V S P C Q H S G H K T R T K R T L S R L C R P V L S M S K	509
B-WEAU	I F V G E I Y K R W I I L G L N K I V R M Y S P V S I L D I R O G F K E P F R D Y V D R F Y K T L R	539
B-WEAU	P S Q G E K S I K D G S S S O I X K S S E C I A L S A F N T S D K D O K N P L E I M S T G S I K L S	495
A-U455	T-----G-----C-----G-----G-----G-----T-----T-----T-----A-----G-----C-----G-----C-----TA-----A-----T-----C-----	1158
A-92UG037.1	T-----G-----C-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----C-----	1073
A-Q31	T-----G-----C-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1167
B-NM	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1709
B-AD9	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1703
B-WR27	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1082
B-RF	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1217
B-696	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1702
B-BC8G3	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1250
B-JRFL	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1020
B-JACBP	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1703
B-OYI	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1249
B-CAH1	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1705
B-NYS	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1703
B-pNL43	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1703
B-IDB2	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1703
B-LA1	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1249
B-YU2	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1701
B-YU10	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1704
B-ACH320.2A.2.1P	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1705
B-ACH320.2A.1.2T	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1703
B-SF2	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1710
B-D31	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1249
B-NANC	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1649
B-HAN	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1172
B-C18MBC	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1522
C-ETM2220	TG-T-----G-----C-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1077
C-92BR025.8	TG-T-----G-----C-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1048
D-842R085	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1220
D-NOK	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1238
D-Z226	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1706
D-ELI	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1249
D-94UG114.1	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1052
F-93BR020.1	TG-T-----G-----G-----C-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1034
H-90CR056.1	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1050
AC-ZAH184	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1076
AC-92RM009	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1047
ADI-MAL	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1281
AE-93TE153	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1720
AE-CN240	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1268
AE-90CP402	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1745
AGI-2321	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	917
AG-92NG083.2	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1069
AG-92NG083.1	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1070
AG-1BM3	T-----G-----G-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1234
BF-93HR029.4	TG-T-----G-----T-----C-----G-----G-----G-----T-----T-----T-----A-----G-----C-----C-----C-----TA-----A-----T-----T-----T-----C-----	1055
O-WP5180	T-----G-----C-----G-----A-----G-----G-----G-----C-----C-----G-----A-----C-----T-----A-----G-----CT-----A-----T-----T-----G-----G-----A-----C-----T-----T-----C-----AT-----	1727
O-ANT70	T-----G-----C-----G-----A-----G-----G-----G-----C-----C-----G-----A-----C-----T-----A-----G-----CT-----A-----T-----T-----G-----G-----A-----C-----T-----T-----C-----AT-----	1793
SIVCPZANT	AG-----G-----G-----C-----G-----G-----C-----A-----G-----G-----G-----G-----C-----G-----N-----T-----A-----A-----T-----C-----A-----G-----A-----G-----G-----A-----A-----T-----	1124
SIVCP2GAB	AG-----G-----G-----C-----G-----G-----C-----A-----G-----G-----G-----G-----C-----G-----N-----T-----A-----A-----T-----C	

[illegible]

HIV1 Complete Genomes

B-WEAU	GCCTGGCCATTAAGCAAGAGTTTGGCTGAAGCAATG AGCCAACTAACCAATACA GCTACCATGATGATGACAGAGGCAAT TTACAGGCCCAAGAAAACTAATAAGTOTTTCATTTGTCACAAAGA	1981
B-WEAU	A R P G B K S F G S S M H P S N R Y I Y H D D A E R Q I S E P K R M V S V F Q L M Q R	594
B-WEAU	F Q H K A R V L A E A M S Q V T M T T M H M Q R G N F R S P R K T I R C F M C G R E	632
B-WEAU	G P A I K D E F M L K Q I A K S Q I T L P S S C R B A I L G A Q E X L L S V S I V A K	576
A-U155	A-----G-----G-----T-----CA-C-Q-----G-----A-----C-----C-----	1431
A-92UG037.1	A-----G-----G-----T-----CA-C-----A-----A-----C-----A-G-AG-A-GG-----C-----	1346
A-Q23	A-----G-----G-----T-----CA-C-----A-----A-----C-----A-G-AG-A-----GA-----C-----	1440
B-NM	A-T-T-----Q-----Q-----T-----A-----A-----A-----AT-A-----G-T-----C-----C-----	1988
B-AD8	A-----G-----T-----A-----A-----A-----AT-A-----GG-----G-----	1982
B-WR27	A-----KK-----GG-T-----AA-G-----A-----A-----A-----A-A-----G-----G-----M-----H-----	1361
B-RF	A-----A-----A-----T-----A-----A-----C-----A-----T-----GA-----A-----T-----G-----C-----T-----	1496
B-896	A-----A-----C-----G-----T-----A-----A-----A-----G-----A-----G-----C-----	1981
B-BC963	A-----C-----G-----T-----C-----A-----C-----A-----A-----G-----G-----	1529
B-JRFL	A-----C-----C-----G-----A-----A-----A-----A-----G-----G-----	1299
B-JRCSP	A-----C-----C-----A-----A-----A-----A-----A-----G-----G-----	1982
B-OYI	A-----T-----TC-----G-----A-----A-----A-----A-----G-----G-----	1525
B-CAH1	A-----T-----T-----A-----A-----A-----A-----G-----G-----C-----T-----	1984
B-NY5	A-----C-----C-----A-----A-----A-----A-----A-----G-----A-----G-----	1982
B-pNL43	A-----C-----C-----A-----A-----A-----A-----A-----G-----G-----	1982
B-KXB2	A-----G-----T-----A-----A-----A-----A-----A-----G-----G-----	1982
B-LA1	A-----G-----T-----A-----A-----A-----A-----A-----G-----G-----	1528
B-YU2	A-----T-----T-----A-----A-----A-----A-----A-----G-----G-----	1980
B-YU10	A-----T-----T-----A-----A-----A-----A-----A-----G-----G-----	1983
B-ACH320.2A.2.1A	A-----T-----T-----A-----A-----A-----A-----A-----G-----G-----	1984
B-ACH320.2A.1.2A	A-----Q-----G-----C-----C-----A-----A-----A-----G-----G-----	1982
B-972	A-----C-----C-----C-----A-----A-----A-----A-----G-----G-----	1989
B-D31	A-----T-----TG-----A-----A-----A-----A-----G-----G-----	1529
B-MANC	A-----A-----A-----T-----A-----A-----A-----A-----G-----G-----T-----	1928
B-HAN	A-----G-----G-----T-----G-----A-----A-----A-----A-----G-----G-----	1451
B-C184BC	A-----A-----A-----C-----G-----TG-----A-----A-----A-----A-----G-----G-----	1801
C-BTH2220	A-T-C-C-----G-----G-----AC-----A-----A-----A-----A-----G-----T-----A-----G-----A-----C-----G-----	1353
C-92BR025.8	A-T-----C-----G-----G-----A-----AC-----A-----A-----A-----A-----G-----A-----G-----T-----A-----G-----A-----C-----G-----	1324
D-842R085	A-----G-----G-----C-----C-----G-----G-----CT-----A-----G-----A-----A-----A-----A-----G-----A-----T-----G-----C-----	1502
D-NDR	A-----C-----G-----G-----G-----T-----GCT-----A-----G-----AG-----A-----A-----A-----A-----G-----G-----C-----	1520
D-Z236	A-----C-----G-----G-----C-----T-----GCT-----A-----G-----AG-----A-----A-----A-----A-----G-----G-----C-----	1980
D-ELI	A-----C-----G-----G-----C-----T-----GTT-----A-----G-----AG-----A-----A-----A-----A-----G-----G-----T-----	1531
D-94UG114.1	A-----A-----T-----G-----G-----C-----G-----AAT-----A-----G-----T-----A-----A-----A-----A-----A-----G-----AG-----TC-----C-----	1334
F-93BR020.1	A-TA-----G-----G-----C-----C-----G-----T-----A-----A-----A-----A-----A-----A-----G-----T-----G-----A-----C-----T-----	1310
H-90CR056.1	A-TA-----G-----G-----C-----C-----AAT-----A-----AG-----A-----A-----A-----C-----A-----G-----A-----TT-----G-----A-----C-----C-----	1332
AC-ZAM184	A-----G-----C-----A-----G-----AAT-----C-----A-----A-----A-----A-----T-----A-----A-----GA-----C-----G-----	1355
AC-92RM009	A-----G-----G-----C-----C-----AC-----A-----A-----A-----A-----A-----A-----G-----AG-----G-----T-----C-----	1323
ADI-MAL	A-----A-----T-----G-----G-----C-----C-----T-----ACT-----G-----A-----A-----A-----A-----A-----A-----G-----AG-----A-----GA-----C-----	1960
AE-93TH253	A-TA-----C-----G-----G-----C-----C-----CA-----C-----G-----AT-----A-----AG-----C-----GA-----C-----C-----	1993
AE-CM240	A-TA-----G-----G-----C-----C-----CA-----C-----G-----T-----A-----A-----A-----A-----C-----A-----G-----AG-----A-----GA-----C-----C-----T-----G-----	1541
AE-90CF402	A-T-----A-----G-----G-----T-----CACC-----AT-----A-----A-----A-----A-----A-----A-----G-----AG-----A-----AA-----C-----	2018
AGI-2321	A-----A-----A-----G-----G-----G-----C-----T-----G-----GCA-----AG-----A-----A-----A-----A-----A-----A-----G-----G-----C-----	1199
AG-92NG083.2	A-----A-----A-----G-----G-----G-----C-----T-----GG-----G-----GCA-----GCA-----AG-----A-----A-----A-----A-----A-----G-----G-----C-----	1354
AG-92NG003.1	A-----A-----C-----A-----G-----G-----GG-----C-----GG-----TCA-----AG-----A-----A-----A-----A-----A-----A-----G-----G-----C-----	1152
AG-IBNG	A-----T-----G-----G-----T-----CA-----C-----G-----C-----A-----G-----A-----A-----A-----A-----A-----A-----G-----AG-----A-----A-----C-----	1507
BF-93BR029.4	A-----A-----A-----A-----A-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----G-----A-----A-----G-----	1334
O-MVP5180	AACT-----C-----G-----A-----A-----AC-----A-----A-----GCTTCT-----C-----C-----G-----TT-----A-----AGG-----AGGATACA-----AG-----AG-----AT-----C-----A-----GC-----G-----AA-----CCA-----ATAG-----A-----GGC-----C-----A-----A-----C-----A-----	2021
O-AMT70	A-AACT-----G-----C-----GC-----A-----A-----GCTAC-----C-----C-----G-----TT-----GA-----AGG-----AGGATACA-----AG-----AG-----AT-----C-----A-----GC-----A-----AA-----CCA-----TTAG-----A-----GG-----A-----A-----A-----C-----A-----	2046
SIVCPZANT	C-----A-----C-----G-----G-----G-----A-----T-----GCTTCT-----CT-----AT-----G-----CHG-----GGA-----CG-----AG-----CT-----TC-----GGA-----AG-----AGG-----A-----G-----C-----C-----C-----A-----T-----C-----C-----T-----	1412
SIVCPZGAB	A-----ATCP-----G-----G-----C-----A-----C-----TCAATG-----TCAG-----CA-----GGG-----AGA-----AGATG-----TT-----CT-----C-----AG-----AC-----A-----GG-----GC-----G-----A-----G-----AA-----A-----A-----C-----T-----G-----	2049
p24 Capsid	P-----G-----H-----K-----A-----R-----V-----L-----A-----E-----A-----M-----S-----Q-----V-----T-----M-----T-----T-----M-----H-----M-----Q-----R-----G-----N-----P-----R-----S-----P-----R-----H-----T-----I-----K-----C-----F-----N-----C-----G-----X-----E-----Gag, Gag-Pol	

HIV1 Complete Genomes

		Gag-Pol -1 ribosomal slip site (Gag continues in frame 2, Gag-Pol in frame 1)		
B-WEAU	MSGACACATAGCCGAAATTCAGGGCCCCAGGAAAGAGGCTCTTGGAAATGTGGACAGAGAGGACCAAAATGAAAGATGTACT.....GAGAGA...CAG.....GCTAATTTTATAGGGAAGATCTGCTCTTCCAAAGGG	loop	2116	
B-WEAU	A T H S O K L Q G P S E K G L L E M W T G R T F N E R L Y I	stem-loop-stem potentiates ribosomal	-1	slip
B-WEAU	G H T A R N C F A P R K K G C W R C G O E G H O M K D C T	5 F F R E D L V F P K G	634	
B-WEAU	K D T S P E I A G P L G K R A V G N V D R J D T K S X V I	A N F L Q K I M S S O F G	677	
A-U455	---C---A---T---A---C---GA---A---C---C---A---R---I---S---L---I---F---S---G---R---S---G---L---P---K---R---		615	
A-92UG037.1	---TC---A---T---T---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1566	
A-Q23	---C---A---T---T---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1481	
B-INI	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1575	
B-AD8	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2323	
B-WR27	G---G---N---M---A---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2117	
B-RF	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1496	
B-894	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1637	
B-BCSG3	---G---T---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		2116	
B-JRFL	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1664	
B-JRCSF	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1434	
B-DYI	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2117	
B-CAN1	C---G---T---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1660	
B-WY5	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2119	
B-pNL43	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2117	
B-KXB2	---G---C---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		2117	
B-LAI	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2117	
B-LAI	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1663	
B-YU2	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2115	
B-YU10	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2110	
B-ACH320.2A.2.1	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2119	
B-ACH320.2A.1.2	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2117	
B-SF2	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2124	
B-D31	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1663	
B-NAMC	Q---G---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		2063	
B-HAN	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1586	
B-C16MBC	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1936	
C-BTH2220	---G---C---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1480	
C-92BR025.8	---G---T---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1459	
D-842R085	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1637	
D-MOK	---G---C---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1685	
D-2256	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		2123	
D-ELI	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1666	
D-94UG114.1	---G---CG---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1469	
F-93BR020.1	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1445	
H-90CR056.1	G---G---C---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1467	
AC-ZAM184	---G---A---A---G---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---A---		1490	
AC-92RW009	---C---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1458	
ADJ-MAL	---C---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1695	
AB-93TH253	---C---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		2128	
AE-CH240	---C---G---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1676	
AE-90CP402	---C---C---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		2193	
AGI-2321	TC---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1331	
AG-92NG083.2	TC---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1489	
AG-92NG003.1	TC---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1487	
AG-IBNG	---C---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1642	
BF-93BR029.4	---G---A---A---G---G---O---C---C---A---A---T---C---A---C---A---C---A---C---		1469	
O-MYF5180	G---T---A---A---C---TC---A---A---A---G---T---C---C---AAART---GA---A---A---A---A---A---A---A---A---A---		2156	
O-AMT70	G---T---A---A---C---TC---A---A---A---G---T---C---C---GAART---GA---A---A---A---A---A---A---A---A---A---		2181	
SIVCPZAMT	G---C---T---CT---A---A---A---A---G---A---C---G---C---A---A---GCAAC---A---T---GAAA---TA---ACCCACCC---CA---GTGGTG---		1556	
SIVCPZGAB	---T---C---T---T---AA---A---A---A---G---A---COG---G---A---G---C---C---A---GA---shift back---TG---AGGT---C---GC---GC---		2184	
Gag p7 NC	G H T A R N C F A P R K K G C W R C G O E G H O M K D C T E R O A N F L Q K I M S S O F G	Pol TF F F R E D L V F P K G Pol		

See introduction for ribosomal slip references.

B-NEAU
1000 1000

I-34
DEC 97

		Gag CDS end (frame 2) <--		
B-MEAN	CAGAAGCAGGAGCCA.....ATAGACAGGAGCTG.....TATCTTTAACTTCCTCCTCAATCACTCTTGGCAACGACCCCTCTCACAATAAAGATAGAGGGCACTGAAGGAAGCTCTATTAGATACAGGAGCA			2335
B-MEAN	EAGAGAG.....RQGA.....SFFNFPOITLWQRPLVLTIKIEGOLKRALLDTG			699
B-MEAN	OKQEP.....IDREL.....VPLTSLKSLFGLNDPSSQSRSRGNLSRLVSTIQE			746
B-MEAN	RSRS.....ISRL.....ILSLPSNRLSLATTFPRHNRKDRQAATRGSSIRVRS			677
A-U455	QC--A--TG--A--G--A--R--AC--GT--T--G--A--G--A--G--TA--			1782
A-92UG037.1	C--AG--AC--G--C--A--AAC--CCA--C--GT--T--T--G--A--G--A--AA--			1703
A-Q23	T--AG--A--G--A--A--GQC--C--A--C--GT--T--T--G--QA--G--A--A--			1794
B-IN	GAAGCAGGAGA.....A--G--A--C--C--G--A--T--G--A--A--			2351
B-ADB	G--G--A--A--G--G--A--G--A--			2336
B-WR27	A--G--A--A--G--G--A--A--			1715
B-RP	AA--A--A--G--A--A--			1853
B-996	G--A--C--G--G--A--G--A--			2339
B-BCSB3	GG--GA--AA--C--G--G--A--G--A--			1883
B-J8FL	G--AA--AA--G--G--A--A--			1653
B-JRC8P	GAAGCAGGAG.....G--AT--G--A--A--			2348
B-OYT	G--GA--C--G--A--A--			1879
B-CAM1	G--A--G--G--A--A--			2338
B-WY5	G--A--G--G--A--A--			2336
B-PUL43	G--A--G--G--A--A--			2336
B-HXB2	G--A--G--G--A--A--			2336
B-LAI	G--A--G--G--A--A--			1918
B-YU2	G--A--G--G--A--A--			2334
B-YU10	G--A--G--G--A--A--			2337
B-ACH220.2A.2.1	G--GG--A--C--G--G--A--A--			2338
B-ACH220.2A.1.2	G--GG--A--C--G--G--A--A--			2334
B-SF2	G--A--G--G--A--A--			2343
B-D31	G--A--G--G--A--A--			1882
B-HANC	A--G--A--C--G--G--A--A--			2282
B-HAN	G--A--C--G--G--A--A--			1802
B-C18MBC	G--G--CAG-GATCTG--G--A--G--A--			2191
C-ETH220	C--TG--A--G--A--G--A--T--T--A--G--A--G--A--G--C--C--			1728
C-92BR025.8	G--A--G--A--A--C--G--G--A--G--A--G--A--G--C--C--			1675
D-842R085	A--A--AGAGGAC--A--A--C--G--G--A--G--A--A--			1858
D-WDK	A--AG--A--A--G--A--G--A--A--			1871
D-2226	A--AG--A--A--C--G--A--G--A--A--			2339
D-ELI	A--A--AG--A--A--G--A--G--A--A--			1893
D-94UG114.1	A--A--AG--A--A--G--A--G--A--A--			1691
F-93BR020.1	C--AG--A--G--GA--TAC--CC--C--G--A--A--G--A--G--A--			1664
H-90CR056.1	C--AGCTG--AG--A--A--CC--C--G--G--A--G--A--G--A--G--			1683
AC-2AM184	C--TG--A--G--A--C--T--T--G--A--A--G--A--A--			1697
RC-92RM009	T--A--AG--A--G--A--T--T--G--A--GA--T--G--A--GA--			1668
ADT-MAL	A--AG--A--A--AT--G--A--G--A--GA--A--G--A--A--C--			1911
AB-93PH253	AG--AG--A--AT--CCTCC--C--GT--T--G--A--GA--A--A--			2323
AE-CM240	C--AG--A--A--AT--CCTCC--GT--T--G--A--GA--A--G--A--			1910
AB-90CF402	C--AG--A--A--A--AT--CCTCC--GT--T--G--A--GA--A--G--A--			2372
AGL-2321	C--A--G--GG--A--A--ATAC--CC--G--G--A--A--G--A--G--CA--TA--C--			1550
AG-92NR003.2	C--A--G--AG--G--A--A--CC--C--G--G--A--A--G--A--G--A--TA--C--			1705
AG-92NR003.1	T--G--A--GG--A--ATCA--CC--A--C--G--GA--G--A--G--A--TA--C--			1703
AG-IBNG	CAC--A--G--GG--A--ATAT--CC--G--G--A--A--G--A--TA--C--			1661
BF-93BR029.4	A--A--G--A--A--G--G--A--A--G--A--G--A--A--C--C--			1688
O-MVP5180	GT--A--GG--GATC--GG--A--C--A--TG--C--C--G--CA--AA--A--GC--G--T--G--T--ATGT--G--T--C--G--G--			2366
O-AMF70	G--A--A--AGG--GGCCCG--C--C--G--TG--C--C--G--CA--AA--A--GC--G--T--G--T--ATGT--G--T--C--G--G--			2391
SIUCP2ANT	G--GCTCAAG--GGG--GG--A--A--CC--C--GTA--G--C--G--A--AA--GA--GGA--G--TCTC--C--A--AAATGTC--G--T--			1775
SIUCP2GAB	A--GC--AAG--A--GGA--AGCAATCTA--ACC--C--G--A--C--G--AG--A--A--ATGT--G--T--C--G--T--			2397
Pol TP	SFAGA.....MRQGA.....VSFFNFPOITLWQRPLVLTIKIEGOLKRALLDTG			Pol
		Pol TP end / \ Pol Protease start		
Gag p6	OKQEP.....IDREL.....VPLTSLKSLFGLNDPSSQSRSRGNLSRLVSTIQE			Gag CDS p6 end

HIV1 Complete Genomes

[illegible]

HIV1 Complete Genomes

[illegible]

HIV1 Complete Genomes

[illegible]

Accession	Sequence	Position
B-WEAU	CTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAATACCATCGCTTCAGGCTTAAAAAGGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTACCCCTTAGAGTAAGAACTTCAGGAAGTACACTGCA	2935
B-WEAU	L N K R T Q D F N E V Q L G I P H P S G L K K K K S V T V L D V G D A Y P S V P L D E D F R K Y T A	899
B-WEAU	L T R R L K T S G G K F L G B Y H L L Q G S K R K K N O S O Y W M W V M H I F O Y P S M K T S G S T L H	921
B-WEAU	T S S E N S R L L Q S S I R N T T S F R V K R R I S N S T G C G S C I F F S T L R S R L O E V M C	864
A-U455	-C-A--A--G--C-G-A-AAG-G-TC-----A--G-C-----T-T-----AG--T-A--T--G	2382
R-92UG037.1	-C-A--A--A--T-----G--G--AG-G-C-----A--G-C-----T-T--C--G--T-A--T--	2303
A-Q23	-C-A--A--A--T-----G--G--AG-G-TC-----A--G-C-----T-T--C--G--T-A--T--	2394
B-MH	-----A-----G-----G-----T-----T-----A-----A-----T-----	2951
B-AD8	-----CQ-----CQ-----CQ-----T-----A-----A-----A-----T-----	2936
B-WR27	-----CG-----CG-----CG-----G-----W-----T-----Y-----N-----G-----T-----	231
B-RP	-----G-----G-----G-----T-----T-----T-----T-----A-----G-----T-----	2453
B-996	-----C-----C-----CQ-----T-----A-----A-----A-----A-----T-----	2935
B-BCGQ3	-----G-----G-----CG-T-A-----A-----CG-----C-----T-----A-----A-----T-----	2483
B-JRFL	-----A-----A-----CG-----G-----G-----T-----A-----A-----A-----T-----	2253
B-JRCBF	-----G-----A-----CG-----CG-----CG-----T-----A-----A-----A-----T-----	2949
B-CYI	-----G-----C-----AG-----A-----C-----T-----A-----A-----A-----T-----	2479
B-CAM1	-----A-----T-----G-----G-----T-----T-----A-----A-----A-----T-----	2939
B-MYS	-----T-----T-----C-----C-----T-----T-----A-----A-----A-----T-----	2936
B-pH43	-----T-----T-----G-----C-----C-----T-----T-----A-----A-----A-----T-----	2936
B-NXB2	-----CG-----CG-----CG-----T-----T-----T-----T-----A-----A-----A-----T-----	2936
B-LAI	-----CG-----CG-----CG-----T-----T-----T-----T-----A-----A-----A-----T-----	2518
B-YU2	-----CG-----CG-----CG-----T-----T-----C-----T-----T-----A-----A-----T-----	2934
B-YU10	-----CG-----CG-----CG-----T-----T-----C-----T-----T-----A-----A-----T-----	2936
Frameshift		
B-ACH320.2A.2.1	-----T-----C-----CG-----C-----G-----C-----T-----A-----A-----A-----T-----	2939
B-ACH320.2A.1.2	-----T-----C-----CG-----C-----C-----T-----A-----A-----A-----T-----	2936
B-SF2	-----A-----G-----CG-----CG-----T-----C-----T-----A-----A-----T-----	2943
B-D31	-----A-----A-----CG-----A-----T-----T-----A-----A-----A-----T-----	2482
B-MANC	-----A-----T-----CG-----CG-----C-----C-----T-----C-----A-----A-----T-----	2482
B-HAN	-----A-----T-----CG-----C-----A-----C-----T-----C-----A-----A-----T-----	2791
B-C10ADC	-----C-----A-----T-----G-----C-----A-----G-----C-----T-----T-----G-----A-----A-----T-----	2329
C-ETH2220	-----C-----A-----T-----G-----C-----A-----G-----C-----T-----T-----G-----A-----A-----T-----	2275
C-92BR025.8	-----C-----A-----T-----G-----C-----A-----G-----C-----T-----T-----G-----A-----A-----T-----	2275
in frame TAA stop codon		
D-B42R095	-----G-----G-----G-----A-----G-----G-----A-----C-----A-----T-----TC-----T-----T-----	2459
D-NDK	-----T-----G-----G-----G-----C-----G-----C-----C-----T-----T-----T-----A-----T-----C-----	3471
D-2226	-----T-----G-----G-----G-----C-----G-----C-----C-----T-----T-----T-----A-----T-----	3939
O-ELX	-----T-----G-----G-----G-----C-----G-----C-----C-----T-----T-----T-----A-----T-----C-----C-----	3482
D-94U3114.1	-----T-----T-----C-----G-----G-----C-----C-----C-----T-----T-----C-----A-----A-----T-----C-----	2291
F-93BR020.1	-----A-----T-----G-----G-----AG-----G-----G-----G-----C-----A-----G-----T-----T-----	2264
H-90CR056.1	-----G-----G-----C-----AG-----T-----T-----G-----G-----C-----T-----A-----A-----A-----T-----	2283
AC-ZAM184	-----C-----A-----G-----AG-----A-----G-----A-----G-----C-----C-----P-----C-----A-----A-----T-----	2297
AC-92RM009	-----C-----A-----T-----C-----G-----C-----AG-----G-----G-----G-----C-----C-----C-----T-----T-----GAG-----A-----A-----T-----	2268
ADI-HAL	-----A-----T-----T-----G-----G-----G-----T-----G-----C-----T-----C-----T-----T-----T-----	2511
AE-93TH253	-----C-----G-----T-----G-----AG-----T-----T-----G-----A-----A-----T-----T-----T-----AG-----T-----A-----T-----	2923
AE-CH240	-----C-----A-----G-----G-----AG-----T-----T-----A-----A-----A-----T-----T-----T-----AG-----T-----A-----T-----	2510
AE-90CF402	-----C-----A-----T-----T-----AG-----TC-----A-----A-----A-----C-----T-----T-----G-----T-----A-----T-----	2972
AGI-2321	-----A-----T-----G-----A-----G-----C-----CG-CA-----G-----G-----A-----A-----C-----T-----C-----A-----T-----C-----	2150
AG-92NGDB3.2	-----A-----G-----G-----T-----C-----CG-G-----G-----G-----A-----A-----C-----T-----A-----A-----T-----	3309
AG-92NG003.2	-----C-----A-----G-----C-----T-----CG-G-----G-----G-----A-----A-----C-----T-----A-----A-----T-----A-----	3303
AG-3BRG	-----C-----T-----G-----G-----T-----CG-G-----A-----G-----G-----A-----G-----T-----A-----A-----T-----A-----	2461
BF-93BM029.4	-----G-----G-----G-----CG-----G-----G-----G-----T-----A-----A-----A-----T-----T-----	3288
O-MVP5180	T-A-----A-----A-----T-----G-----G-----T-----T-----AGGG-----T-----GC-A-GCC-----T-----T-----CT-A-----A-----A-----T-----C-----TCG-----T-----CC-----T-----A-----A-----A-----C-----	2986
O-AMT70	T-A-----A-----A-----G-----G-----A-----GC-----T-----C-----GGGG-----T-----G-----GC-A-GC-----T-----T-----CT-A-----A-----A-----T-----C-----TCG-----T-----CC-----T-----A-----A-----A-----T-----	2991
SIVCP2AMT	T-A-----A-----A-----T-----T-----G-----A-----A-----T-----T-----AG-----C-----A-----GC-A-G-----G-----T-----A-----A-----A-----C-----C-----CA-----C-----G-----T-----A-----A-----T-----A-----T-----	2375
SIVCP2QAB	-----A-----G-----G-----G-----C-----T-----C-----AG-----C-----A-----G-----G-----A-----C-----C-----C-----T-----G-----C-----G-----A-----T-----T-----	2997
Pol p51 RT	L N K R T Q D F W E V Q L G I P H P S G L K K K K S V T V L D V G D A Y P S V P L D E D F R K Y T A	Pol

HIV1 Complete Genomes

B-WEAU	TTTACCATACCTAGTATAAACAATGAAACACCGGGATTAGATATCGTACAAATGTGCTTCCACAGGGATGGAAGGATCCACAGCAATATCCAAAGTAGCTGACAAAAATATTAGAGCCTTTTAGAAAAACAAATCCAGACATAGTT	3085
B-WEAU	F T I P S I M N E T P G I R Y O V N V L P O G W K G S P A I F O S S N T K I L E P F R K Q M P D I V	949
B-WEAU	L P Y L V S T N R H O G L D I S T M C F M R D G K D H O Q V S R V A S O K V S S L L L E N K I O T S L	967
B-WEAU	I Y H T S V K O S N T R D S I S V Q C A S T Q X E R I T S N I P K S H D K N I R A F S X T K S R H S	989
A-U455	-C-----G-----AG-C-G-----G-----C-----C-----TC-C-----	2512
A-92UG037.1	-C-----C-----G-----A-C-G-----G-----GGCC-----C-----TC-A-----	2453
A-Q23	-C-----C-----C-G-----AG-C-G-----G-----C-----C-----TC-A-----A-----	2544
B-WN	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3101
B-AD8	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3086
B-WR27	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2465
B-RP	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2603
B-896	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3085
B-BCSG3	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2633
B-JAFL	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2403
B-JACSF	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3098
B-OYE	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2629
B-CAM1	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3088
B-NYS	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3086
B-pRL43	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3086
B-NXB2	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3086
B-LAI	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2668
B-YU2	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3084
B-YU10	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3086
B-ACK320.2A.2.1	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3089
B-ACK320.2A.1.2	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3086
B-SF2	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3093
B-D31	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2632
B-MANC	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3032
B-HAH	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2552
B-C18MBC	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2941
C-6TH2220	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2478
C-92ER025.B	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2425
D-84ZR085	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2608
D-MDK	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2621
D-Z226	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3089
D-ELI	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2632
D-94UG114.1	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2441
F-93BR030.1	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2414
H-90CR056.1	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2433
AC-ZAM184	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2447
AC-92RM009	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2416
ADI-MAL	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2561
AE-93TH253	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3073
AE-CN240	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2660
AE-90CF402	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3122
AGI-2321	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2300
AG-92NG081.2	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2455
AG-92NG083.1	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2453
AG-1BMQ	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2611
BP-93BR029.4	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2438
Q-MVP518D	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3116
Q-AMT70	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3141
SIVCPZANT	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	2525
SIVCPZAB	-C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----	3147
Pol p51 RT	F T I P S I M N E T P G I R Y O V N V L P O G W K G S P A I F O S S N T K I L E P F R K Q M P D I V	Pol

B-WENO	ATCTATCATACATGGATTTGTATGTAGGATCTGACTTAGAATAGGGCAGCATAGAACAAAAATAGAGGAGCTGAGACACATCTGTTGAGGTGGGATTTACCAACACAGACAAAAACATCAAAAAGACCCCTCCATCTTTGG	3235
B-WENO	I V O Y N D D L V V G S D L E I G Q H R T X I E E L R O H L L R W G F T T P D K K F O K D P P F L W	999
B-WENO	S I N T N M I C K S D L T S K I G S I E O K S R S S D M I C S G G D L F P H O T K N I K K T L R S F G	1011
B-WENO	V L S T H G S P V C R I S L R N R A A S N K M R G A E T T S V S V G T Y B T R O X T S K R P S I P I	956
A-U455	-----C-----G-----T-----A-----G-----AT-A-GCT-----A-----C-----C-TT-----G-----G-----A-----T-----	1682
A-92UG037.1	-----C-----T-----G-----G-----A-AT-A-G-----A-AA-----C-----T-----G-----G-----A-----T-----	2603
A-Q23	-----T-----G-----G-----A-AT-A-GCT-----A-----C-----C-TT-----G-----G-----A-----T-----	2694
B-MN	-----G-----A-----G-----A-----C-----G-----G-----A-----T-----	3251
B-AD8	-----A-----G-----A-----C-----G-----G-----A-----T-----	3236
B-MR27	-----A-----G-----A-----C-----G-----G-----A-----T-----	2615
B-RF	-----T-----A-----G-----A-----G-----G-----G-----G-----A-----T-----	1753
B-896	-----C-----G-----T-----A-----G-----G-----G-----A-----T-----	3235
B-BC853	-----G-----A-----A-----A-----G-----G-----G-----A-----T-----	2783
B-JRFL	-----G-----A-----A-----A-----G-----G-----G-----A-----T-----	2553
B-JRCSF	-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	3248
B-OVI	-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	2779
B-CAM1	-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	3238
B-WVS	-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	3236
B-pHL43	-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	3236
B-HXB2	-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	3236
B-LAI	-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	2818
B-YU2	-----G-----C-----A-----A-----G-----G-----G-----A-----T-----	3234
B-YU10	-----G-----C-----A-----A-----G-----G-----G-----A-----T-----	3236
B-ACH320.2A.2.1	-----T-----A-----A-----A-----G-----G-----G-----A-----T-----	3238
B-ACH320.2A.1.1	-----T-----A-----A-----A-----G-----G-----G-----A-----T-----	3236
B-SF2	-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	3243
B-U31	-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2782
B-NANC	-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	3182
B-NAN	-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2702
B-CLMBC	-----C-----T-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	3091
C-ETH2230	-----T-----C-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2628
C-92BR025.8	-----T-----C-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2575
D-842R085	-----T-----C-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2758
D-MOK	-----T-----C-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2771
D-2226	-----T-----C-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	3239
D-ELI	-----T-----C-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2782
D-94DG114.1	-----T-----C-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2591
F-93BR029.1	-----C-----G-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	2564
H-90CR056.1	-----T-----C-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2583
AC-2AM184	-----C-----C-----C-----C-----A-----A-----G-----G-----G-----A-----T-----	2597
AC-92RM009	-----T-----C-----C-----C-----A-----A-----G-----G-----G-----A-----T-----	2568
ADI-MAL	-----A-----C-----G-----T-----A-----A-----A-----A-----G-----G-----G-----A-----T-----	2811
AB-93TH253	-----A-----C-----T-----C-----A-----A-----G-----G-----G-----A-----T-----	3223
AB-CN240	-----A-----C-----T-----C-----A-----A-----G-----G-----G-----A-----T-----	2810
AE-90CP402	-----C-----T-----A-----G-----G-----G-----G-----G-----A-----T-----	3272
AGI-2321	-----A-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2450
AG-92NG083.2	-----C-----A-----G-----G-----G-----G-----G-----A-----T-----	2605
AG-92NG003.1	-----G-----A-----A-----A-----G-----G-----G-----G-----G-----A-----T-----	2603
AG-IBNG	-----C-----A-----T-----T-----G-----G-----G-----G-----G-----A-----T-----	2761
BF-93BR029.4	-----C-----A-----T-----T-----G-----G-----G-----G-----G-----A-----T-----	2588
O-NVP5180	-----T-----G-----A-----C-----A-----A-----T-----CC-T-G-CAG-A-----A-----G-----G-----G-----C-----ATT-----T-----G-----G-----T-----A-----ATCA-----T-----C-----T-----T-----G-----G-----A-----C-----TT-----A-----	3266
O-ANT70	-----T-----G-----G-----CC-A-----A-----T-----CCCC-GACAG-A-----A-----G-----G-----T-----ATT-----T-----G-----C-----A-----ATCA-----C-----T-----C-----T-----G-----G-----A-----C-----T-----A-----	3291
SIVCPZANT	-----T-----C-----C-----C-----TA-G-----TACTGCA-----GA-----TG-----AA-----T-----T-----ACG-GTC-----C-----AGAG-----T-----T-----G-----G-----A-----C-----TT-----A-----	2675
SIVCPZQAB	-----T-----C-----G-----CC-A-----G-----G-----TC-T-----T-----AT-A-----A-----GG-G-----A-----A-----T-----C-----T-----AA-----G-----C-----A-----C-----G-----G-----A-----C-----TT-----A-----	3297
Pol p51 RT	I V O Y N D D L V V G S D L E I G Q H R T X I E E L R O H L L R W G F T T P D K K F O K D P P F L W	Pol

[illegible]

[illegible]

[illegible]

B-WEAU	ACCACAGAAAGCAAGTATATGGGAAAGAGCTCTTAACTTAACTACCCATACAAAGAAACATGGGAAACATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCTGAGTGGGAGTTTGTCAATACCCCTCCCTTAOTGAAATTA	3823
B-WEAU	T T E S I V I M G K T P K P K L P I O K E T W E T N W T E V M O A T N I P E W E F V N T P P L V K L	1193
B-WEAU	F Q K A S S Y Q E A L L N L M Y P Y K K K N G K N G G Q S I G K P P G P L S G S L S I P L P S S N Y	1187
B-WEAU	N H R K N S N M G R D S S I S T T H T K R N M G N M V D R V L A S H L D S S V G V C Q V P S L S E I	1136
A-1455	T-----TC-----G-----G-----TG-----G-----T-----A-----TC-----A-----	3270
A-92UG037.1	GT--TG-----G-----C-----TG--C-----G--T-----A-----TC-----A-----	3191
A-Q23	GT--TG-----G-----C-----TG--C-----G--T-----A-----TC-----A-----G	3282
B-NN	G-----G-----AC--T-----T-----A-----G-----	3839
B-AD8	G-----G-----TG-----	3824
B-MR27	G--N-----G-----N--T-----G-----T--C-----Y-----G-----	3203
B-RF	G-----G-----GG-----G-----T-----G-----G-----A-----G-----	3341
B-896	G-----T-----A-----A-----G-----G-----T-----A-----	3823
B-BCS03	G--AT-----T-----T-----T-----	3371
B-JRFL	G--AT-----T-----T-----	3141
B-JRCSF	G--AT-----T-----T-----	3836
B-OYT	G--CA-----G-----G-----A-----A-----	3367
B-CAN1	G--CA-----T-----G-----TG-----T--C-----	3826
B-MY5	G-----G-----T-----G-----G--G-----A-----G-----	3834
B-pNL43	G-----T-----G-----G--G-----G-----T-----G-----	3824
B-HXB2	-----G-----G-----G-----T-----	3824
B-LA1	-----G-----	3406
B-YU2	G-----A-----	3822
B-YU10	G-----A-----	3824
B-ACH320.2A.2.1	G-----G-----	3826
B-ACH320.2A.1.2G	G--CA-----G-----G-----TG-----T-----	3824
B-SF2	T-----T-----G-----G-----TG-----T-----	3831
B-D31	-----G-----G-----	3370
B-NAMC	-----A-----TC--C-----CG-----	3770
B-NAN	G-----G-----G-----G-----G-----	3290
B-C18MBC	G-----G-----CG-----C--A-----	3679
C-WH2220	G--CTG-----G--T-----C--G-----G-----C-----G-----A-----T-----C-----A-----	3216
C-92BR025.8	G--CTG-----A-----G--T-----C--G-----G-----C-----G-----A-----T-----T-----A-----	3163
D-842R095	G--T-----G-----G-----T-----C-----A-----T-----A-----	3346
D-NDK	G-----G-----G-----T-----A-----T-----A-----	3359
D-Z226	T-----G-----G--T-----G-----GT-----T-----A-----T-----A-----	3827
D-BLI	T-----G-----G-----G-----G-----G-----T-----A-----T-----A-----	3370
D-94UG114.1	G--CA--T--T-----G-----G-----G-----G-----G-----T-----A-----C-----T-----T-----	3179
F-93BR020.1	T--TCT-----C-----G-----G-----TT-----G-----T-----C-----C-----TC-----A-----C-----	3158
H-90CR056.1	T--T-----A--T-----G-----G-----T-----G--C-----A-----A-----T--C-----ATC-----A-----	3171
AC-ZAM184	G--TG-----G-----G-----C-----G-----G-----T-----T-----TC-----A-----C-----	3185
AC-92BR009	G--TG-----G--T-----C--G-----G-----C-----C-----T-----C-----C-----A-----C-----	3156
AD1-MAL	G--CA-----A-----G-----G-----G-----G-----C-----A-----T-----C-----A-----C-----	3399
AE-93TH253	G-----C-----G-----G-----G-----TG-----G--T-----A-----T-----TC-----A-----	3811
AE-CH240	G-----C-----G-----G-----G-----TG-----G--T-----A-----T-----TC-----A-----	3398
AE-90CF402	G-----C-----G-----G-----G-----TG-----G--T-----A-----T-----TC-----A-----	3860
AGI-Z321	G--T-----G-----G-----G-----G-----G--C-----G-----T-----A-----A-----	3038
AG-92NG083.2	G-----GG-----C-----T-----T-----G-----GT-----C-----G-----G-----TC-----A-----C-----	3193
AG-92NG003.1	G-----G-----G--AGT-----T-----AGG-----GT-----A-----G-----T-----TC-----A-----G-----	3191
AG-TUNG	G--T--TG-----C-----G-----G-----G-----TG-----G--T-----A-----TC-----A-----	3349
BF-93BR029.4	-----T-----G-----G-----G-----G-----T-----	3176
O-KVP5180	T--TCA--GCT--T-----G--ATTA-----C--GG--G--AG--TACT--G-----T-----T-----G--A-----G-----GC--A--C--A--GA--C-----	3854
O-ANT70	T--TCA--TCT--A--T--C-----ATTG-----G--G--AG--CACY--G-----T-----G--G--C-----A-----A-----GC--A--C--A--GA--C-----	3879
STVCP2AM7	GGATT-----T--C-----G--GTA--A-----C--T-----ACT--GG-----T-----YG--C-----T-----C-----G-----A--A-----A--A--T--G--C--C-----	3272
STVCP2GAB	G--T--T-----T-----A--A-----GGT-----AG-----G-----GT-----G-----G--A--C-----G--A-----C-----A-----A-----A-----C-----	3885
Pol p51 RT	T T E S I V I M G K T P K P K L P I O K E T W E T N W T E V M O A T N I P E W E F V N T P P L V K L	Pol

HIV1 Complete Genomes

B-WEAU	TGGTATCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGCGACCTAACAGGAGACTAAATTAGGAAAAGCAGGATATGTTACTAACAGAGGAAAGACAAAGAGTTGTCTCCCTAAGTACACACAACAATCAG	3973
B-WEAU	W Y Q L E K E P I V G A E T F Y V D Q A A N K E T K L G K A G Y V T N R G R O R V V S L T D T T N O	1243
B-WEAU	Q T S S R X N P S S R Q K L S M S M G O L T R R L N S R K O D M L L T E E D K B L S P S L T O O I R	1231
B-WEAU	K V S V R B R T B S R S R N F L C R H G S S Q G D F T R X E R I C V S Q R X T K S C L P N S H M K S	1182
A-U455	-----C-----C-----C-----G-A-----T-G-A-----GC-----G-----C-----G-----AG-----T-----G-----A-----	3420
A-92UG037.1	-----C-----C-----C-----G-A-----T-G-A-----GC-----G-----C-----G-----AG-----T-----G-----A-----	3341
A-Q23	-----C-A-----C-----C-----G-----T-G-A-----GC-----G-----G-T-AG-----AG-A-T-----G-----A-----	3422
B-MN	-----C-----C-----C-----T-----G-----AA-----G-----AG-----AG-----G-----A-----	3989
B-ADB	-----C-----C-----C-----T-GA-A-----AA-----G-----A-----TC-----	3974
B-WR27	-----C-A-----A-----A-----GT-G-----G-----G-----A-----A-----A-----	3353
B-RF	-----C-----A-----A-----A-----T-GA-----G-----G-----A-----A-----A-----	3451
B-B96	-----C-----C-----C-----C-----G-----C-----C-----C-----G-----A-----A-----G-----	3973
B-BC8G3	-----C-----C-----G-----A-----C-----A-----G-----G-----A-----A-----A-----A-----	3521
B-JRFL	-----C-----C-----C-----C-----G-----C-----C-----C-----T-----A-----A-----G-----	3291
B-JRCSP	-----C-----C-----A-----C-----G-----G-----G-----G-----A-----A-----A-----	3986
B-OV1	-----C-----C-----C-----C-----T-G-----G-----G-----G-----A-----A-----A-----	3517
B-CAM1	-----C-----C-----C-----C-----T-G-----A-----G-----G-----G-----A-----TC-----G-----	3976
B-NY3	-----C-----C-----A-----G-----A-----A-----C-T-G-----A-----A-----G-----T-----A-----C-T-----G-----	3974
B-pNL4.3	-----C-----C-----C-----C-----A-----C-T-G-----A-----A-----G-----A-----C-----C-----G-----	3974
B-HXB2	-----C-----C-----C-----C-----C-----G-----G-----A-----T-----A-----A-----A-----G-----	3974
B-LA1	-----C-----C-----C-----C-----G-----G-----A-----A-----T-----A-----A-----A-----	3556
B-YU2	-----C-----C-----C-----C-----A-----G-----G-----AG-----AG-----A-----A-----	3973
B-YU10	-----C-----C-----C-----C-----A-----G-----G-----AG-----AG-----A-----A-----	3974
B-ACH320.2A.2.1	-----C-----C-----G-----G-----T-G-----G-----G-----A-----A-----A-----G-----	3976
B-ACH320.2A.1.2	-----C-----C-----G-----G-----T-G-----G-----G-----A-----A-----A-----A-----	3974
B-SF2	-----C-----C-----C-----C-----T-G-----G-----G-----A-----A-----A-----A-----	3981
B-D31	-----C-----C-----C-----C-----T-G-----G-----G-----A-----A-----A-----A-----	3520
B-MRAC	-----C-----C-----C-----C-----T-G-----G-----G-----A-----A-----A-----A-----	3920
B-HAN	-----C-----C-----C-----C-----T-G-----G-----G-----A-----A-----A-----A-----	3440
B-C18NEC	-----C-----C-----A-----A-----A-----GA-----C-----C-----A-----A-----	3829
C-ETH2220	-----C-----C-----C-----C-----C-----T-----A-----T-G-----A-----A-----G-----G-----T-----G-----G-----A-----T-----T-----A-----	3366
C-92BR025.8	-----C-----C-----C-----C-----C-----A-----T-G-----A-----T-----A-----G-----G-----G-----G-----A-----T-----TA-----A-----	3313
D-842R085	-----C-----C-----C-----C-----G-----T-----A-----T-----A-----G-----G-----G-----G-----A-----C-----T-----T-----	3496
D-NDK	-----C-----C-----C-----C-----G-----A-----T-----GA-----G-----G-----G-----G-----A-----C-----TT-----C-----G-----	3509
D-2226	-----C-----C-----C-----C-----G-----A-----G-----T-----GA-----G-----G-----G-----G-----A-----C-----TT-----T-----T-----	3977
D-ELI	-----C-----C-----C-----C-----G-----A-----G-----T-----GA-----G-----G-----G-----G-----A-----C-----TT-----G-----G-----	3520
D-94UG114.1	-----C-----C-----C-----C-----G-----A-----G-----T-----GA-----G-----G-----G-----G-----A-----C-----TT-----T-----T-----	3329
F-93BR020.1	-----A-----C-----G-----C-----C-----A-----A-----T-----T-----GA-----C-----AA-----G-----G-----A-----CG-----A-----G-----T-----	3302
H-90CR056.1	-----C-----A-----C-----G-----C-----C-----A-----A-----T-----T-----G-----A-----C-----AA-----G-----G-----T-----AG-----A-----A-----G-----A-----	3321
AC-2AM184	-----C-----A-----G-----C-----C-----A-----G-----T-----GA-----C-----C-----G-----G-----G-----G-----A-----A-----G-----G-----A-----	3335
AC-92RN009	-----C-----C-----C-----C-----T-----G-----A-----TCC-----A-----A-----G-----G-----G-----G-----G-----G-----A-----T-----T-----A-----	3306
ADJ-NAL	-----C-----A-----C-----C-----C-----G-----G-----T-----GA-----A-----AAG-----G-----G-----G-----AG-----A-----T-----A-----	3549
AE-93TH253	-----C-----A-----A-----C-----C-----G-----G-----GT-----G-----GC-----G-----G-----C-----G-----G-----AG-----AA-----T-----G-----A-----	3961
AE-CH240	-----C-----A-----A-----C-----C-----G-----G-----GT-----G-----GC-----G-----G-----C-----G-----G-----AG-----A-----T-----G-----A-----	3548
AE-90CP402	-----C-----A-----A-----C-----C-----A-----G-----G-----GT-----G-----GCR-----G-----G-----C-----G-----G-----A-----A-----T-----G-----A-----	4010
AGI-2321	-----C-----A-----G-----G-----A-----A-----G-----G-----A-----CAM-----G-----G-----A-----G-----A-----G-----A-----GA-----TA-----A-----	3388
AG-92M083.2	-----A-----A-----C-----CC-----A-----A-----T-----G-----A-----A-----G-----G-----C-----A-----A-----A-----TA-----A-----C-----A-----	3343
AG-92NG003.2	-----GA-----A-----C-----CC-----A-----A-----T-----A-----A-----A-----G-----C-----G-----A-----A-----A-----A-----A-----CAG-----A-----	3341
AG-18NG	-----C-----A-----C-----C-----A-----A-----T-----T-----GA-----GA-----G-----G-----C-----G-----G-----AG-----T-----A-----	3499
BF-93BR029.4	-----C-----C-----C-----C-----T-----G-----A-----A-----GA-----G-----G-----G-----G-----G-----A-----C-----C-----G-----	3326
O-MVP5180	-----C-----A-----C-----T-----T-----G-----C-----T-----A-----A-----T-----A-----C-----G-----G-----G-----AG-----ACA-----A-----G-----ACA-----AA-----AAAGT-----GAA-----G-----C-----A-----	4004
O-ANT70	-----CAG-----C-----A-----GT-----T-----TA-----G-----G-----C-----AT-----A-----A-----T-----T-----GA-----A-----C-----C-----G-----AG-----ACA-----G-----A-----AGA-----AA-----AAAAT-----GA-----G-----C-----C-----A-----	4029
SIVCP2AM7	-----A-----TC-----TT-----GC-----C-----T-----TCC-----AG-----T-----C-----T-----T-----T-----GAA-----CT-----C-----G-----G-----G-----C-----G-----AG-----G-----AGT-----G-----AAAAC-----CAAA-----G-----C-----C-----A-----	3422
SIVCP2GAB	-----CAGT-----C-----T-----CC-----ACCA-----T-----AT-----A-----A-----T-----T-----G-----A-----A-----ACT-----G-----G-----A-----AG-----A-----G-----A-----CA-----TAG-----T-----GAAA-----T-----C-----T-----	4035
Pol p51 RT	W Y Q L E K E P I V G A E T F Y V D Q A A N K E T K L G K A G Y V T N R G R O R V V S L T D T T N O	

Pol p51 Reverse Transcriptase and /A Pol p15 RNase H start

HTV1 Complete Genomes

B-WBAU	AAGACTGAGTTTACAAGCTATTCATCTAGCTTTGCCAGGATTGGGATTAGAGTAAACATAGTGACAGACTCCCAATTATGCATTAGGANTCATTCAAGCACAACCAGATCAGAGTGAATCAGAGTTAGTCAAGTCAAAATAATAGAACGTGA	4123
B-WBAU	K T E L O A I M L A L O D S G L E V N I V T D S O Y A L O I O A O P D O S E S E L V S O I I S O L	1293
B-WBAU	R L S Y X L F L L L C R I R D S K S T S S O T P N M H S F R K N O I R V N Q S S S V R S S G N S S	1271
B-WBAU	R D S Y V Y S Y S S S F A G F G I R S K R S D R L P I C I R N H S F T T R S E S I R V S O S N N R T V	1230
A-0455	-----A-----T-----A-----C-----C-----A-----C-----T-----A-----A-----G-----G-----G-----CAG-----AA-----A-----GA-----C-----	3570
A-92UG037.1	-----A-----T-----A-----C-----C-----A-----C-----A-----G-----G-----G-----CAG-----CA-----A-----GA-----C-----	3491
A-Q23	-----AC-----T-----A-----C-----T-----C-----A-----A-----C-----A-----A-----G-----G-----C-----G-----CA-----A-----GA-----C-----	3582
B-MN	-----A-----A-----A-----G-----A-----A-----A-----A-----A-----A-----G-----G-----C-----G-----CA-----A-----GA-----C-----	4139
B-AD8	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4124
B-WR27	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3503
B-RF	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3641
B-896	-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4123
B-BCSG3	-----G-----A-----A-----T-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3671
B-JRFL	-----A-----A-----A-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3461
B-JRCEP	-----A-----A-----A-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4136
B-OTX	-----A-----A-----A-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3667
B-CAM1	-----A-----A-----A-----T-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4126
B-NYS	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4124
B-gML43	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4124
B-NXB2	-----A-----A-----A-----T-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4124
B-LAT	-----A-----A-----A-----T-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3706
B-YU2	-----A-----A-----A-----T-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4122
B-YU10	-----A-----A-----A-----T-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4124
B_ACH320.2A.2.1	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4126
B_ACH320.2A.1.2	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4124
B-SF2	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4131
B-D31	-----A-----A-----A-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3670
B-MANC	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4070
B-HAN	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3590
B-C18NC	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3979
C-EYM2220	-----A-----G-----C-----G-----G-----A-----A-----C-----A-----A-----G-----A-----G-----G-----C-----T-----A-----A-----A-----A-----	3516
C-92BR025.8	-----A-----A-----G-----A-----A-----C-----A-----A-----G-----A-----G-----A-----T-----A-----A-----A-----A-----	3463
D-842RD85	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3646
D-NDK	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3659
D-2226	-----A-----A-----A-----T-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4127
D-ELI	-----A-----A-----A-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3670
D-94UG114.1	-----A-----A-----A-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3479
F-93BR020.1	-----G-----A-----A-----A-----GT-----A-----A-----A-----C-----A-----A-----G-----A-----A-----G-----A-----A-----A-----A-----	3452
H-90CR056.1	-----A-----A-----A-----T-----A-----T-----A-----A-----A-----G-----G-----A-----A-----A-----G-----C-----A-----A-----A-----A-----	3471
AC-2AM180	-----G-----G-----A-----A-----T-----A-----CT-----T-----G-----A-----A-----A-----G-----G-----T-----A-----A-----G-----A-----A-----A-----	3485
AC-92RM009	-----A-----A-----A-----A-----G-----A-----A-----C-----A-----C-----A-----A-----C-----A-----A-----A-----A-----A-----A-----	3456
ADX-HAL	-----A-----A-----A-----C-----CT-----A-----A-----C-----A-----A-----G-----A-----G-----T-----A-----A-----A-----A-----A-----A-----	3699
AE-93TH253	-----A-----A-----T-----G-----C-----T-----C-----A-----C-----A-----C-----A-----A-----G-----G-----T-----G-----A-----A-----A-----A-----	4111
AE-CM240	-----A-----A-----T-----G-----C-----T-----C-----A-----C-----A-----C-----A-----A-----G-----G-----T-----G-----A-----A-----A-----A-----	3698
AE-90CF402	-----A-----A-----T-----G-----C-----T-----C-----A-----C-----A-----C-----A-----A-----CC-----A-----A-----T-----G-----G-----A-----A-----	4160
AG1-Z321	-----AC-----G-----A-----A-----A-----A-----A-----T-----A-----A-----A-----G-----A-----T-----G-----G-----A-----A-----A-----A-----	3338
AG-92MG083.2	-----G-----A-----A-----T-----A-----A-----A-----C-----AA-----CC-----A-----A-----A-----A-----G-----A-----G-----A-----A-----A-----	3493
AG-92MG003.1	-----A-----A-----A-----C-----A-----A-----G-----A-----A-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3491
AG-1BM0	-----A-----A-----A-----T-----A-----T-----C-----A-----A-----A-----A-----A-----A-----G-----A-----G-----A-----A-----A-----A-----	3649
BF-93BR029.4	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3476
O-MVP5150	-----G-----A-----A-----ATG-----G-----ATTAA-----C-----A-----CAAGBAGC-----A-----A-----A-----T-----G-----C-----ATCCT-----C-----ACA-----C-----CCCTA-----TCAG-----G-----GG-----AC-----	4154
O-ANT70	-----AC-----A-----A-----ATG-----G-----ATTAA-----C-----A-----CAA-----GA-----AC-----T-----A-----A-----T-----A-----G-----CG-----CTCCT-----C-----TACA-----C-----CCCTA-----TCAG-----G-----GG-----AC-----	4179
SVCPZAMT	-----C-----AG-----A-----A-----G-----G-----T-----A-----G-----C-----AG-----AAC-----GGCCC-----C-----T-----C-----T-----A-----C-----G-----T-----G-----GTACC-----A-----CCCC-----GGAG-----CC-----GA-----A-----	3572
SVCPZGAB	-----C-----AG-----A-----A-----G-----T-----G-----T-----G-----C-----A-----ATCA-----C-----G-----T-----A-----T-----A-----G-----T-----G-----GTACC-----A-----CCCC-----GGAG-----CC-----GA-----A-----	4185
p15 RNAse	K T E L O A I M L A L O D S G L E V N I V T D S O Y A L G I O A O P D O S E S E L V S O I I S O L	Pol

Ps1 p15 RMase N end \ Ps1 p31 Interram start

I-49
DEC 97

B-WBAU	ACACATTTAGAGGGGAAAAATTATCTGTGGTAGCAGTTCATGTAGCCAGTGGAATATATAGAAGCAGAACTTATCCAGCAGAGACAGGCCAAGAACAGCATACTTTATCTTAAAAC TAGCAGGAGATGGCCAGTAAAAACAATACATACA	4573
B-WBAU	T H L E G K I I L V A V H V A S G Y I E A E V I P A E T Q G O E T A Y F I L K L A G R W P V K T I M T	1443
B-WBAU	H T S R E K L S N G F M G F V D I F K Q X L L F D O R O G K K Q H T L S S N S O B R D G G S K Q V I Q	1401
B-WBAU	Y T F R Q K N Y F G S S C S Q W I Y R S R G Y S S R D R A R N S I L Y L K T S R R R M A S K N M N T Y	1369
A-D455	-C-A--G-A--A-----C-----C-C-----A-G-Q-----AC-T-----GT-----C-	4020
A-92W037.1	--C-A--G-A-T-----C-----G-T---C-CG-----A--A-G-G-----C-AC--G-----GT-G--C--	3941
A-Q23	-A-G--G-R-TG-----C-----C-----A--A--Q-----C-AC--T-----T-G--C--	4032
B-MN	-A--G-----C-----C-----G-G-Q-----C-----T-----	4589
B-ADB	-A-----C-----G-----A-G-----C-----T-----C-----	4574
B-WR27	-A-----G-----G-----T-----F-----N-W-----	3955
B-RF	-C-A-----C-----A-A-G-----T-----GT-----	4091
B-096	-A--G-----G-----G-----C-C-T-----C-----	4573
B-BCS03	-C--A-----G-G-----G-----C-----T-----	4121
B-JRP1	-A-----A--G-----C-----T-----	3891
B-JRCBP	-A-----A--G-----C-----T-----C-----C-----	4586
B-OVI	-A-----G-----G-----T-----	4117
B-CAM1	-A--G-----G-----C-----T-----C-----	4576
B-NV5	-A--G--T-----G-----A-----CC-----T-----G-----	4574
B-pML43	-A--Q--T-----A-----CC-----T-----G-----	4574
B-HXB2	-A--G-----A-----G-----T-C-T-----T-----	4574
B-LAI	-A--G-----A-----G-----C-T-----P-----	4156
B-YU2	-A--G-----G-----C-----P-----	4572
B-YU10	-A--G-----G-----C-----T-----C-----	4574
B_ACH320.2A.2.1	-A-----A-----T-----C-----	4576
B_ACH320.2A.1.2	-A-----A-----T-----C-----	4574
B-SF2	-C-A-----G-----T-C-----T-----	4581
B-D31	G-A--G-----G-----T-----	4120
B-MANC	-A--G-----G-----C-----T-----C-----	4520
B-HAN	-A-----T-----C-----T-----C-----G-----	4040
B-C18HBC	-C-A-----C-----C-----G-----T-----G-----G-----	4429
C-ETH2220	-C-A--G-----C-----C-C--G-----A--A-----C-AC-T-----G-----C-GGT-----	3966
C-93BR025.0	-A-----C-----C-----G-----C-----A-----AC-T-----C-GT-----	3911
D-042R095	-C-A-----G-----C-----A--G--G-----C-----T-----GT-G-G-----	4096
D-MDX	-C-G-A--G-----C-----A--G-----G-----C-----T-----GT-G-----	4109
D-Z226	-A--G-----C-----A--G-----T-----T-----T-----T-----GT-G-----	4577
D-BLE	-C-A-----G-----C-----A--G-----T-C-T-----T-----T-----GT-G-----	4120
D-94UG114.1	-A-----GG-----C-----T-----A-----A--G-----C-C-T-G-----GT-G-----	3929
F-93BR020.1	-A-----C-----T-----G-CC-----C-----A--A--G-----C-CC-AC-GT-----	3902
H-90CR056.1	-G-A--C-Q--T-----C-----C-C-----A--AA-G-----CC-G-G-----A-C-----GT-----	3921
AC-ZAM184	-C-A-----G-----C-A-----C--G-----C-C-----A--A--G-G-----A-----GT-----	3935
AC-92RW009	-C-G-A-----A-----C-----T-----G-----A-----A--G-G-----AC-T-----C-----GT-----	3906
ADI-MAL	-C-A-----A--A-A-----C-----C-----A--A--G-G-----AC-T-----GT-G-----C-----	4149
AE-93TH253	-C-A-----G-C-----C-C--G-----A--A--G-G-----C-GC-T-----G-----GT-----C-----	4561
AE-CN240	-C-A-----G-C-----C-C--G-----A--A--G-G-----C-CC-----GT-----C-----	4148
AE-90CF402	-G-C-A-----G-C-----C-C--G-----A--A--G-G-----C-CC-T-----GGGT-----C-----	4610
AGI-Z321	-A-----C-----C-----C-----A--A--G-----A-----T-----G-----GT-G-----C-----	3788
AG-92NG083.2	-A-----A-----A-A-----C-----C-----A--A--G-----A-----T-----G-----GTG-----	3943
AG-92MG003.1	-C-A-----G-C--TA-A-----C-----C-----A--A--G-G-----CC-OC-----GT-----C-----	3941
AG-IENG	-A--G-----T-A-----C-----C-----A--A--G-G-----A-----T-----G-----GT-----C-----	4099
BF-93BR029.4	-C-A-----G-----G-----O-----C-----A-----T-----T-----	3926
O-MVE5180	-A-C-----C-AA-A--T-T-C-----A--AC-T-----G-G-A-----A--A--G-----T-C-T-CC-G-----T-----C-----T-C-----GT-----	4604
O-AMT70	-A-G-A-----C-AA-A--T-T-C-----G-A--G-TC-----G-A-----A--A--G-----T-C-----CC-G-----G--T-C-----T-T-----GT-----	4629
STVCPZANT	-C-----A--T--GA-A--A--G-----G-----T-T-----C-TC-----T-----A--GG-AT-----A--AGT-----CC-G-----T-----CA-C-----C-T-----	4622
STVCPZGB	-C-CC-----C-G-G-A-A--G-----C-----T-----T-----C-----G-----T-----T-----C-----	4635
Integratee	T H L E G K I I L V A V H V A S G Y I E A E V I P A E T G O E T A Y F I L K L A G R W P V K T I M T	Fol

HIV1 Complete Genomes

[illegible]

[illegible]

B-NEAD	C A A A T T A C A A A A T T C A A A A T T T C G G G T C T A T T A C A G G G A C A C A G A G A T C C A C T T T G G A A A G G A C C G C A A A G C T T C T C T G G A A A G G T G A A G G G C A G T A G T A A T A C A A G A T A A T A G T G A C A T A A A .	5002
B-NEAD	O T T K I O N F A V Y V R D S R D P L W K Q P A K L L W K G R G A V V I O D N S D I F .	1586
B-NEAD	R L O K F K T P G S I T Q T A R I H F G R D Q Q S F S G R V R G O S S V K I I V T S F .	1530
B-NEAD	T N Y K N S K F S G L L O G O O R E T L E R T S X A S L E R G R G S S N T R S S S H K I .	1503
A-0455	-----T-----C-CA-----A-A-----G-----G-C-----T-----G-----	4449
A-9200037.1	-----T-----A-----T-G-----A-----A-A-----G-----G-C-----T-----G-----	4370
A-023	-----C-----T-----A-----A-----A-----A-----G-----G-C-----T-----G-----	4461
B-MH	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5016
B-AD8	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5003
B-NR27	-----M-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4382
B-RF	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4520
B-896	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5002
B-BC903	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4550
B-JRFL	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4320
B-JRC9F	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5015
B-OYI	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4546
B-CAM1	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5005
B-NY5	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5003
B-pNL43	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5003
B-HXB2	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5003
B-LAI	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4585
B-YU2	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5001
B-YU10	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5003
B-ACH320.2A.1.1	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5005
B-ACH320.2A.1.2	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5003
B-862	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5010
B-B31	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4549
B-NRXC	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4949
B-HAN	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4469
B-C18MB	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4858
C-RTN2230	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4395
C-92BR025.8	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4341
D-84CR085	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4925
D-MOK	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4538
D-2226	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5006
D-BLI	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4549
D-940G114.1	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4358
F-93BR020.1	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4331
H-90CR056.1	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4350
AC-2AM184	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4364
AC-92RM009	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4335
ADI-NAL	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4578
AE-93TH253	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4990
AE-CH240	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4577
AE-90CF402	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5039
AGI-2321	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4217
AG-92NG083.2	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4372
AG-92NG083.1	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4370
AG-IBNG	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4529
BP-93BR029.4	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4355
G-NVP5180	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5033
O-ANT70	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5058
STVCP2ANT	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	4472
STVCP2QAB	-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----	5064
Integrase	O T T K I O N F A V Y V R D S R D P L W K Q P A K L L W K G R G A V V I O D N S D I F .	Pol

HIV1 Complete Genomes

B-WEAU	5002
B-WEAU	1586
B-WEAU	1530
B-WEAU	1503
A-U455	4449
A-92UG037.1	4370
A-Q23	4461
B-MN	5019
B-AD8	5003
B-WR27	4382
B-RP	4520
B-896	5002
B-BCSG3	4550
B-JRPL	4320
B-JRCSF	5015
B-CYE	4546
B-CAM1	5005
B-NT5	5003
B-pML43	5003
B-HXB2	5003
B-LAI	4585
B-YU2	5001
B-YU10	5003
B_ACH320.2A.2.1	5005
B_ACH320.2A.1.2	5003
B-SF2	5010
B-D31	4549
B-MANC	4949
B-MAN	4469
B-C18MBC	4858
C-ETH2320	4395
C-92BR025.0	4341
D-04ZR085	4525
D-NDK	4538
D-Z226	5006
D-ELI	4549
D-94UG114.1	4358
F-93BR020.1	4331
H-90CR056.1	4350
AC-ZAM184	4364
AC-92RM009	4335
ADY-MAL	4578
AE-93TH253	4990
AE-CW240	4577
AE-90CT402	5039
AGI-2321	4217
AG-92NG083.2	4372
AG-92NG003.1	4370
AG-IBNG	4529
BF-93BR029.4	4355
O-MVE5180	5033
O-ANT70	5058
SIVCPZANT	4622
Large insert in SIVCPZANT is not significantly similar to any sequence in GenBank as of 10/27/97. BLAST search only "hits" self.	
SIVCPZGAB	5064

	Vif CDS start	Pol CDS end
	--> (frame 2)	(frame 1) <--
B-WEAU	GTAGTCCCAAGAGAAAGCAAGATPCATTACCAATTATGGAACACAGATGCCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATAGACATG	5101
B-WEAU	V V P R R K A K I T R D V G K O N A G D D C V A S R Q D E D S M M	1618
B-WEAU	S C Q E E K Q R S L G I M E N R W Q V M I V W Q V D R M R I R Y W	1882
B-WEAU	I S A V X X S X D N S Q L W R T D Q R S S L C G K S T G S G L E M	1530
A-U455	A	4548
A-920G037.1	A	4469
A-Q23	A	4560
B-NM7	A	5117
B-AD4	A	5102
B-VR27	A	4481
B-RP	A	4619
B-896	A	5101
B-BC903	A	4649
B-JRFL	T	4419
B-JRC6F	T A	5114
B-OYI	G	4645
B-CAM1	A	5104
B-NYS	C	5102
B-pML43	C	5102
B-HXB2	A	5102
B-LAI	A	4684
B-YU2	A	5100
B-YU10	A	5102
B-ACH320.2A.2.1	A	5104
B-ACH320.2A.1.2	A	5102
B-SF2	A	5109
B-D31	T	4648
B-MANC	T T G	5048
B-HAV	GCA	4568
B-C18MBC	T	4957
C-ETH222D	A G G A	4494
C-92BR025.9	A G Q T A C	4440
D-842R065	A	4624
D-NDR	A	4637
D-2226	A T T C	5105
D-ELI	A T	4648
D-94U0114.1	A T	4457
F-93BR020.1	T	4430
H-90CR056.1	A G Q A	4449
AC-ZAN164	A T C	4463
AC-92RW009	A	4434
ADI-MAL	A	4677
AB-93TH253	A	5089
AE-CH240	A	4676
AE-90CF402	A	5138
AGI-Z321	A	4316
AG-92HG083.2	A C A	4471
AG-92HG003.1	A T A	4469
AG-IBNG	A	4627
BP-93BR029.4	A T	4454
O-MVP5180	G A G A A C A	5132
O-ANT70	A G G A C A G	5157
	Pol continue for HIV-1 G group	
SIVCPZANT	CAAGGACGGGTGAGGTGCCCCGTGCGAGCACTGATCTCGTCCAGTTCCTCCAG-CA-ATGAGGA--GATCTTC-CAG-ACT-AGCA--ACTGCATCGGTGCGC-C-CGC--	4772
SIVCP2GAB	A	5163
Integrase	V V P R R K A K I T R D V G K O N A G D D C V A S R Q D E D S Pol CDS end	
	Vif protein start M E N R W Q V M I V W Q V D R M R I R T W Vif	

B-WEAU	GAAGGTTTAGTAAACACCATATGTATTTTCAAGGAAGGCTAAGGATGCTTATAGACATCACTACGAAAGCAGCTATCCAGAAATAGTTTCAGAAAGTACATATCCCACTAGGG	GAAGGTAATATTAGTAAACACATA	5245
B-WEAU	E K F L G K T P Y V Y F K E S S G W V L E T S L R M H S S K N K P R B T Y F P T R S	E R S T S N M N I	1662
B-WEAU	K S L V K H H M Y I S K K A K G W S Y R H H Y E S T H F R I S S E V H I P L G	E G K L V I T T Y	1610
B-WEAU	G K V S S N T I C I F O R K L R D G L I D I T T K A L I O E S V Q K Y I S H S F	F K V N S G S O M	1570
A-U455	-----T-C-----G-C-----C-A-T-----T-----T-OA-----T-G-----C-----C-V-G-----G-G-----	-----C-G-----G-G-----	4692
A-92N037.1	-----C-C-----T-----C-----CG-G-----A-T-----T-----T-CG-----A-G-----C-----A-----T-C-G-A-----G-G-----	-----T-C-G-A-----G-G-----	4613
A-Q23	-----C-----T-----C-G-C-----C-----A-AA-G-----T-----T-----GG-----A-G-----C-----C-----T-C-C-----G-G-----	-----T-C-C-----G-G-----	4704
B-MN	-----T-----T-----A-----C-----T-----T-----T-----C-----C-----T-----C-G-G-----G-----A-----	-----C-C-G-----G-----A-----	5261
B-AD8	-----G-----G-----AA-----GT-----A-----T-----TGA-----C-----C-----C-----G-----G-----	-----C-C-G-----G-----	5246
B-WR27	-----MC-----GG-----G-A-----NT-----AN-----T-----A-----A-----T-C-----TG-----G-----	-----T-C-TG-----G-----	4625
B-RF	-----G-----G-----G-----G-----T-----T-----A-----A-----G-----C-----C-----T-AA-GG-----B-----	-----T-AA-GG-----B-----	4763
B-886	-----GG-----GG-----T-----T-----A-----C-----C-----C-----T-----T-C-----G-G-----	-----T-C-----G-G-----	5245
B-BCSG3	-----T-----T-----G-----A-----T-----T-----T-----C-----A-----C-----C-----T-C-G-G-----	-----T-C-G-G-----	4793
B-JRPL	-----C-----GG-----AT-----T-----T-----T-----C-----A-----C-----C-----T-C-G-G-----	-----T-C-G-G-----	4563
B-JRCSP	-----C-----GG-----AT-----A-----T-----T-----A-----G-----A-----A-----T-C-----G-G-----	-----T-C-G-G-----	5258
B-OY1	-----G-----G-----T-----T-----T-----T-----C-----C-----C-----T-----C-CC-----G-----G-----	-----T-C-CC-----G-----G-----	4789
B-CAN1	-----C-----G-----GG-----AA-----C-----C-----T-----T-----C-----G-----C-----G-----G-----	-----G-C-G-----G-----G-----	5248
B-MY5	-----A-----G-----AC-----T-----T-----T-----A-----A-----A-----T-----C-----	-----T-C-----	5246
B-pNL43	-----A-----G-----AC-----T-----T-----T-----A-----A-----A-----T-----C-----	-----T-C-----	5246
B-HXB2	-----G-----GG-----G-----T-----T-----T-----C-----C-----C-----T-----C-G-G-----	-----T-C-G-G-----	5246
B-LAI	-----G-----GG-----G-----T-----T-----T-----C-----C-----C-----T-----C-G-G-----	-----T-C-G-G-----	4828
B-YU2	-----GG-----G-----T-----T-----T-----TC-----C-----C-----T-----C-G-----	-----T-C-G-----	5244
B-YU10	-----GG-----G-----T-----T-----T-----TC-----C-----C-----T-----C-G-----	-----T-C-G-----	5246
B-ACH320.2A.2.1	-----AG-----G-----A-----T-----T-----T-----C-----C-----C-----G-----C-G-G-----	-----G-C-G-G-----	5248
B-ACH320.2A.1.2	-----AG-----G-----A-----T-----T-----T-----C-----C-----C-----G-----C-G-G-----	-----G-C-G-G-----	5246
B-SF2	-----G-----A-----T-----T-----T-----T-----G-----C-----C-----T-----C-G-----	-----T-C-G-----	5253
B-D31	-----G-----G-----GG-----G-AA-----T-----A-----T-----T-----A-----C-----C-----T-C-G-G-----	-----T-C-G-G-----	4792
B-MARC	-----C-----GG-----A-----A-----T-----T-----T-----A-----G-----C-----C-----T-----T-----	-----T-C-----G-----A-----	5192
B-HAN	-----B-----TG-----CG-----AAC-----GT-----T-----T-----A-----A-----G-----C-----C-----G-----	-----C-G-----G-----	4712
B-C10MC	-----C-----G-----G-----AT-----T-----T-----T-----A-----A-----C-----C-----C-----G-----G-----	-----C-G-G-----G-----	5101
C-ETH2220	-----T-----G-----C-----G-----G-----T-----GT-----T-----T-----C-----GA-----AGG-----T-----	-----G-C-G-----A-----A-----	4638
C-92WR025.8	-----T-----G-----C-----G-----G-----GT-----A-----T-----T-----T-----CA-----A-----C-----T-----	-----G-C-G-----T-----	4584
D-B42RD05	-----T-----C-----C-----A-----T-----C-----T-----C-----C-----C-----A-----A-----T-----T-----A-----	-----T-C-G-C-G-----G-----	4768
D-MDE	-----T-----G-----CA-----T-----T-----T-----C-----CAC-----C-----A-----C-----C-----A-----	-----C-G-C-G-----G-----	4781
D-Z226	-----G-----G-----GCA-----T-----T-----T-----C-----C-----C-----A-----C-----T-----A-----	-----C-TGC-----B-----G-----A-----	5249
D-SLI	-----G-----CA-----T-----T-----T-----C-----C-----C-----A-----C-----T-----A-----	-----C-G-C-G-----A-----	4792
D-94DG114.1	-----G-----A-----TG-----T-----CT-----C-AA-----A-----C-----T-----A-----	-----C-GCC-----G-----A-----	4601
F-93BR020.1	-----T-----C-----C-----A-----T-----T-----TT-----GG-----A-----C-----C-----G-----G-----	-----AC-C-G-----G-----	4574
F-90CR055.1	-----C-----GT-----C-----G-----GA-----T-----T-----T-----T-----G-----C-----T-----A-----	-----C-CGG-----C-----C-----	4593
AC-ZAN184	-----C-----T-----G-----C-----T-----TC-----T-----T-----GA-----GG-----G-----C-----G-----	-----G-C-G-----G-----G-----	4607
AC-92RW009	-----T-----C-----G-----G-----G-----C-----T-----T-----T-----GA-----A-----C-----C-----T-----	-----C-G-----A-----A-----	4578
ADI-MAL	-----C-----T-----G-----C-----AAAT-----T-----T-----T-----GG-----A-----G-----C-----G-----	-----T-C-G-----G-----G-----	4821
RE-93YH253	-----C-----C-----C-----ACAG-----T-----T-----T-----CAG-----AGG-----G-----A-----	-----C-C-G-----G-----	5233
AE-CH240	-----C-----T-----C-----AAAG-----T-----T-----T-----CAG-----AGG-----C-----A-----	-----G-C-G-----C-----	4810
AE-90CF402	-----C-----T-----GC-----GCCAA-----T-----T-----T-----CAG-----A-----G-----C-----C-----	-----T-C-G-----A-----G-----	5282
ADI-Z321	-----G-----T-----AG-----AAAC-----T-----T-----C-----A-----A-----C-----C-----	-----C-C-----G-----G-----	4460
AG-92NG083.2	-----C-----T-----G-----C-----A-----T-----T-----T-----GG-----G-----G-----A-----A-----	-----T-C-C-C-----G-----G-----	4615
AG-92N009.1	-----T-----T-----AA-----T-----AT-----T-----T-----T-----GG-----A-----G-----C-----	-----G-C-G-----G-----G-----	4613
AG-1BM3	-----C-----T-----G-----C-----T-----T-----T-----TT-----T-----GG-----A-----G-----T-----	-----T-C-GG-----G-----G-----	4771
BF-93BR029.4	-----T-----C-----G-----C-----AA-----T-----T-----T-----T-----GG-----G-----A-----	-----C-G-----A-----	4598
O-MVP5180	-----C-----CC-----G-----T-----AA-----C-----G-----T-----G-----CGC-----AAC-----CG-----G-----T-----TC-----GGA-----A-----G-----C-----G-----CG-----GT-----T-----C-----CA-----	-----C-G-TA-----GG-----C-----C-----	5276
O-ANT70	-----C-----CC-----G-----T-----A-----C-----GG-----T-----GA-----GA-----CG-----AAC-----GG-----C-----T-----T-----TC-----GAA-----G-----T-----AGT-----T-----T-----G-----	-----TG-----CCC-----TC-----G-----C-----	5301
----- Pol CDS end for HIV-1 O group			
SIVCP2AMT	TCC-----G-----C-----A-----GGGAGA-----T-----AGT-----CT-----ACC-----AG-----T-----T-----ATGA-----C-----AG-----A-----G-----AGA-----A-----CCAACATG-----TAAA-----G-----G-----G-----GTG-----T		4922
SIVCP2GAB	T-----T-----T-----T-----GG-----A-----G-----A-----GA-----T-----T-----T-----TCA-----C-----A-----G-----TGC-----T-----A-----C-----A-----TT-----CA-----A-----GATTA-----TTC-----A-----A-----G-----G-----C-----T-----		5310
Uif	K S L V K H H M Y I S K K A K G W S Y R H H Y E S T H F R I S S E V H I P L G	E G K L V I T T Y	Vif

HIV1 Complete Genomes

B-WEAU	TTGGGCTCTGCATACAGGAGAGAGACTGGCAATTGGGCTCAAGGAGTCTCCATAGAAATGGAGGAAACAGAGATATAGCACAAGTAGACCTTGACCTAGCAGACCAACTAATTCATCTGTATCATTTGATTTGTTTTCAGAGTCCGC	5395
B-WEAU	L G S A Y R R R L L A F G S R S L R R R R T E I G H T E R P S P S R P T N E S V S F S I F F R V R	1709
B-WEAU	L G L H T G E R D M H L G Q G V S T R W R K Q R V S T Q V D P D L A D Q L I H L Y H F D C F S E S A	1660
B-WEAU	L G V C T O S R R T G I M V Y R S P S N G C N R D I A H R S T L T S O T M S F I C I L I V F Q S F	1616
A-U455	-----A-----C-----T-----G-----CTGA-A-----T-----C-----C-----T-----C-----A-----T-----	4862
A-S2UG037.1	-----G-----A-----C-----T-----G-----CT-A-A-----A-----G-----C-----T-----A-----C-----C-----T-----	4763
A-Q23	-----A-----C-----T-----G-----CT-A-A-----A-----ACT-----G-----A-----C-----T-----C-----T-----	4854
B-MH	-----A-----A-----G-----A-----A-----A-----C-----T-----C-----C-----T-----C-----T-----	5411
B-ADB	-----G-----A-----A-----A-----G-----A-----A-----T-----A-----T-----A-----T-----	5396
B-YR27	-----G-----A-----A-----AC-----A-----AC-----G-----T-----A-----AC-----T-----	4775
B-RF	-----G-----A-----AG-----A-----A-----C-----CT-----A-----T-----A-----T-----	4913
B-B96	-----G-----A-----A-----G-----G-----G-----T-----C-----A-----C-----T-----	5395
B-BC903	-----T-----T-----G-----A-----T-----T-----C-----T-----C-----A-----T-----	4943
B-JRF1	-----G-----G-----A-----A-----T-----T-----C-----A-----T-----A-----T-----	4713
B-JRC9P	-----G-----G-----G-----C-----AG-----A-----T-----C-----A-----T-----A-----T-----	5408
B-OYI	-----A-----A-----C-----C-----T-----A-----G-----A-----ACA-----T-----A-----T-----	4939
B-CAN1	-----G-----C-----T-----C-----A-----G-----A-----G-----CT-----A-----T-----	5396
B-NYS	-----G-----A-----A-----A-----A-----C-----CT-----A-----T-----A-----T-----	5396
B-PNL43	-----G-----A-----A-----A-----C-----CT-----A-----T-----A-----T-----	5396
B-KXB2	-----G-----A-----A-----A-----T-----T-----C-----C-----C-----T-----	5396
B-LAI	-----C-----G-----A-----A-----A-----T-----T-----C-----C-----C-----T-----	4978
B-YU2	-----C-----G-----A-----A-----A-----T-----T-----C-----A-----T-----	5394
B-YU10	-----C-----G-----A-----A-----A-----T-----T-----C-----A-----T-----	5396
B-ACB320.2A.2.1	-----C-----G-----A-----A-----T-----G-----T-----T-----A-----T-----	5396
B-ACB320.2A.1.2	-----C-----G-----A-----A-----T-----G-----T-----T-----A-----T-----	5396
B-SF2	-----A-----C-----G-----G-----A-----A-----G-----C-----T-----A-----T-----	5403
B-D31	-----A-----C-----A-----A-----A-----G-----CT-----A-----T-----	4942
B-HAMC	-----G-----C-----C-----A-----G-----C-----T-----G-----T-----A-----T-----	5342
B-HAN	-----C-----G-----C-----AG-----A-----A-----C-----T-----A-----T-----	4862
B-C18NBC	-----A-----G-----AG-----C-----A-----A-----C-----A-----AC-----T-----C-----A-----T-----	5251
C-ETH2220	-----T-----A-----T-----T-----ATTGAGA-----C-----A-----G-----G-----C-----A-----C-----T-----G-----A-----T-----	4788
C-92BR025.8	-----T-----A-----T-----G-----ACTGAGA-----G-----G-----G-----A-----C-----T-----G-----C-----T-----	4734
D-842R085	-----A-----A-----C-----G-----AG-----A-----G-----G-----A-----A-----T-----A-----G-----C-----T-----	4918
D-NDK	-----A-----A-----C-----G-----AG-----A-----G-----G-----A-----A-----T-----A-----G-----C-----T-----	4931
D-Z226	-----A-----A-----C-----G-----AG-----A-----G-----G-----A-----A-----T-----A-----G-----C-----T-----	5399
D-ELI	-----A-----A-----C-----G-----AG-----A-----G-----G-----A-----A-----T-----A-----G-----C-----T-----	4942
D-94UG114.1	-----A-----G-----G-----A-----A-----A-----G-----G-----A-----A-----T-----A-----G-----C-----T-----	4751
P-93BR020.1	C-----G-----T-----C-----A-----C-----G-----C-----GGG-----G-----A-----A-----G-----G-----G-----C-----A-----A-----T-----A-----T-----	4724
H-90CR056.1	C-----A-----A-----A-----A-----C-----G-----ACTGA-----A-----G-----G-----G-----G-----A-----C-----T-----A-----T-----	4743
AC-ZAM184	-----CT-----A-----A-----C-----A-----T-----G-----C-----GA-----A-----G-----T-----A-----T-----G-----T-----C-----A-----T-----	4797
AC-92BM009	-----T-----A-----G-----T-----ATTGAGA-----AG-----G-----G-----G-----C-----A-----C-----T-----G-----C-----T-----	4728
ADI-NAL	-----A-----A-----A-----C-----T-----G-----C-----GA-----A-----C-----T-----G-----G-----C-----CT-----A-----T-----	4921
AE-93TH253	-----G-----AG-----A-----T-----C-----GAGA-----A-----A-----T-----G-----G-----AC-----AT-----C-----C-----T-----A-----	5383
AE-CH240	-----A-----AG-----A-----C-----C-----GAGA-----A-----A-----T-----A-----G-----AC-----AT-----GC-----C-----T-----	4970
AE-90CP402	-----A-----A-----C-----T-----C-----GAQA-----A-----G-----A-----T-----G-----G-----C-----AT-----C-----T-----	5432
AGI-4321	-----A-----A-----G-----G-----CTGAG-----G-----C-----A-----C-----G-----G-----A-----A-----C-----T-----C-----A-----T-----	4610
AG-92NG083.2	-----G-----A-----A-----C-----T-----G-----T-----C-----GA-----A-----T-----A-----A-----AC-----T-----G-----T-----C-----A-----T-----	4765
AG-92NG003.1	-----C-----G-----G-----G-----A-----C-----GA2A-----A-----A-----T-----A-----G-----G-----C-----C-----T-----A-----C-----	4763
AG-18HG	-----G-----C-----C-----T-----G-----A-----A-----C-----GA-----A-----G-----A-----A-----T-----T-----G-----C-----C-----A-----T-----	4921
BF-93BR029.4	-----G-----T-----C-----G-----G-----G-----G-----G-----A-----G-----G-----G-----C-----A-----A-----T-----A-----T-----	4748
O-MVP5180	AT-AATGC-----G-----GGAA-----C-----A-----T-----G-----TAGT-----CAAT-----CA-----GAG-----AA-----GA-----T-----T-----AAC-----AGGA-----G-----A-----C-----T-----CACA-----A-----A-----A-----	5426
O-ANT70	AT-AATGC-----G-----TGAA-----A-----A-----T-----G-----TAGT-----CT-----TA-----AG-----AA-----GA-----T-----AAC-----AGGA-----G-----A-----C-----T-----ACC-----A-----CA-----A-----	5451
STVCPZANT	AT-A-----ATGT-----C-----CCCA-----C-----A-----A-----T-----T-----GAG-----ATGTGGA-----AG-----T-----A-----AAC-----T-----T-----A-----G-----A-----G-----AA-----T-----CCCC-----TCAA-----	5072
STVCP2GAB	CCT-AAGCC-----G-----C-----A-----C-----T-----TC-----CT-----GATCT-----GT-----G-----G-----T-----CTT-----AC-----TAG-----T-----G-----C-----CTCT-----AT-----C-----G-----AA-----T-----	5460
Vif	M G L H T G E R D W H L G Q G V S I E W R K Q R Y S T Q V D P D L A D Q L I H L Y H F D C F S E S A Vif	

B-WEAU	TATAAGAANTGCCATATTAGGACATTTAGTATTTCCTAGGGTGAATATCAAGCGGGCATAACAGGTAGGATCTCTACAATACTTGGCA	CTAACAGCATTAAACACCAACCAAAAGTAAGCCACCTTTGCCTAG	5533
B-WEAU	Y K K C H I R T F S Y S S V S I S S R A S Q G R I S T I L G A	I N S I M M T P K K D K A T F A S	1750
B-WEAU	T R N A I L O H L V I P R C E Y Q A G H N R V G S L Q V L A	L T A L I T P K K I K P P L P S	1706
B-WEAU	L S E M P Y S D I S L F L G V N I K Q G I T A S D L Y N T M	S O M S S M O K R S S H L C L	1653
A-0455	C-G-A-A-AA-QA-T-T	G-A-G-C-G-GC	4980
A-92UG037.1	C-G-A-G-AG-GC-T-A-A-C-T-A	G-A-G-TC-G-Q	4901
A-Q23	C-G-A-G-AG-G-AA-A-A-T-A	G-A-G-C	4992
B-MN	C-G-A-AG-G-TT-T-A-G	G-A-G	5549
B-AD9	C-G-A-T-CAG-G-A-A-G-C	G-G-C	5534
B-WR27	C-A-G-G-A-A-A	C-A-R	4913
B-RF	C-G-ATC-A-G-G-G-C	G-C-A	5051
B-096	C-GT-A-GC-T-A-A-G-C	G-C-G-G-C-T	5533
B-BCS03	C-A-G-A-A-A-G-C-A	G-G-G	5081
B-JRFL	C-A-G-A-A-A-G-G	A-T	4851
B-JRCSF	C-G-A-G-A-A-G-G	A-G	5546
B-OYI	C-G-A-A-G-A	G-G-A	5077
B-CAM1	C-G-G-GA-GC-G-G	G	5536
B-NY5	C-G-A-G-A-G-G	G-A-C	5534
B-pNL43	C-A-G-A-A-G-G	A-C	5534
B-HXB2	C-G-T-CA-GC-G	G	5534
B-LA2	C-G-T-A-GC-A	G	5116
B-YU2	C-G-T-AG-G-G	C	5532
B-YU10	C-T-AG-G-A-G	C	5534
B-ACH120.2A.2.1	C-G-G-G-A	A	5536
B-ACH120.2A.1.2	C-G-G-A-A	A-C	5534
B-SF2	C-A-T-AG-G-A	G-C	5541
B-D31	C-G-A-G-GC-T-G	G-G	5080
B-KAMC	C-A-A-GC-G-A	G-G-C	5480
B-NAM	C-G-A-G-A-G	G-C-T	5000
B-CLN8C	C-G-T-AG-GG-A-G	G	5309
C-ETH2220	C-G-A-T-AG-GC-C-T	G-A-G-GC	4926
C-92BR025.8	C-G-A-AG-GCT-C-T	G-G-A-C	4972
D-842R095	C-A-A-G-A	G-G-C	5056
D-NDK	C-A-A-G-T-G-A	G-G-T	5069
D-2226	C-A-A-C-G-A-T-G	G-G-T	5537
D-ELI	C-A-G-A-G-A-G	G-G-T	5080
D-94DU114.1	C-A-AG-TA-A-T-A	G-G-G-A	4889
D-93BR020.1	C-G-A-AA-A-GC-A-C	G-T-G	4852
K-90CR056.1	C-G-A-G-G-GA-A-C-C-A-AC-A	T-G-G-G-G	4881
AC-2AM184	C-G-A-G-AA-G-A-T-G-T	T-G-G-G-G-T-T-C-G-C-C	4895
AC-92RND09	C-G-A-A-GC-C-T-A	G-A-G-A-T	4866
ADI-KAL	C-A-A-G-A-T	T-G-T	5109
AB-937H253	C-G-GA-AG-GA-G-C-T-A	G-A-C-G-G-C	5521
AE-CM240	C-G-A-AG-GA-G-C-T-A	G-A-C-G-G-C	5108
AE-90CF402	C-G-G-A-T-AA-GA-C-T-A	G-A-C-A-A-C	5570
AGI-Z321	C-G-A-AG-GC-A-T-TC	TCC-A-G-C-G-A-G	4748
AG-92M3083.2	C-G-A-G-GA-GC-C-C-A-A-T-T	TCC-A-G-C-G-A-G	4903
AG-92M3003.1	CG-G-A-C-G-AG-GA-TC-A-T-A	C-A-G-C-C-A	4901
AG-JBMG	C-G-A-G-AA-GA-T-G	G-AT-T-G-G-C	5059
BF-92BR029.4	C-G-A-AG-A-GC-T-G	A-G	4866
O-NVP5180	A-C-G-G-TC-G-SAG-GC-GA-C-A-T-C-TG-A-GTC-GA-A-T-A-C	T-G-A-G-G-G-A-GT-GA-AT-T-CC-A-C	5564
O-ANY7D	AG-C-G-A-CC-GAG-AC-GA-C-A-C-CTA-A-GTC-GA-A-CTCC-A-T	G-G-G-A-G-G-GC-GG-T-CC-A-C	5589
SIVCPZAMT	AG-C-C-A-A-CC-G-AG-A-AT-GA-ATAC-CC-C-CA-GAAG-A-C-C-C	GA-TTCAG-G3-TAC-C-GGGTCC-GAAG-AG-C	5222
SIVCPZGAB	C-C-GA-T-GC-T-AGCG-C-T-CA-A-A-C-GSC-CT-A-G-T	A-T-T-GAG-G-GAT-GA-A	5598
VIF	I R N A I L G H L V I P R C E Y Q A G H N K V Q S L Q V L A	L T A L I T P K K I K P P L P S	1750

	Vpr CDS start --> (frame 1)	Vif CDS end (frame 2) <--	
B-WEAD	TGTTAAGAACTGACAGAGATGGAACAAGCCCAAGGCCCCACAGAGGGAACCATACAAATGAGTGGACACTAGAGCTTTTATAGAGGAAGCTTAAGAGTGAAGCTGTAGACATTTTCCTAAGGTAAGGCTCCATAGCTTA		5693
B-WEAD	C S E T D R G S M E D A P E D O G P O R E P Y N E M T L E L L E E L K S E A V R H F F R V W L H S L		1798
B-WEAD	V K K L T E D R W N K P O K T K G H R G S H T M N G H S S F S R N L E V K L L D I F L G V G S T A S		1753
B-WEAD	V L R N S O R I D G T S P R R P R A T E G A T O S M D T R A F R G T S E S S C S T F S S G M A P S L		1696
A-0455	-----T-A-----G-----G-GC-----A-G-T-----T-G-----A-G-----CA-----A-----CAG-----G-A-----		5130
A-92UG037.1	-----T-AG-----G-----G-----A-----TGT-----T-G-----A-T-----CA-----CC-----G-A-----		5051
A-023	-----G-T-T-A-----A-----G-----T-----C-----TGT-----G-----T-T-----CA-----CC-----T-G-A-----		5142
B-MN	-----CA-----T-----G-----G-----A-----G-----A-----A-----A-----CC-----G-----G-----		5699
B-AD8	-----CA-----T-----G-----G-----A-----A-----A-----A-----A-----CC-----G-----G-----		5684
B-WR27	-----N-----C-----A-----G-----A-----A-----A-----A-----C-----T-----G-----		5063
B-RF	-----C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----		5201
B-696	-----C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----		5683
B-BCSG3	-----CA-----A-----A-----A-----A-----A-----A-----A-----A-----A-----G-----		5231
B-JRPL	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----C-----		5001
B-JRCSE	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----C-----		5696
B-DYI	-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----G-----		5227
B-CAM1	-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----G-----		5686
B-NYS	-----G-----C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----		5684
B-PHL43	-----G-----C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----		5684
B-HXB3	-----C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----A-----T-----G-----		5684
B-LAI	-----C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----A-----T-----G-----		5266
B-YU2	-----A-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----CC-----A-----G-----		5682
B-YU10	-----C-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----CC-----A-----G-----		5684
B_ACH320.2A.2.1	-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----CC-----T-----G-----		5686
B_ACH320.2A.1.2	-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----CC-----T-----G-----		5684
B-SF2	-----GCA-----C-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----CC-----		5691
B-D31	-----GCA-----C-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----C-----A-----T-----		5230
B-MANC	-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----G-----A-----A-----		5630
B-HAN	-----CA-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----C-----A-----A-----		5150
B-C19NBC	-----GC-----C-----C-----C-----C-----C-----C-----C-----C-----C-----C-----G-----A-----G-----		5539
C-BTH2220	-----GT-----T-AGT-----A-----G-----C-----A-----A-----G-----G-----C-----CAG-----C-----C-----ACC-----A-----		5076
C-92BR025.8	-----T-AGT-----G-----A-----G-----A-----A-----A-----A-----A-----C-----C-----A-----C-----C-----ACC-----T-----G-----		5022
D-84ZR085	-----G-----A-----A-----A-----G-----C-----G-----G-----G-----G-----G-----G-----G-----C-----A-----T-----		5206
D-BDK	-----G-----G-----A-----A-----A-----G-----G-----G-----G-----G-----G-----G-----C-----A-----T-----		5219
D-Z2E6	-----G-----GT-----A-----A-----A-----G-----A-----G-----G-----G-----G-----G-----A-----A-----		5687
D-BLI	-----G-----G-----A-----A-----A-----C-----G-----G-----G-----G-----G-----G-----G-----A-----A-----		5230
D-94UG114.1	C-----GG-----G-----A-----C-----GA-----G-----G-----G-----G-----G-----G-----G-----C-----G-----CC-----T-----C-----		5039
F-93BR020.1	-----CC-----AGT-----A-----C-----G-----G-----A-----A-----A-----T-----T-----G-----G-----A-----A-----CC-----		5012
H-90CR056.1	-----GA-----G-----AGT-----G-----G-----C-----C-----C-----C-----C-----G-----GA-----G-----G-----CAA-----		5031
AC-2AN184	-----G-----T-AG-----G-----G-----G-----G-----C-----GTGT-----A-----G-----G-----C-----CAG-----C-----CAG-----A-----G-----C-----G-----		5045
AC-92RN009	-----GT-----T-AGT-----A-----A-----G-----G-----A-----A-----A-----C-----C-----CAG-----C-----C-----ACC-----GA-----		5016
ADI-MAL	-----G-----G-----A-----A-----A-----C-----C-----C-----C-----C-----T-----A-----A-----G-----G-----CAA-----C-----C-----T-----		9259
AE-93TH353	-----A-----T-----A-----A-----A-----T-----T-----A-----A-----C-----T-----A-----G-----G-----A-----A-----CCC-----G-----		5671
AE-CM240	-----G-----T-----A-----A-----A-----G-----A-----A-----C-----T-----A-----G-----G-----A-----A-----CCC-----AG-----		5252
AE-90CF402	-----G-----T-----A-----A-----A-----C-----A-----C-----A-----T-----A-----G-----G-----A-----C-----CCC-----G-----		5720
AGI-2321	-----AGT-----G-----A-----A-----A-----A-----A-----A-----A-----G-----G-----A-----A-----CC-----G-----A-----		4898
	in frame TGA stop codon for Vpr		
AG-92NG083.2	-----GG-----T-AG-----A-----G-----A-----A-----A-----C-----T-----A-----G-----A-----G-----A-----A-----CCC-----G-----		5051
AG-92NG093.1	-----GT-----A-----A-----G-----G-----G-----T-----C-----A-----G-----A-----A-----ACA-----T-----G-----A-----		5038
	frameshift		
AG-IBNG	-----G-----GT-AG-----A-----C-----AG-----G-----G-----G-----G-----A-----G-----A-----G-----ACA-----C-----CCG-----G-----A-----		5209
BF-93BR029.4	-----CC-----G-T-----A-----A-----TG-----A-----T-----G-----A-----GCT-----T-----C-----G-----A-----G-----G-----AGAA-----A-----A-----C-----CCT-----T-----A-----AGC-----GT-----		5016
O-MVP180	-----CC-----T-----A-----A-----A-----T-----GA-----T-----G-----A-----GCT-----AA-----C-----T-----G-----C-----C-----A-----G-----A-----AGCA-----A-----A-----C-----CCT-----A-----CGC-----G-----		5714
O-AN770	-----C-----GTCT-----TCT-----A-----A-----QC-----G-----TG-----G-----A-----A-----ATG-----G-----TTGT-----CAC-----A-----A-----A-----A-----G-----A-----A-----C-----ACA-----CCTACAT-----A-----ACA-----		5739
SIYCP2AMT	-----CGCC-----T-----A-----A-----G-----T-----CA-----A-----A-----T-----C-----G-----G-----T-----ACC-----G-----A-----T-----A-----A-----A-----A-----C-----ACC-----T-----G-----CAA-----		9371
	frameshift		
SIYCP2GAB	-----CGCC-----T-----A-----A-----G-----T-----CA-----A-----A-----T-----C-----G-----G-----T-----ACC-----G-----A-----T-----A-----A-----A-----A-----C-----ACC-----T-----G-----CAA-----		5748
Vif	Vpr protein CDS start M E D A P E D O G P O R E P Y N E M T L E L L E E L K S E A V R H F F R V W L H S L Vpr V K K L T E D R W N K P O K T K G H R G S H T M N G H S S Vif protein CDS end		

[illegible]

	Tat CDS start (frame 2) -->	<-- Vpr CDS end (frame 1)		
B-WEAU	GCAG...GAAATGGAGCCAGTAGATCATAGCTAGAGCCCTGGAGCATCCAGGAGTCAAGCTAAGACTCTCTTGTACCAATGCTATGTGTAAGATGTTGCTTTTCATTGCCAAGTTTGTTCATGACAAAAGGCTTAGGCATCTCCFA			5960
B-WEAU	A F N G A S R S S T R A L E A S R K S A S D C L Y O L L L E K M L L S L P S L F J H D K R L R L L L			1889
B-WEAU	O E N E P V D H R L E P W K H P G S Q P K T A C T N C Y C K R C C F H C O V C F M T X G L G I S Y			1943
B-WEAU	S # K N S O G Y T O G S P G S T O B V S I R L L V P L A I V R D V A P T A K F V E S O K A S A S P			1782
A-U455	-GC-...-G-...-C-AC-...-A-C-G-...-CA-...-G-C-T-...-GTG-...-G-...-C-...-TC-...-A-...-T-...			5415
A-92UG037.1	-TC-...-GG-...-T-G-...-C-C-...-C-G-...-A-C-...-A-G-T-C-...-GTG-...-A-...-TGC-...-TC-...-AC-...-G-...			5339
A-Q23	-GC-...-G-...-C-AC-...-T-C-G-...-CA-...-CA-A-T-C-...-AG-...-T-A-...-T-...-C-TT-...-AC-...			5427
B-MN	...-C-AC-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5984
B-AD8	A-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5966
B-WR27	...-T-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5348
B-RP	...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5486
B-896	A-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5968
B-BC803	...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5516
B-JRFL	...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5286
B-JRC5F	...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5981
B-OYT	...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5512
B-CA81	T-...-G-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5971
B-NYS	...-G-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5969
B-pNL43	...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5969
B-HX82	...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5970
B-LAI	...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...-C-...			5551
B-YU2	-GAA-...-C-AC-...-G-...-A-...-A-...-AG-...-A-...-T-CA-A-...-C-...			5970
B-YU10	-GAA-...-C-AC-...-G-...-A-...-A-...-AG-...-A-...-CA-A-...-AG-...			5972
B-ACH320.2A.2.1	-GAA-...-C-A-...-A-...-A-...-AG-...-A-...-CA-A-...-AG-...			5974
B-ACH320.2A.1.2	-GAA-...-C-A-...-A-...-A-...-AG-...-A-...-CA-A-...-AG-...			5972
B-SF2	-GAA-...-C-AT-...-G-...-A-...-A-...-AG-...-T-C-CG-...-CA-G-...			5979
B-D31	...-C-...-G-...-A-...-C-...-A-...-C-...-CA-A-...-CA-A-...			5515
B-NANC	A-...-C-...-T-...-T-...-A-...-A-...-AG-...-G-...-T-A-...-T-...			5915
B-HAN	...-C-...-C-...-G-...-CC-...-AG-...-C-...-CA-G-...-A-...			5435
B-CHMBC	...-C-...-G-...-C-...-AG-...-C-...-AG-...-C-...-TC-...-T-...			5824
C-ETH2220	...-C-AC-...-A-...-C-...-ATC-A-T-...-A-...-A-...-T-T-...-C-TC-...-T-...			5360
			frameshift	
C-92BR025.8	...-C-AC-...-C-...-A-...-C-AT-...-T-C-...-CA-A-...-T-T-...-C-TCA-...-T-...			5307
D-842R085	...-T-...-C-ACA-...-T-...-G-...-A-...-T-...-AG-...-A-...-C-A-...-G-...			5491
D-MOK	...-T-...-C-AT-...-T-...-G-...-AT-G-TC-...-AG-...-A-...-C-A-...-G-...			5504
D-2226	...-T-...-C-ACA-...-C-...-A-...-AG-TC-...-AG-...-A-...-C-A-...-G-...			5972
D-KLI	...-T-...-C-AC-...-G-...-G-...-A-...-G-TC-...-AG-...-A-...-C-...-T-A-AC-...			5515
D-94UG114.1	AT-...-CC-AT-...-G-...-C-...-A-...-A-...-G-...-A-...-TT-C-G-...-Q-A-G-...			5324
			defective stop codon	
F-93BR020.1	-T-...-A-T-...-C-ACT-...-T-...-C-...-C-...-GA-T-...-T-G-...-TT-CTGG-...-C-T-CA-...-G-...			5297
H-90CR056.1	-T-...-C-...-C-A-...-C-...-C-A-...-A-...-T-...-AG-C-...-A-...-A-G-C-...-TT-A-AG-...-A-T-...			5316
AC-ZAM184	-T-...-GG-...-T-G-...-C-C-...-C-...-A-...-A-...-G-T-...-AG-...-A-...-C-G-C-...-TC-A-AC-...-A-...			5330
AC-92BR009	...-C-...-C-...-C-...-C-...-A-...-C-...-AT-C-T-...-CAC-A-...-A-...-T-T-...-C-CA-G-...-T-...			5301
AD1-HAL	...-T-...-C-ACT-...-C-...-G-...-G-...-GC-...-AT-G-T-...-AG-C-...-A-...-A-G-C-...-A-G-...			5544
AE-93TH253	-GC-...-G-...-TC-...-C-AC-...-T-...-G-...-CA-...-G-...-G-T-C-...-A-...-GG-...-C-A-...-C-TC-...-A-...			5957
AE-CH240	-GC-...-G-...-A-TG-...-C-AC-...-T-...-G-...-CA-...-G-...-G-T-C-...-A-...-GG-...-C-A-...-C-TC-...-A-...			5537
AE-90CF402	-GC-...-G-...-G-...-C-AC-...-A-...-G-...-CA-...-G-...-G-T-C-...-TG-...-GG-...-C-A-...-C-TC-...-A-...			6009
AGI-5321	-GG-...-C-...-C-AC-...-T-...-G-...-C-A-...-A-...-AG-...-T-...-C-...-TTCA-...-GC-G-...			5183
AG-92NG083.2	-T-...-GG-...-C-G-...-C-AG-...-T-...-G-...-CA-...-C-C-...-A-...-A-T-...-GTG-C-...-GG-...-T-...-C-...-TT-A-AC-...			5338
AG-92NG003.1	-GC-...-G-...-TG-...-C-C-...-C-...-CA-...-A-...-A-T-C-...-T-C-...-GG-...-T-G-...-C-TC-...-AC-G-...			5323
AG-IEHQ	-GC-...-G-...-TG-...-C-C-...-C-...-CA-...-G-...-G-T-C-...-TG-C-...-G-...-T-G-...-C-TC-...-AC-G-...			5493
			frameshift	
BF-93BR029.4	...-G-...-C-...-G-...-C-...-G-...-AT-G-...-AG-...-T-...-CA-...-G-...			5921
O-MVP5180	-G-GAA-...-T-...-C-GAG-GGCC-...-T-...-C-T-...-C-T-...-G-CA-...-CC-A-...-CC-...-AT-...-C-...-C-...-A-...-T-T-...-CA-AG-...-G-T-...-G-A-...-C-...			6011
O-AM70	-G-GAA-...-T-...-C-GAG-GGCC-...-T-...-C-C-...-C-T-...-CC-TCC-...-AT-...-C-...-C-...-A-...-T-T-...-G-A-G-...-G-T-...-G-A-...-C-...			6027
STVCP2AWT	...-C-...-C-...-CGC-GA-AC-CCT-...-A-TTA-...-C-TCCTC-ACA-AGCA-...-GC-A-...-C-AT-...-C-T-...-T-C-...-T-AC-...-CTC-C-...-C-CA-A-...-G-G-...-T-...			5632
STVCP2GAB	AG-T-...-CC-...-T-A-...-C-GAC-...-G-...-A-...-C-...-A-GA-...-A-T-...-AT-...-C-...-GCT-...-T-A-...-C-TAT-TA-...-C-CA-A-...-A-...			6033
Vpr	A-...-R-N-G-A-S-R-S-S-...-Vpr protein CDS end			
	Tat protein CDS start A R P V D H R L E P W K H P G S Q P K T A C T N C Y C X R C C F H C O V C F M T X G L G I S Y Tat			

HIV1 Complete Genomes

[illegible]

HIV1 Complete Genomes

[illegible]

HIV1 Complete Genomes

[illegible]

[illegible]

[illegible]

HTV1 Complete Genomes

[illegible]

5-WEAU	AAGAGAAAGACTGAATATGCACCTTTTATAAAGTGTGTAATGCCAATAGATCAT	GATAATACA	AGCTATACGTTGATAAATTTGTAATCTCTCAACCATTTACACAG	6843
B-WEAU	K R K T E Y A L F Y K L D V M F I D H	D N T	S Y T L I N C R S S F I T Q	2152
B-WEAU	R R L M A H F P I M L V M S C Q G I		F A Y V D K L S I L M N Y T	2093
B-NSAU	S R R D S I C T F L S F S C N A M R S			2047
A-U455	A-C-AA-T-T-T-G-A-G-A-A-CA-A-A	ACT-C-AT	T-GGC-A-A-TR	6278
A-92UG037.1	ATCG-AG-T-T-T-G-A-A-TA-GGT	AATAACAG-G-AT	CTG-CA-A-TA-G-C	6218
A-Q23	C-AA-T-T-T-G-A-G-A-TR-G-A	AATC-GGG-GT	GMA-CA-A-TA-G	6276
B-MN	TGC-AA-C-G-A-G-AT	G-C	G-AA-G-TA-GT	6862
B-ADU	GTC-AA-C-G-A-G-A	T	G-G-TA-GT	6817
B-WR27	GTGC-AA-C-G-A-G-A-G	AATAATAC-C	T-G-G-C-CA	6229
B-RF	CGC-AAA-CG-A-GA-GGTAATATTAGCCCTAAGATAATAGTACCA	TAGC	TATGGT-A-A-C-T-GT	6390
B-896	GT-AA-A-G-G-A-AA	ACTA-T	AG-G-A-G-CA-GT	6837
B-BC903	AGC-AA-C-C-G-A-A	GATAAGA-B-C	AA-G-G-G-CA-GT-C	6365
B-JRPL	G-GTGC-AA-T-G-A-A	A-C	G-G-B-G-CA-GT	6140
B-JRCSF	GTGC-AA-T-G-A-A	AAGA-C	AA-G-A-G-CA-GT	6838
B-OYT	TGC-GAAC-T-A-G-A	A-G-T	AA-T-G-A-C-CA	6390
B-CAM1	GTGG-AA-G-A-A-G	C	T-A-C-CA-GT	6837
B-WY5	T-C-A-AA-A-G-A-A-G		A-A-G-CA-GT	6773
B-pvL43	GTGC-AA-T-C-A-A	C	G-B-C-CA-GT	6822
B-HX83	GTGC-AA-T-A-A	A	A-C-G-CA-GT	6833
B-LAI	GTGC-AA-T-A-A	A	A-C-G-CA-GT	6426
B-YU2	GT-C-AA-C-G-A-A	G-T	G-G-G-CA-GT	6821
B-YU10	GT-C-AA-C-G-A-A	G-T	G-G-G-CA-GT	6821
B-ACH320.2A.2.1	GCGC-TA-C-A-AATATACT	AATACCGCT-C	G-G-G-CA-GT	6846
B-ACH320.2A.1.2	GTTC-AA-AA-C-G-GA-G-GTAATGCT	ACTACCAAAATTAAMC-C-G	A-T-G-C-C-GT	6862
B-SF2	TTC-AA-A-CG-C-G-A-GCTAGTACT	ACTMCC-CT-C	A-G-C-CAGA-GT	6850
B-D31	GTGC-AA-AC-G-A-A	C-T	T-G-G-G-CA-GT	6378
B-NANC	GTGC-T-AA-G-A-A-G	A-G-T	T-GA-G-CA	6790
B-HAN	GT-C-AA-C-AC-G-A-AATAAGACT	AGTAATAGGGATA-C-C	T-T-C-CAGA-TGT	6304
B-C18NCC	GTGC-AAC-AC-G-A-G	AATGATA	G-A-G-CA-T	6693
C-BTN2240	A-G-AG-C-T-A-G-A-C-TA-A-C	G-TC-T	GAT-GA-A-TA-TA	6236
C-92BR025.8	G-AG-T-C-G-A-G-A-C-TA-GA-GAG	AGTAATACAC-TC-GGT	GAT-GA-A-TA-G-A	6186
D-842R085	G-CAA-T-C-G-G-A-A-R-GAGGT	AAAAATGAAATCAATGATAC-T-GGT	CT-G-A-TA-G	6375
D-NDR	GG-CAA-T-A-G-CR-AAT	AATAGGCCA-G-T	AT-G-A-G-TA	6352
D-2226	C-ACAA-T-C-G-A-G-A-C-CA-GAT	AGTAATACCA-G-C	AT-GA-A-TA-G	6838
D-ELI	AGC-CAA-T-C-G-A-G-A-CR-GAT	AGTAATACCA-G-C	AT-GA-A-TA-G	6381
D-94UH114.1	AG-CAA-T-C-G-GG-AAA-A-G	AATGATAGT-C-C	GAC-A-TA-G	6163
F-93BR020.1	CAGCT-AA-T-C-A-G-A-A-CA-CA-G	GAT-C-G	CA-C-GAC-A-G-TG	6172
R-90CR056.1	CAGC-AA-T-C-G-G-A-T-A-C	A-G-T	CAG-G-C-A-TR-GT	6164
AC-7AM184	CGC-GA-TG-T-G-GA-A-T-CA	AGTA-T	TG-GAC-A-G-G-CG	6176
AC-92RN009	A-C-GA-TC-T-GG-A-G-A-TA-AGCAATAGT	AATACCGTAG-C-AT	CAG-G-A-TA-G	6177
ADI-MAL	A-G-CAA-AC-C-C-G-A-A-G	AGT-GT	T-B-C-A-TA-GTA	6422
AB-93TH253	AGC-AG-TC-G-A-GA-A-GA-AG	A-G-GT	GAG-G-A-TA-T-GT	6835
AE-CN240	AGC-AG-TCC-C-G-A-G-A-T-AG	AGGAGAC-G-GT	GAG-G-A-TA-T-GT	6413
AE-90CP403	A-AGC-AG-TC-C-GTC-G-A-A-TAG-AG-AGTGTACAAATATAATACAGTATAG-GGACA	ANTATAGTCAT	AG-T-GA-A-C-TR-T-GT	6899
AGI-2321	CA-G-AA-T-T-G-A-A-G-A-T-AG-AAATAGT	AGTATGGTA-G-GT	GAA-GAC-A-TR-T-G-A	6045
AG-92N3083.2	AG-GAA-C-G-C-G-GA-A-TAG-A	GGGA-G-T	T-G-C-A-C-TGT	6194
AG-92N3083.1	A-A-A-G-G-C-G-GG-A-C-GG	AATAATAT-TCCA-AT	A-G-C-A-TGT	6207
AG-IBNG	A-A-AGATGC-G-G-A-TA-G-A	AATA-GT	CAG-G-A-TA-G	6356
BF-93BR029.4	CAGCT-AA-T-C-GG-A-G-A-CAGCA-GATAATAGT	AGCAATGATATAG-GC-GT	GAA-C-G-C-A-TA-GG	6200
O-MVP5180	A-AGG-CAA-C-G-T-A-C-GT-TCA-C-G-GTAGG-TA-G-CTCA	AATGCAGTAA-GGA	CA-T-A-CT-C-A-TT-C-AG	6866
O-ANT70	A-AGG-AAA-C-G-T-A-C-GT-TCA-T-G-GA-C-GA-G-GACAAGCAAG	ACAAATAGGACCA-C-GC-A	TG-A-A-CT-C-A-TT-C-G-A	6865
SIVCPZANT	A-C-A-AAATGA-A-CA-A-GGA-C-T-AGGTGC-GG-C	AACAT-G-C-AC	T-TAC-ATGGC-C-C-CA-A-T-C-A	6516
SIVCPZGAB	AG-ACAG-TC-T-A-GTGGAG-GG-AAACG	GGA	CA-G-A-A-T-C-TR-TA-G-A	6894
ENV 00120	K R K T E Y A L F Y K L D V M F I D H	D N T	S Y T L I N C K S S T I T Q	Env

HIV1 Complete Genomes

[illegible]

HTV1 Complete Genomes

B-NEAU GTAGTGTCAACCCAAATTGCCTGTTAAATGGCAGCTTAGCAGAAGAAGACATAQTAAATTAGATCTGAATAATTTCACGGACAATGCTAAAAACATAAATAGTACAGCTGAATGTATCCATAGAAATTAATGTACAAAGACCCAAACAAATACA
B-NEAU V_V_S_T_Q_L_L_N_G_S_L_A_E_E_D_I_V_I_R_S_E_N_F_T_D_N_A_K_N_I_I_V_Q_L_N_V_S_I_B_I_M_C_T_R_P_H_N_N_T
B-NEAU S_C_O_P_N_C_C_S_M_A_V_S_O_K_X_T_S_E_L_D_L_X_L_K_T_M_L_K_T_S_S_Y_E_S_M_V_P_S_K_L_I_V_Q_D_P_T_T_I_O
B-NEAU E_S_V_N_P_I_T_A_V_X_M_O_S_E_A_R_H_B_N_S_S_S_K_F_H_G_Q_C_I_K_H_M_S_T_A_E_C_I_F_R_N_6_L_Y_K_T_O_O_O_Y
A-U455 -Q-A-T-T-C-A--A-G-A-AGG-AA-AA--A-C--TGCAAATC-TG-A-C--T-C--TT--
A-92DG037.1 ---T-C---T-----T-G-A-GG-A-G-----A-AA--TC-----A-T-C-AO-TG-AC-----C---T---T---
A-Q83 ---A---T-----G-----A-A-T-AC-----A-AA--C--TT-----A-TGCCA-C-TG-GAC--G--TC--T---A-
B-MN ---A---T-C---GG-----G-----T-T---C-C---T---A-TG-C---T---A-
B-ADB ---T-C---GG-----AGT---A---A---T---A-A-TG-----
B-NR27 ---T-A---GG-----G-C---A---C---G-AG-TG-----GG---G-T
B-RP ---T-C---GG-----T---C---C---TG-C---
B-B96 ---G---T-C---A---A---TG-T---A-A-G-TG-----C---A-A
B-BC933 ---A---T-C---C---GG-----CC-T---A---C---A-A-G-TG-----C---A-A
B-JRFL ---A---T-C-C---GG-----C---A---C---A-A-TG-----
B-JRCDF ---A---T-C-C---A-GG-T---C---T---C---A-TG-A---G-G
B-OYI ---T-C---C---GG-A---AGC---AA---T---A-AA-TG-----
B-CAM1 ---G---C-T-C-AT---A---GU---AA---C---A-GA-C-TG-----T---
B-NYS ---T-C---G-GG---G---A---T---C---A-TG-----
B-pNL43 ---A---T-C---TG---CC---A---C---CAC-TG-----
B-HXB2 ---A---T-C---GG---TC---C---CAC-TG-----
B-LAI ---A---T-C---G---GG---CC---A---CCA-TG-----
B-YUZ ---A---T-C---G---G---AA---CT---C-A-TG-T---
B-YU10 ---A---T-C---G---G---AA---CT---C-A-TG-T---
B_ACR320.2A.2.1 ---A---T-C---GG---CC---T---A---C---A-TG-T---
B_ACR320.2A.1.2 ---T-C---GG---C---A---C---A-TG-C---
B-SF2 A---T-C---GG---A---C---C---A-TG-C---C---
B-D31 ---T-C---C---GG---C---C---G-A-TG-----TT---
B-MANC ---T-C---C---GG-G-C---C---A---C---T---G-G-T---
B-HAN ---T-C---C---T---A---AG---C---A-T-A---C---T---C-A-TG-----
B-C18MBC ---A---T-C-T---GG---C---T---A---C---GCAA-TG-----C-GT
C-ETH2220 ---G-A---T-T-C-AT---T---A---GT-G-C-A---T---C-G-AA---C-TA-G-TG-----G---G-T---
C-92BR025.8 ---T-C---C-A---T---A---C-A-A---TC-CA---T---T-C-A-TG-----G---G---G---A-GGA
D-84ZR085 ---G---T-C---G-C---A---C---AA---T---T---CA-TG-----G-T---A
D-MDR ---G---T-C---G-CA---C---AA---T---C---T---T---AT---
D-Z226 ---G---T-C---G-CA---CC---AA---T---AG-TG-C---G-T-GA-T---
D-ELI ---G---T-C---G---GB-CA---C---AA---C---T---T---A-TG-A---CC-G---G-T-TA---
D-94UG114.1 ---T-C-T---G---A-A---C---AA---T---AG-TG-CC---C-T-G---T-T---
F-93BR020.1 ---G-A-C---T---T---G---G---C---C---A-T-A-T---A-C---G-C-T---A-TG-C-G---A-TG-A---C-T---G
H-90CR056.1 ---G-A---T-C---A---A-C---C-G-CA---A-A---A-T-A---A-C---T-GAC-C-AG-A-C---CA---C-G-T---T---G
AC-ZANI184 ---T-C---C-G---A-T-A---G-GA-A-G---T-C---G-C-A-C-TG-C---G-C---OG-----
AC-92BRND09 ---A---T-GC---G-A---A-T-AA---C-C---A-T-C-AGA-TG-C---T-C---T---
ADI-KAL ---Q---T-C---G-A-G---A-C---A---A---A---A-TG-AC-----G-TGA
AE-93TM253 ---A---T---A-G-A---C---C-AA---C---C---G-C-TC-A-XU-AC---C---T-T---A
AE-CH240 ---G-A---T---G-A---C---G-C-AA---C-C---G-C-T-AA-TG-----C---C---TC---
AE-90CF402 ---G-A---T---G-A---C---G-C---C---C---G-C-T-AA-TG-----C---C---TT---G-AGT
AGI-2321 ---A---T-C---C---T---A---GG-AGG---G-C---T-A---T---TGCCAAGC-TG-A-C---C---TG---T---T---
AG-92M3083.2 ---G-A---T-C-A-C-G---T---T-AG---A---A---C---GT---G---T---AATAGT---TC---T---T---
AG-92M3003.1 ---G-A---T---A-AC---T---G-A---CC---A---C---GT---T---A-A-T---G---C---T---
AG-I8W3 ---G-A---T-C---G---G-AG---G-C---A-AA---C---C---T-GC-AATC-TG-A---C---T---
BF-93BR029.4 ---G-A---T---T---A-T-A---C---C---A-T-A-T---A-G---A-T---TG-GCG---
O-MV5180 AC-AAGT-T-C-AA-AC-G---G-CA-CT-TAG---A-T-A-AG---TGGGA---A-T-A-ATCA-A-G-T-C---ACC-A-CATCT-T---A-C-G-CC-C-T---GAGGA-TTCAGAG
O-AMY70 AC-AAGT-T-GC-AA-A---G-CA-CT-TA---G-A-A-AG---G-TGG-AA-G-A-TTT-AGG-GA---T-C---GACC-A-CTCTA-A-A-C-G-CC-GA---AC-A-TAG-C
SIVCPZAM7 A-G---G-ATGG-A-A---A-C-TAGCA-AC-A-CACTTC---G-A-TGAA-GTGGAAA-AT-ATC---TGCTTGA-G-T-T-G-A-AGATTC-A-AA-T-AC---CA---T---AGA-T-GG
SIVCPZAMB ---G-TA-C-T-G-A-TA-T---A-T---T---G-A---ACTG-A-OTG-G-AAGT-AA-A---A-G-TGT-TGG---AGTA-A-G-AG-TAGTC-A---CAT---AGGA
Bnv gp120 _V_V_S_T_Q_L_L_L_N_G_S_L_A_E_E_D_I_V_I_R_S_E_N_F_T_D_N_A_K_N_I_I_V_Q_L_N_V_S_I_B_I_M_C_T_R_P_H_N_N_T

l> v3 loop start

B-WEAU	AAAAAATAATACCTTA	GGGCGAGGAGAGTCTTTATCA	ACNGGAGNA	ATAATAGGAGTATA	AGAGGAGCATTGTAACCTTAGTAGACACAGTTGGAGTAACACCTTTAAACAG	7263
B-WEAU	R K K I T L	G P G R V L Y T	T Q E	I I G D I	R R A R C N L S R T S W N N T L K Q	7292
B-WEAU	E K K S L S	D O G E Y F I	R R E	S S E I	D E H I V T L V S O V Q I T L F N R	7215
B-WEAU	K K K M M P S	T R R E T L Y	R R E	M R R Y S	T E T L S P S S N K L E S M F K T	2169
A-U455	T CG AGGTATAGTATA	T ACA C T C GT	TA	G	TG C GGGAC GA A C	6704
A-92UG037.1	G GTG CG A	ACA AC T C G	T T	G	TG C G GT CAA A G C	6838
A-Q23	GT CG A	ACA C T C G	T C	G	TG C C T A A G C G	6896
B-IN	GG CA A	C T	AA A T	AC	A A A G A G	7282
B-AD6	GT CA A	C T	A C	A	C A A T A	7237
B-WR27	G GG CA A	C T	ATAG	G C	G A T A G C AA G T G A	6643
B-RF	GT AAG	A G	C	AA	G CAA G C A	6810
B-694	G GGT T A	C T G	G A A C	A	A A A C	7257
B-BCBQ3	CTT ACG	TA	G	A	G AA G G	6785
B-JRFL	GT CA A	C T	T	A	A A A G	6560
B-JRCSP	GT CA A	C T	A	A	G CAA G	7258
B-OYT	C GG T AA	C T C	AA C	A	G CA G G A G	6810
B-CAN1	GT G A	G AC G G	ACAG	A	T AA C	7257
B-NV5	GGT G A	G AC C G	G A A	A	G AA G	7193
B-pNL43	GT CGG A CCAG	AGG C T GT	T A	A G	A A A G GC	7246
B-HXB2	G CGG A CCAG	AGA C T GT	T A	A G	A A A G	7256
B-LA1	GT CGG A CCAG	AGG C T GT	T A	A G	A A A G GC	6849
B-YU2	GT A A	C T G	A	A	A CAA G A	7241
B-YU10	GT A A	C T G	A	A	A CAA G A	7243
B-ACH320.2A.2.1	GGT CA A	A C T G	A C	A	G AA	7266
B-ACH320.2A.1.2	GGT CA A	CCT G	G A A	A	G CAA	7282
B-SF2	GT CTA A	C T C	AG	AA	G CAA G	7270
B-D31	C GG CG A	G A C T	A A	A	G G AA G G	6798
B-NANC	GT TACA	AG T C GT	A CC	G C	A G AA G G A C	7210
B-HAN	GGT CA A	C G G	AG	G	T A A G A C C A	6724
B-C18MBC	GT G A G	C T	A	T	A A G AC G G	7110
C-ETH2220	G GT GGA	ACA AC T C G	C	A	A GA GA AA A C C A	6856
C-92BR025.8	GT GGA	ACA C T C G	C	A	A GCA C A C G	6806
D-842R085	C GG C C A	A CA CG C	A GT T	C C A GATA	G A T TA G GT AA C CG	6795
D-NDK	CC GG C T A	T A CA TC C	ATA	A AGGAG A C T C	G A T AA C G GAA AG C	6781
D-Z226	C GG C T A	TT CA CG C	A	C G A TATA	G A T TA ATGAA G A C	7255
D-BL1	C G C C A	T CA TC C	T AG	TC G TC ATA	G A T TA G CAA G A C A	6798
D-94UG114.1	C GT C CG A	CA C C T	A	G	A C G G A A C	6580
F-93BR020.1	G T	AC C T C G	C C	ABG	G G CAA GG G GC A	6593
H-90CR056.1	C GT CA	G AC C T C G	T C	C	TA GAC G C C	6584
AC-ZAM184	G GT CGG T	ACA CCT C	AT	C	G TA C ACGA AA GC C A	6593
AC-93RW009	GTG CA A	ACA C T G	T C	G G	CTG C A G AA GA C A	6597
AD1-HAL	G GGG CA C	C CA C G	G	G	CTA A GA GAA G A C	6839
AE-01TR253	C GG C T A	A AC C G	C	C A	G AA C T G GA A G AA AGT T	7255
AE-CN240	C GT A	AC T C G	T	AA	T G AA A G AA AGT	6833
AE-90CP402	T GTGC GGA	AC T CC	A C	AT T	AA T G AA A A AG A	7319
AGI-2321	GT CA A	AC C C C	GA T C	AG	C TG C A GAA G G A TCCA	6465
AG-92NG083.2	GT C AA C	ACA CGT C G	T T	A	TG T AA GGG G TG GA T	6614
AG-92NG003.1	GT GAA C	T ACA CGT C G	T	G A A	TG C A CAA C GG G TG C GA	6594
AG-1BMQ	GGTG CA A	CA C T C G	T C	G	TG C A CAA C C T	6776
EF-92BR029.4	GT C AA	AC C T	C C	AAG	G G AA G G G A	6620
O-MVP5180	GT C G T TA AC	T AT TGG GCGG TG	CTTA	G GTAACA C TCACCAAGATCA	GGT TT CATA A AG TGTA G A TG CC C A	7300
O-AM770	T C G G G GAA	T AT GCGTGACAGC TGGGA T	G G	C GC A C GC	GGC TT C GTA A GCC TGA A TA	7292
STVCPZANT	GT G TC CAAA	A TGACTT C AC	GT A AT	GC C C CT	GAA GTTC CAG C A AG GCTA G AC AG ACGT CA A	6936
STVCPZGAB	GGG G GG GCGGA	G A TGACCT AT	T A A T	G G CC	GA A A G G C CA CGA GG GG G A	7314
Env gp120	R K K I T L	G P G R V L Y T	T G E	I I G D I	R R A R C N L S R T S W N N T L K Q	Env

V3 loop tip

V3 loop end <-1

B-WEAU	ATAGTT	GAAGAAATGAGAGAAATA	AAACAAATTTAAGAAT	AAACAAATAGTTTITTAACAAATCCCTCAGGAGGGGACCCAGAAATTGTATATGCAC	AGTITTAATTGTGAGGCGGAATTTTCTAC	7386	
B-WEAU	I V	E K L R E I	K Q F K N	K T I V F K Q S	S G G D P E I V N M	S F N C G G E F F Y	7387
B-WEAU	S	K L S E K I	F M N L R	K Q S F L N M P O E G T	P Q K L S C	V L I V E G N F S Y	7388
B-WEAU	D S	I K K R N S	I T I S E	F N N S F S T I L R R G P R N C N A	I F S L V R G I F L		7389
A-U455	G-C	C-AGA	A-G	A-C-GCTAGC	TAT	AC-CA-T	6821
A-920G037.1	G-C	C-A	T-T-GC-C-C	T-T-A-C-TAGC	TTT	AC-CA-T	6755
A-Q23	G-C	C-AC	T-C-GG-C	A-A-GCTA-C	T-T-G-CAC-CA-T	A	6813
B-MN		AGC-A		C-T			7399
B-ADB	C	AC-A	GG-AAT	C-T			7357
B-WR27	C	CC-A	T	C-T-C	T-G		6766
B-RF	G	AC	G-C	C-GTC		C-T	6927
B-896		AT	A-G	CC-T			7374
B-BC003		AT	G	C-T		T	6902
B-JRFL		AT	G	C-T-C			6677
B-JRCSF			T	C-CT-C	T		7375
B-OYI	C	AC-A	G	CC-G-T-G	T		6927
B-CAM1	C	AC-A	GG	A-C-T	G		7374
B-WYS		AC-A	G	C-T			7310
B-pNL43	C	AGC	GGA-AAT	A-C-G		C	7366
B-MXB2	C	AGC	GGA-AAT	A-C-G		C	7376
B-LAI	C	AGC	GGA-AAT	A-C-G		C	6969
B-YU2	C	AT-A	GG-AAT	A-C-T-C		CA	7361
B-YU10	C	AT-A	GG-AAT	A-C-G-C		CA	7363
B-ACH320.2A.2.1		AGC-A	GG	C-GT-C-T			7383
B-ACH320.2A.1.2		AT	C-GG	AAA-T	G-T		7399
B-SF2		A	G-GG-AAT	C-T		A	7390
B-D31		A	G-GG-C	C-T		C	6915
B-MARC			A-GGA-C	A-C-T	CT-T		7327
B-HAN	T	AG	T-G	A-T-G	A	C	6847
B-C18NBC		AG		C-G-C		CA-GTTTTC	7233
C-ETHC220	G-AAG	CA-A-G	C-CCCT	AA-G-C	T	AC-CA-T	6773
C-92BR025.B	G-G	GC-G	C-CCCT	G-AAA-GC-A-GCA	A-T	AC-CA-T	6723
D-842R085	G-C	AG-G-A-C	TT-G-CC-A	C-A-A-C	T	AC-CA	6912
D-NDK	G-C	AC-C-G-A-C	TTT-C-A	C-AC-G-C		AC-TCA	6898
D-2226	G-C	AT-G-A-C	TTT-C-A	C-A-A-C		AC-CA	7372
D-BLI	G-C	AG-G-ACC	TTT-C-A	C-T-AGG-C		AC-CA	6915
D-94UG114.2	G-C	G-A-C	TTT-CC-G	C-A-C-G		AC-CA	6697
F-93BR020.1	G-AAG	C-G-G-GTCT	T-T-CCCT	GC-AAA-CTC	T	AC-G-T	6710
H-90CR056.1	G	AC-C-G-AT	C-G-C	G-AGC-G-C-AA	ATG-G	AG-CA-T	6701
AC-2AM184	G-C	A-C-CA-A-C	T-C-CCCT	AGC-GCTA-C-T	T	AC-CA-T	6710
AC-92R009	G-CA	TC-C	T-C-G-CATT	C-A-GA-C	G-TT	AC-CA-T	6717
ADI-MAL	G-C	T-C-G-AGC	TTT-C-A	C-A-A-TTC		AC-CA	6956
AE-93TH253	G-AC	A-G	C-T	GG-AC-AGC-G-G-C-A	A-T-T	AC-T	7372
AE-CH040	G-AC	A-G	C-T-G	A-C-C-C	A-T-T	AC-T	6947
AE-90CF402	G-ACA	AG-G	CC-TGGGACA	TG-AGC-CG-C	A-T	CAC-T	7439
AGI-8321	G-C	C-C-A	C-GTA-ACA	AGA-C-GAT-A-C-GTA-C	TGT	AC-CA-T	6598
AG-92NG083.2	G-CACA	C-C-GC-GA	ATC-A-T	G-AC-ACC-CTC-TG	T	AC-CA-T	6731
AG-92NG003.1	G-CAG	C-C-C-GA-C-G	GTC	GT-ACC-CTC-G	T	AC-CA-T	6708
AG-IBNG	G	AC-C-AGACG	T-C	CC-A-C-GCTA-CC-T	TGT	AC-CA-T	6893
BF-93BR029.4	G-AGG	C-G-AGCTT	T-CCCT	GC-AAA-CTC	T	AC-T	6737
O-NUP5160	C-C	AT-GG-ATT-A-TC-TGTA-C	ACAG	GTT-C-A-A-C-GCAG-A-TAGT-T-A-TG	G-AAGCATTTA	CA-C-CAT-A-C-T-T	7426
O-AMT70	C-C	GG-ATT	T-G-A-C	ACAGTATG-TT-AC-GACA-C-T-CAG-AGC-T-A-T-T	GG-AACGATTTA	CAC-C-CAT-A-C-T-T	7428
SIVCPZANT	C	AGCAGCTTCTT-COG-GCATT-GA	A-G-AG-C-C	AAAACC-TG-G-A-ACAA-ATGGAC-T-CA-AT-A-T-T	G-AAA-G-T	T-G-CCA-A-C-T-T	7068
SIVCPZQAB	G-AAA	GCTTT-C-CC	TC-CA-TGGNACA	GCAACC-AC-ACAC-C-CAG-G-T-A-T	GG-CAC-CAT-T	TG	7437
Env gp120	I V	E K L R E I	K Q F K N	K T I V F K Q S	S G G D P E I V N M	S F N C G G E F F Y	Env

B-WEAU	TGTAATTCACACACCTGTTTAACTACTTGG.....CATGCTAAT.....GGTACTTGGAAAG...AATACCTGAAGGGGACAGAT.....AACAATATCACACTCCCATGCGAATTAACAATTTATTAACAGG	7503
B-WEAU	C N S T Q L F M S T W.....M A N.....G T W K.....N T R G A D.....N N I T L P C R I R Q I I N R	2372
B-WEAU	V T O M S C L I V L S.....M L S.....V L G S.....I L K G Q J.....T I S H S R A B S N K L S T G	2281
B-WEAU	L S F M T A V S S T L S.....C S I.....S Y L E S.....Y S R G R F.....O V M T P M O N K T N V K O	2233
A-U455	-C-CA-T-GGC-----C-T-----AA-GGCA-C-T-TCA-----GGCATG-----C-A-----GG-C-A-T-----A-----T-----T-T-----	6932
A-92UG037.1	-A-T-GGC-----GTA-----A-----C-ACC-GC-----GCAAG-----T-A-----GG-C-A-T-----G-----T-T-----	6883
A-Q23	-A-T-CCCT-----C-----T-----T-A-A-C-CTTG-----GACACT-ACAGCAC-CAGGAGTCAATG-----C-A-T-----G-----T-T-----	6939
B-INI	-A-T-CA-----A-----A-----AA-----T-----AC-----T-A-----A-----T-----T-----	7516
B-AD6	-A-----A-----AATTTA-----G-C-TGG-----AA-TTACAC-ATCG-----GG-ACT-AA-G-A-----G-C-----T-----T-----T-----	7486
B-WR27	-A-----A-----A-----A-----GAAGGA-----A-----GAAGG-----G-ACTT-----T-----A-T-----G-----T-----T-----	6877
B-BF	-A-----A-----A-----A-----GAAGG-TCA-----ACACT-AG-A-----G-----CA-----A-----G-----T-----T-----	7041
B-896	-A-G-----A-----A-----T-----G-AGG-ACA-----GGCACT-AA-G-A-----G-----TA-----A-----T-----T-----	7488
B-BCSG3	-G-----A-----A-----GCTGGGA-AA-C-TGGAAT.....AGTA-G-GAA-G-TCAQ-----GACACT-A-G-A-----G-----T-----T-----	7031
B-JRFZ	-A-----A-----A-----A-AA-----AC-GAAG-----TCAAT-C-----AATAC-----T-----T-----	6791
B-JRCSF	-A-----A-----A-----A-----A-----A-----GAA-----TCA-G-GGCACT-AA-G-A-----G-CC-----T-----T-----T-----	7489
B-OYI	-A-T-----A-----A-----A-----A-----ACA-G-GCA-----GCACT-AA-T-AC-----G-----T-----T-----T-----	7035
B-CAN1	-A-----A-----C-----CTGTTA-----G-C-TGGAAT.....A-----GAAGG-TTA-----ACACT-AAAG-A-----T-----T-----T-----	7500
B-MYS	-GA-----A-----A-----CTGTTA-----AG-C-TGGAAT.....A-----GAA-G-TCAQ-----ACA-T-AA-----AC-----T-----T-----	7433
B-DNL43	-A-----A-----A-----TTTA-AG-C-TGG-----A-----GAAGG-TCA-----ACACT-AA-G-AG-----G-CA-----T-----T-----	7493
B-HXB2	-A-----A-----A-----TTTA-AG-C-TGG-----A-----GAAGG-TCA-----ACACT-AA-G-AG-----G-CA-----C-----T-----T-----	7502
B-LAI	-A-----A-----A-----TTTA-AG-C-TGG-----A-----GAAGG-TCA-----ACACT-AA-G-AG-----G-CA-----T-----T-----T-----	7095
B-YUZ	-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7465
B-YU10	-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7500
B-ACH320.2A.2.1	-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7500
B-ACH320.2A.1.2	-G-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7510
B-SF2	-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7501
B-D31	-G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7023
B-MANC	-AT-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7450
B-HAN	-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	6958
B-CL8MBC	-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7362
C-BTN2220	-C-A-T-A-C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----	6878
C-92BR025.8	-C-A-T-AGC-----A-ACACACCTA-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----	6840
D-842R095	-C-A-T-GGA-----G-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----	7023
D-HDK	-A-T-AGA-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----	7006
D-ZZ26	-A-T-GGA-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----	7495
D-BLI	-A-T-GGA-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----	7041
D-94UG114.1	-C-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	6817
F-93BR020.1	-A-GATG-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	6809
H-90CR056.1	-C-A-T-GG-----G-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----	6821
AC-ZAH184	-C-A-T-A-C-----A-AC-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----	6818
AC-92BR009	-A-T-GGCT-----C-C-AGT-----A-AGA-----C-----C-----C-----TCA-----GGCAC-AAIT-A-----G-----T-----T-----	6934
ADI-MAL	-A-T-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7076
AE-93TH253	-C-A-----GA-----A-----C-----ATA-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7480
AE-CM240	-C-A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7055
AE-90CF402	-C-A-----GCA-----A-----ATAAA-GGA-----ACC-TGCA-G-GTT-----GGCAC-AACT-----G-----T-----T-----T-----	7556
AGI-Z321	-C-A-T-GGC-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----	6702
AG-92HG083.2	-C-A-T-GGA-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	6824
AG-92HQ003.1	-A-T-GGAT-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	6795
AG-IBNG	-C-A-T-A-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7010
BP-93BR029.4	-C-A-T-GGA-----A-----A-----A-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	6821
O-MVP5190	-CA-T-TGG-A-----CTA-----TTATCAACTG-A-A-STCCGGATGCCAG-AG-TCAAAAG-AOC-----GAGACCAATTA-A-----G-----T-----T-----	7559
O-ANT70	-CA-G-TA-----TA-----C-TT-----TCATG-AACGGA-----ACC-C-----T-GTGTG-A-A-TTA-TCA-G-----G-----T-----T-----T-----	7541
SIUCPCAMT	-G-AT-----CTTG-----C-----C-CA-----TAC-----GG-A-----G-----CTC-TCAAC-AAC-G-----C-----T-----T-----T-----	7167
SIUCPCQAB	-A-T-GC-AA-A-----C-----G-C-----A-----A-----A-----A-----A-----A-----T-----T-----T-----T-----T-----	7521
Env gp120	C N S T Q L F M S T W.....M A N.....G T W K.....N T R G A D.....N N I T L P C R I R Q I I N R	Env

HIV1 Complete Genomes

B-NEAU	TGGCAGGAAGTAGGAAAAGCAATGTATGCCCCACCCATCGAAGGACAAATTAGATGTTTATCAAAATATTACAGGATGOTGGTAGTACT.....GAA...GAGAACCCAGACCGAGATCTTCAGACCTGGAGGA	7444
B-NEAU	W Q E V G K A M Y A P P I E G O I R C L S N I T G L L L L T R D G G S S . . . E . E N O T E I F R P G G	2419
B-NEAU	G R K S E K Q C M P H P S K D K L D V Y Q I L O G Y V S Q E M V V V . . . I I I R T R P R S E D L R E	2324
B-NEAU	V A G S R K E S N V C P T H R R T N S M P I R Y V R V T T N K R Y W S S F . . . I I I E P D R D L O T W R	2275
A-0465	---AG---C---T---C---GT---A---G---GA---C---AC---C---AC---A---T---CRA-A-AT---C---T---	7070
A-92UG037.1	---AG---C---T---C---GT---A---AG---GA---C---AC---A---GGT---A---AGC---GTG-C-GT---A---C---	7001
A-Q23	---AG---C---C---T---CC---GT---A---AG---GA---C---AC---A---G---A---AGGA---A---C---TGT-A-AT---A---C---	7077
B-MN	---T---T---T---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7660
B-AD8	---A---T---AG---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7627
B-WR27	---G---T---T---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7015
B-RP	---T---AGT---A---A---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7182
B-896	---A---T---AC---	7629
B-BCSG3	---T---AGC---T---AC---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7172
B-JRFL	---T---AG---T---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	6929
B-JRCS7	---G---T---T---TA---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7627
B-OY1	---T---AGT---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7176
B-CAM1	---T---AG---AC---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7638
B-MY5	---T---AGT---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7574
B-DNL43	---T---AGT---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7630
B-MXB2	---A---T---AGT---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7640
B-LAI	---T---AGC---	7233
B-YU2	---T---AG---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7601
B-YU10	---T---AG---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7403
B-ACH320.2A.2.1	---G---T---AGT---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7644
B-ACH320.2A.1.2	---G---T---AGT---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7654
B-SF2	---T---T---G---T---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7642
B-D31	---T---AGT---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7146
B-HANC	---T---AGC---	7597
B-HAN	---T---T---G---T---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7099
B-C180BC	---T---A---A---AC---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7512
C-ETN2220	---G---AG---CG---T---T---AT---A---TG---AG---C---AC---C---C---C---C---C---C---C---C---C---	7019
C-92BR025.8	---GG---C---CG---T---T---AT---T---A---C---AG---C---A---C---C---C---C---C---C---C---C---C---	6978
D-842R085	---T---T---G---AT---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7161
D-MUK	---AG---T---T---A---A---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7144
D-Z224	---G---T---T---AT---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---C---	7633
D-BLE	...GT---C---GCA---G---A---T---T---A---A---C---CP---C---A---T---C---C---C---C---C---C---C---	7176
D-94UG114.1	---G---CG---T---T---TTC---C---AC---C---AC---T---G---G---GCA---T---A---T---AGT---G---AT---C---	6958
P-93BR020.1	---AG---CG---T---T---AAT---T---C---A---C---CC---AAC---TC---G---G---CTG---A---CTACT---AT---C---	6944
H-90CR056.1	---AG---CG---T---T---A---C---TG---G---AC---A---G---TT---C---AG---ACCG---TCT---GCAG---A---ATT---T---	6959
AC-2AM184	---AG---CG---T---T---T---C---GT---A---AG---AC---C---AA---A---GT---G---AC---GGGACAGT---ACC---TGG---T---	6965
AC-92RNG09	---AGGAC---C---C---T---C---GT---A---C---G---C---AC---A---A---A---T---A---T---CTACR---T---A---C---	6972
ADI-MAL	---A---AC---T---T---C---GTC---C---AC---C---CA---A---A---A---AGTGACA---T---GTG---C---AT---C---A---	7220
AE-93TH253	---G---C---AAT---T---AGT---A---C---AT---G---AA---G---G---G---G---A---TGGTAC---TA---TGAG---C---	7624
AE-CH240	---G---C---T---T---AGT---AG---AT---G---AA---G---G---GT---A---A---T---G---G---G---T---A---C---	7193
AE-90CF402	---AG---CG---T---T---AGT---A---GT---AT---G---AA---G---G---T---A---CAA---A---TCAGACT---A---A---AT---	7703
AGI-2321	---AMG---C---T---T---A---GT---A---AG---GA---C---AA---A---A---TGG---A---C---CAGCA---AT---C---	6840
AG-92NG083.2	---A---G---C---TT---C---A---CC---GT---AA---C---A---A---G---A---AC---A---T---C---GTACAGAG---C---	6965
AG-92NG003.1	---AG---G---C---T---C---G---T---C---AG---C---C---C---T---G---A---A---A---T---CTGA---AT---C---	6933
AG-IBMO	---A---T---C---AT---A---GAT---C---AC---C---A---G---A---A---A---G---A---A---GTACA---AT---A---C---	7145
BF-93BR029.4	---G---CGG---G---T---T---C---A---C---CC---AGC---TC---G---G---CAG---A---A---T---C---GAC---GAG---	6959
O-XVP5150	---AT---A---G---AGTCGAG---C---A---T---CCC---CA---CT---A---C---CAT---C---C---A---T---AA---GA---TC---CAGTT---	7696
G-ANT70	---ATAGG---G---CAGTCG---G---C---C---A---T---A---TA---TC---A---C---A---G---A---T---AA---GA---CC---CA---TG---	7682
SIVCPZANT	---GGCAT---TTC---GC---T---CTTAG---C---T---GGAG---GA---TG---TCC---ACT---C---CC---A---T---AA---TA---G---G---	7296
SIVCP2QAB	---AT---AGG---C---G---A---A---C---A---AG---A---C---C---CT---CAAT---T---C---T---AC---T---C---TTC---CAC---CC---GTG---	7662
Env gp120	N Q E V G R A M Y A P P I E G O I R C L S N I T G L L L L T R D G G S S . . . B . B R O T R I F R P G G Env	

Rev Responsive Element, RRE
Stem I

Strain	Sequence	Position
B-WEAU	GGAAATATGAAGGTAATTCGAGAGTGAATTATATAAATATAAGTAGTAAATTTGACCATAGGAGTAGCACCCACC	7779
B-WEAU		2464
B-WEAU		2361
B-WEAU		2313
A-T455		7202
A-92UG037.1		7133
A-Q23		7209
B-NR		7789
B-ADR		7762
B-WR27		7150
B-RF		7317
B-896		7761
B-BGSG3		7304
B-JRFL		7061
B-JRCSF		7759
B-DYI		7311
B-CAM1		7773
B-MY5		7709
B-pML43		7762
B-KGB2		7772
B-LAI		7365
B-YU2		7733
B-YU10		7732
B-ACH120.2A.2.1		7776
B-ACH120.2A.1.3		7786
B-SF2		7777
B-D31		7299
B-NANC		7732
B-HAN		7234
B-C18MBC		7644
C-BTH2220		7148
C-92BR025.8		7110
D-84ZR085		7293
D-MDK		7276
D-2226		7765
D-ELI		7308
D-94UG114.1		7090
F-93BR020.1		7076
H-90CR056.1		7091
AC-2AH184		7097
AC-92RM009		7104
ADI-MAL		7352
AE-93TH253		7756
AE-CA140		7325
AE-90CP402		7835
AGT-2321		6972
AG-92BG083.2		7097
AG-92BG003.1		7065
AG-1885		7277
BP-93BR039.4		7091
D-MVP5180		7843
D-ANT70		7826
SIWCFEANT		7434
SIWCFEAS		7809
Env gp120		

Env
cleaved by cellular protease, most likely furin

RRE Structure

after B. Charpentier,
J. Mol. Evol., 266:950-62 (1997)

HTV1 Complete Genomes

Rev Responsive Element, RRE, continues

The symbol ∇ indicates A, C and G rejected by Ray binding in an RNA footprinting assay (Chamberlin (1997)).

Stem I

Stem II

Stem 11

Section II

Step II

Skem II

Stem IIA

51

II-IV

Stem III-IV:

B-WEAU
 B-WEAU
 B-WEAU
 B-WEAU
 A-U455
 A-92G037.1
 A-Q23
 B-NH
 B-ADB
 B-WR27
 B-RF
 B-896
 B-BCSG3
 B-JRFL
 B-JRC5P
 B-OY1
 B-CAM1
 B-MV5
 B-DNL43
 B-BX2
 B-LAI
 B-Y02
 B-Y010
 B-ACH320.2A.2.1
 B-ACH320.2A.1.2
 B-SF7
 B-D31
 B-MANC
 B-HAN
 B-C18MBC
 C-ETH2220
 C-92BR025.6
 D-842R085
 D-NDK
 D-Z226
 D-ELT
 D-94UG114.1
 F-93BRQ20.1
 H-90CR056.1
 AC-ZAM184
 AC-92RW009
 ADI-NAL
 AB-93TH253
 AB-CN240
 AB-90CP402
 AGI-2321
 AG-92MG081.2
 AG-92MG003.1
 AG-18WG
 BP-93BR029.4
 O-MVF5180
 O-AM770
 STVCP22ANT
 STVCP22AB
 Env gp120

RRE Structure

after B. Charpentier,
J. Mol. Evol., 266:950–62 (1997)

	Stem III-IV	Stem V	Stem V'	Stem I'	End RRE	
B-WEAU	GA	CT	CAC	AGT	CTGGGGCATCAAAACAGCTCCAGGCAAGAGTCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8079
B-WEAU	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	2564
B-WEAU	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	2495
B-WEAU	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	2410
A-U455	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7502
A-92UG037.1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7433
A-Q23	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7509
B-MN	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8089
B-ADU	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8062
B-MR27	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7450
B-RP	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7617
B-896	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8061
B-BC963	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7604
B-JRFL	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7361
B-JRCSE	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8059
B-OV1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7611
B-CAM1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8073
B-NV5	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8009
B-DNL43	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8062
B-HXB2	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8072
B-LA1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7665
B-YU2	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8033
B-YU10	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8035
B-ACB320.2A.2.1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8076
B-ACB320.2A.1.2	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8086
B-SF2	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8077
B-D31	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7599
B-MANC	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8032
B-HAN	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7534
B-C18NHC	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7944
C-BTH2220	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7448
C-92BR025.8	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7410
C-84BR065	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7593
C-MDK	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7576
C-2224	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8065
C-BL1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7608
C-B4UG114.1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7390
C-B3BR020.1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7376
C-90CR056.1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7391
AC-2AM184	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7397
AC-92RM009	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7404
ADI-NAL	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7652
AE-93TH253	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8056
AS-CN240	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7625
AE-90CF402	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8135
AGI-2321	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7272
AG-92NG083.2	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7397
AG-92NG003.1	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7365
AG-18NG	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7577
BF-93BR029.4	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7391
C-MVPS180	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8143
C-ANT70	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8126
STVCP2ANT	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	7734
STVCP2GAB	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	8109
Env gp41	CT	AG	CT	CA	CTGGGGCATTTGGGGTTCCTCTGGAAACTCATCTGCACCACTACCTGCTCTGGAAATGCTAGTGGAGTAAT	

RRE Structure

after B. Charpentier,
J. Mol. Evol., 266:950-62 (1997)

HIV1 Complete Genomes

B-NEAD	AGATCTCAGGATTAC	ATTGGAATTAACATGACCTGGAGTGGGAGAGAGAAATTAACAATTACACAGGCTTAATATACAACTTAATTTGAAGAAATCGCAAGAACCAACAGAAAAAATGAC	8202
B-NEAD	R S Q D Y	I W N N N T W M E W E R E I N N Y T G L I Y N L I E E S O M O O R K N E	2605
B-NEAD	D L R I F	F G I T S P G W S Q R E X L T I T Q A S V T T S L K M P R T N X K K M S	2492
B-NEAD	I S G L I	F I E S H D I D G V G E R N S Q L H R L N I O L N S R I A B P T R K K S	2445
A-U455	A	A	7625
A-92UG037.1	A	A	7556
A-Q23	A	A	7633
B-HN	A	A	8212
B-AD8	A	A	8185
B-WR27	A	A	7573
B-RP	A	A	7740
B-B96	A	A	8184
B-BCSG3	A	A	7727
B-JRFL	A	A	7484
B-JRCSP	A	A	8162
B-OYI	A	A	7734
B-CAM1	A	A	8196
B-NYS	A	A	8132
B-DNL43	A	A	8185
B-HXB2	A	A	8195
B-LAI	A	A	7768
B-YU2	A	A	8156
B-YU10	A	A	8150
B-ACH320.2A.2.1	A	A	8199
B-ACH320.2A.1.2	A	A	8209
B-SF2	A	A	8200
B-D31	A	A	7722
B-HAMC	A	A	8155
B-HAM	A	A	7657
B-C18MC	A	A	8067
C-STH2220	A	A	7571
C-92BR025.8	A	A	7533
D-84ZR085	A	A	7716
D-NDK	A	A	7699
D-2226	A	A	8188
D-ZLT	A	A	7731
D-94UG114.1	A	A	7513
F-93BR020.1	A	A	7499
H-90CR056.1	A	A	7514
AC-2AM184	A	A	7520
AC-92RM009	A	A	7527
ADI-MAL	A	A	7775
AE-93TH253	A	A	8179
AE-CM240	A	A	7748
AE-90CF402	A	A	8258
AGI-2321	A	A	7395
AG-92NG003.2	A	A	7520
AG-92NG003.1	A	A	7488
AG-IBNG	A	A	7700
BP-93BR029.4	A	A	7514
O-MVP5180	A	A	8269
O-ANT70	A	A	8243
SIUCP2ANT	A	A	7884
SIUCP2GAB	A	A	8235
Env gp41	R S Q D Y	I W N N N T W M E W E R E I N N Y T G L I Y N L I E E S O M O O R K N E	Env

HFV1 Complete Genomes

[illegible]

HIV1 Complete Genomes

	Tat Rev intron end (Tat Rev exon start (Tat Exame 2, Rev Exame 3)	Tat CDS end (Exame 2) <--	
B-WEAU	GTAGGCGGGGATCTACACCATTTGTCATTCAGACCCACCTCCAGCCCGAGGGGA...	CCCGACAGGCGCGAGGGAATCGAAGAGAAAGCTCGAGAGAGACAGAGACAGATCCGGAAGATTAGTGGATTCTTAACTACTATC	B499
B-WEAU	V R Q G Y S P L S F O T H L P A P R G	P D R P E G I E E E G G E R D R D R S G A L V D G F L T L I	2704
B-WEAU	L G R D T E K C H F R P T S Q F R G F	P T G P K E S K K K V E R E T E T D P R D S W M D S S F L S	2577
B-WEAU	S S A Q T L T I V T S D P P F S F R G Y	I R Q A R R M R R R R R M R E R Q R Q E R T S G W I L N T Y	2539
A-0455	TC G TTGC AT AGA T	T G A G A CA G A GATC C AGC G GA GC	7922
A-92DG037.1	CC A G TAC GAA A	T G A CA G GAT C C AGC G GC	7853
A-Q23	T TAC AA A T	T G A CA G GATC AGC G GC	7929
B-MN	T G G G C C TT	T G A CA G GATC AGC G GC	8509
B-AD9		T G C AGC G GC	8482
B-NR27	R A T TT	AATC C C A G C	7870
B-RF		G C G TGC A G	8037
B-996	T A G T	CA TCC A C G T	8481
B-BCSG3	G A	TC A C G T	8024
B-JRFL	A G T	C A C G	7781
B-JRCBF	C A G T AA	CA A C G	8479
B-OVI	A G C A A	G G	8031
B-CAH1	G T TT A	G C C AAC G C	8453
B-NY5	A G C G	CC A C G C	8429
B-pNL43	T A G AT	ATC A C C G	8482
B-MXB2	T A G A	ATC A C C GG	8492
B-LAI	T A G A	ATTC A C C G	8085
B-YU2	A G T A	C C TCC G A	8453
B-YU10	A G T A	C C TCC G A	8455
B-ACH320.2A.2.1	A G G	TC A C G	8496
B-ACH320.2A.1.2	A G G	C C TCC G	8506
B-SF2	A G G T	C C TCC G	8497
B-D31	A G	C C AATC A A G	8019
B-MANC	TC A	G A TC C C G	8452
B-HAN	G T AA C	G TC AGC G T	7954
B-C18MC	A G A A	G C C ATC C C G	8364
C-ETH2220	A G TTA CA	T G A CA G ATC C C G A T T	7868
C-92BH025.8	C A G TTAC AA	A T G A CA G ATTC AGC G CCG	7830
D-842R085	G T A G T	A A A CA G GT C ATTC G A C C G	8013
D-NDX	TC G T	A A A CA G ATTC G A C TG T	7996
D-Z286	T TC	C G ATTC G A C C G	8485
D-ELI	TC G T	TC CA C CCG	8028
D-94UG114.1	TC T	A A ATTC G A C C G	7810
Tat stop codon TGA for many strains			
with TGG in place of TAG at WEAU stop			
P-93BRG20.1	A T TA AG A	G CA G A ACC G T GC	7796
R-90CR056.1	T T TTG GAA AC A	TG A C G G	7811
AC-2AM084	T T CTAC CA AGA T	T G A CA G GATC AGC G GC	7817
AC-92RH009	A G TTA AA	T G A CA G ATTC AGC G GC	7824
ADT-MAL	TC G G T A A CCA	TC A G G ATTC G A C C G	8075
AE-93TH253	T T C CTAC ATCAT A A GGA	T C A G TCC AGC G GC	8479
AE-CN240	T T C CTTC ATCAT A A A	A T C A G ATTC AGC G GC	8045
AE-90CF402	T G C TTAC ACCAG A A	A A CA G AT A G AGC G GC	8555
AGI-Z321	T T C TTAC ACCA A A	T T C CA G ATTC C G AGC G GC	7692
AG-92NG083.2	T G C TTAC ACCA A C A	T G AA C TAC AGC G G GC	7817
AG-92NG003.1	T G C TTAC ACCA A A	A A G CA A TCC C AGC G GC	7785
AG-IBNG	T G C TTAC ACCA A A	A A G CA A TCC AGC G GC	7997
BP-93BR029.4	A A G T AG A A	CC G A TCC A C G T G	7811
G-MVF5160	A TCA CC C G G T CTG CA G CA A	G A A C A G A CA G AG GGC AG GGC ACCA CCA AGT G	8566
G-ANT70	A TCA CC C G G T C AA ATCA AGA A	G A GA C A G A CA G G G AGA G GGC ACC C A G GC GT G	8540
SIVCP2GAB	A TCAT TC C G G T CTAC A AA A CA AT	A A GCA A A AAG AGA A GAT G GGC GGC GCA C C G GC CT G	8181
SIVCP2GAB	C TC C G G T TTA T A A	AA GA T G A C G G C CA AG A TG A G G GC	8532
Bnv gp41	V R Q G Y S P L S F O T H L P A P R G	P D R P E G I E E E G G E R D R D R S G A L V D G F L T L I	Env
	Tat CDS intron/exon	Tat CDS end	Rev

ENV
R

I-81
DEC 97

HTV 1 Complete Genomes

Rev CDS end (frame 3) <--

B-MEAM ...CTCTGCGATATTGGAGCCAGGAAGTAAGATAGTCTGTTAGTTGCTTAAATGCCACAGCTATATACAGTAGCTGAGGGGACAGATAGGGTTATAGAAATAGTGCAAAAGAACTTGTAGAGCTATTCTCCACATAGCTAGAAGAATA 8772

B-WEAM ...L L Q Y W S O E L K N S A V S L L L N A T A I A V A E G T D R V I E I V Q R T C R A I L H T I P R R I 2795

B-MEAM ...# S C S I G A R N S R I V L V C L M P O L S O S L R G Q I G I S K S C K R E L V E L F S T Y L E S S 2657

B-MEAM ...t F P A V L E R Q T K R E S C C S F A S C M S Y S S S S G D R S G Y R M S A K N L S S V S P H V S K N 2620

A-U455 ...-T-T-T-G-T-G--A-T--A-C--G-C-TGT-AG--G-T-T--GA--C-T-G--A--C-- 8216

A-921KG037.1 ...-TT--G-T-G--A-T--A-A--G-A-T-A--GCT--C-A--GCT-G--A--C-- 8147

A-Q23 ...-TCA--G-T-G--A-T--A-A-C-AG--G-A-T-A--GCT--CA--T-G--C--GT--C-- 8223

B-MN ...-A--T--G--G--C--C--G-A-T--A--G--C--O-G--C--C-- 8779

B-ADB ...-T--T--G--T--C--C--T--A--Q--A--Q--C--C-- 8755

B-WR27 ...-A--G-T--G--G--A-C-C--G-C--A-C--G-CA--T-T--C-T--C--C-- 8143

B-RP ...-T--G--C--C--A--A--A--T--A--G-CA--T-CT--T--T--C-- 8310

B-B96 ...-A--C--C--C--C--A--A--G--A--G--G--A--C--C-- 8754

B-BCSG3 ...-A--AC--C--T--C--A--G--A--A--G--A--A-- 8297

B-JRFL ...-A--T--C--C--C--A--GC-T-A--A--C-- 8054

B-JRCSF ...-A--T--C--C--A--G--A--G--A--G--C-- 8752

B-OYI ...-TCA--C--C--C--A--G--A--T--A--T--C-- 8304

B-CANL ...-T--G--T--G--A--T--A--G--G-- 8766

B-WYS ...-G--T--A--C--T--G--G--C--T--C-- 8702

B-pHL43 ...-T--T--AC--C--C--G--G--T-A--GC-G--A--G-- 8755

B-HXB2 ...-A--T--C--C--C--G--A--G--G--G-- 8765

B-LAI ...-A--T--C--C--C--G--A--G--G--G-- 8358

B-YU2 ...-C--T--C--C--C--A--T--A--G--T--G--T--GT-- 8726

B-YU10 ...-C--TT--C--C--C--A--T--A--G--T--G--T--GT-- 8728

B-ACH320.2A.2.1 ...-T--G--G--G--G--GT-- 8769

B-ACH320.2A.1.2 ...-A--G--A--A--C--C--C--C--G--A--T--C--C-- 8779

B-SF2 ...-TT--C--C--C--A--G--CA--G--A--A--A--T-- 8770

B-D31 ...-T--T--C--C--C--C--C--G--A--G--G--GT-- 8292

B-MHMC ...-T--TT--C--C--TT--C--G--A--G--GT-- 8725

B-HAN ...-G--C--G--C--T--C--G--G--C--GC-T-A--G--G--A--G--G-- 8227

B-C1BND ...-G--C--G--C--T--C--G--G--C--GC-T-A--G--G--A--G--G-- 8637

C-BIH2220 ...-TC--G--T--G--A--G--A--A--C--A--T--G--A--A--AT--T--A--A--T--G--T--CTG-A-- 8162

C-92BR025.8 ...-TC--T--TA--GT--A--G--A--C--T--G--A--T--A--A--A--A--G--CA--G--T--G--CTG-A-- 8124

D-84ZR085 ...-A--T--GA--G--A--C--G--G--A--A--C--AAG--G--A--G--G--T--C--C-- 8286

D-NUR ...-T--G--G--G--AG--C--G--A--T--A--A--G--A--G--C--TA--G--C-- 8269

D-2226 ...-A--T--G--G--C--AG--C--G--A--T--A--G--A--G--C--G--TA--C--C-- 8758

D-KLI ...-A--T--G--G--C--AG--C--T--G--T--A--A--A--A--G--C--G--TA--C--C-- 8301

D-94UG114.2 ...-TT--CC--T--A--T--A--C--C--G--GT--TA--GT--C--G-- 8063

F-93BR020.1 ...-ACA--G--T--A--C--A--C--T--A--A--G--A--T--G--G--T--G--G--A-- 8069

H-90CR056.1 ...-T--A--A--C--G--A--A--GA--A--A--G--A--T--A--G--A--T--G--G-- 8084

AC-ZAN184 ...-T--TTA--G--T--T--G--A--A--A--G--A--T--A--T--A--A--G--A--T--C-- 8111

AC-92RM009 ...-TG--G--T--T--A--GG--A--A--C--G--A--A--T--A--A--T--A--A--T--A--C--CTATA--C-- 8118

ADI-MAL ...-A--G--T--T--G--A--C--A--A--A--A--A--A--A--A--A--T--G--A--TT--G--T-- 8248

AE-93TH253 ...-T--TTA--G--A--T--A--TC--T--G--TTT--AG--TG--G--G--T--G--CA--G--G--G--C--C-- 8773

AE-CHE40 ...-T--TTA--G--A--T--A--TC--G--T--A--C--G--G--T--C--G--CA--G--G--G--C-- 8339

AE-90CF402 ...-T--TCA--GTT--GA--T--A--C--G--T--A--A--G--G--T--C--T--C--T--C-- 8849

AGI-2321 ...-TG--ATC--C--G--T--A--A--G--A--AGT--A--C--CY--G--A--G--G--T--A--A--C-- 7965

AG-92NG083.2 ...-TT--G--T--G--A--A--G--A--A--T--A--AC--A--C--G--CA--G--A--A--G--C--G-- 8112

AQ-92NG003.1 ...-T--T--G--T--G--A--A--A--G--A--T--A--A--CT--A--C--G--CA--G--G--A--A--G-- 8079

AQ-IBMG ...-TA--ATCA--C--GTT--A--A--A--A--T--A--T--A--CT--A--C--GA--G--G--G--A--G--A-- 8291

BP-93RM029.4 ...-GC--T--A--A--C--A--A--T--A--T--A--C--GCTT--G--G--G--G--TA--G-- 8084

C-NVPS180 ...GCTG-AA--A--CTA--A--T--G--A--ACA--ACC--G--A--T--TT--AG--GT--T--CA--TT--T--CG--CA--C--CTT--GGT--A--TAG--ACA--GAT--C--T--C--A--A--T-- 8860

C-AM770 ...GCTG-AACA--A--C--CTA--A--T--GC--ACA--C--AG--CA--ACTT--AG--G--CA--TT--T--CG--CA--A--C--C--GGCA--A--TAG--A--C--GA--GTA--C--A--G--T-- 8834

SIVCPZAM7 ...A--TA--T--A--G--AA--G--A--T--A--A--CC--G--A--G--T--A--A--T--AA--A--GCTT--T--GTT--ACT--AT--AGAA--C--GC--C-- 8826

SIVCPZGAB ...L L Q Y W S O E L K N S A V S L L L N A T A I A V A E G T D R V I E I V Q R T C R A I L H T I P R R I Env

Env gp41 Rev ...T F A V L E R Q T K R E S C C S F A S C M S Y S S S S G D R S G Y R M S A K N L S S V S P H V S K N

	Env CDS end (Frame 1) <--	Nef CDS start --> (Frame 2)		
B-WEAU	AGACAGGGCTTAGAAAGGGCTTGGCTATAA	GATGGGTGGCATATGGTCAAAACGTAGTGGATGGGCT	..GCTATAAGGAAAGATGAACAGCTGAO	8875
B-WEAU	R Q G L R A L L S	D G W H M V K T S H S M N A I	I Y K C K N E T S S S	8884
B-WEAU	R R A S K G L C V	S M G G I W S R R S G S G W P	A I R E R M K R A R	8889
B-WEAU	K T S L R K G P A I	S M V A Y G O M V V V V D G S	L S G K E S N B L E	8890
A-0455	A	C	AAQ-CA-AGTG-A	8319
A-92UG037.1	G	C	AA-AG-GA-T-CATAGTG	8882
A-023	G	C	G-AG	8326
B-MN	G	C	A	8887
B-ADB	G	C	ATGCC	8858
B-MR27	G	C	TGG	8246
B-BF	G	C	AGATGG	8413
B-896	G	C	G-CAGAG	8865
B-BC903	G	C	TG-AGG	8400
B-JRFL	G	C	TGCC	8106
B-JRC5P	G	C	TGCC	8884
B-OYI	G	C	ATG-AG	8421
B-CAM1	A	C	CTGG	8880
B-MY5	A	C	ATG	8805
B-pML43	G	C	TG-T	8858
B-HXB2	G	C	TG-T	8868
B-LAI	G	C	TGGT	8461
B-YU2	G	C	ATGCC	8888
B-TU10	G	C	ATGCC	8860
B-ACH320.2A.2.1	TC-G	AGC	AGC	8877
B-ACH320.2A.1.2	G	C	TG-T	8882
B-SF2	G	C	ATGE	8884
B-DJ1	G	C	TGGT	8395
B-MANC	G	C	AAAG-T	8860
B-BAN	G	C	T	8329
B-C18MBC	G	C	TGGT-A-A-A-GTAC-T	8721
Nef deletion				
C-RTH220	G-GCA	A	A-G	8265
C-92BR025.8	T-GCA	A	A-G	8234
D-84ZR085	T	A	A-G	8412
D-MOK	G-CT	A	A-G-AA-A	8372
D-2226	CT	C	A-G-A	8861
D-ELI	T-A-T	A	A-G-A-A-A-T	8404
D-94UG114.1	G	A	A-G-A	8106
F-93BR020.1	G	A	A-G	8100
H-90CR056.1	T-AAGC	T	A-G	8188
AC-ZAM164	T-A-G	A	A-G	8214
AC-92RW009	T-GCA	A	A-G	8221
ADI-NAL	C	C	A-G	8451
AE-93TH253	C	A	A-G	8876
AE-CN240	A	C	AA-T-AG	8442
AE-90CP402	G	C	A-T	8975
AGI-2321	A	A	A-G	7995
AG-92NG083.2	A	A	A-G	8214
AG-92NG003.1	A	A	A-G	8182
AG-1BWG	C	C	A-G	8394
BF-93BR029.4	G	A	A-G	8187
O-NVP5180	A-TGC	A	A-G	8977
O-ANT70	AAG	AT-G	A	8954
SIVCPZANT	A-OC-G	A-T	OCAGTCATTA	8941
SIVCFZGAS	A-OC-G	A-T	OCAGTCATTA	8941
Env gp41	R Q G L R A L L S	Envelope CDS gp41 end	M G G I N S K R S G S G W P	Nef

HIV1 Complete Genomes

B-NEAU	.CCAGCAGCAGAAAGGGTGGGAGCAGTATCTCGA.....GACTTGGCAAAACATGGAGCAATCACAAGTAGCAATGAG.....AATAATGCTGATTCCCTCTGTTAAAGCACAAGAGGATGAG.....GAGGTGGG	8997
B-NEAU	I S S R R G G S S I S L L G K T N S M H K I O S I 6 C S L R L V R S T R G S I G G	2856
B-NEAU	P A A E G V G A V E R D L A K H G A I T S S M E N M A D C V M L K A Q E D E E V G	2730
B-NEAU	I Q Q Q K G N E Q V L I T M Q N M E Q S O V A M I I T M L I A S G S R H R M I R M	2686
A-U455	.G-----A-----A-----A-----A-----T-----A-----A-----T-----G-----C-----T-----ACATCATCT-----C-----CAG-----T-----C-----C-----G-----G-----A-----GA-----C-----A-----	8447
A-92UG037.1	TA-----A-----A-----A-----A-----T-----A-----A-----T-----G-----C-----TA-----C-----CC-----AG-----A-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8405
A-Q03	.-----CC-----A-----A-----A-----A-----A-----A-----T-----A-----G-----AG-----C-----CC-----AG-----A-----A-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8448
B-MN	G-----T-----T-----A-----A-----C-----C-----C-----C-----A-----ACAGCAGCT-----CC-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	9016
B-ABE	.-----T-----T-----C-----C-----G-----C-----C-----AG-----C-----C-----C-----C-----C-----G-----G-----A-----A-----A-----A-----	8986
B-NR27	.-----G-----G-----A-----A-----A-----A-----A-----A-----C-----A-----ACAGCAGCT-----C-----C-----T-----C-----C-----G-----G-----A-----A-----A-----A-----	8374
B-RP	.-----T-----T-----C-----C-----A-----A-----A-----A-----A-----A-----C-----AG-----A-----A-----ACAGCAGCT-----C-----C-----AC-----C-----G-----G-----GRTGAG-----	8547
B-896	A-----T-----T-----A-----A-----A-----A-----C-----G-----G-----T-----T-----C-----C-----ATT-----C-----C-----G-----G-----G-----GA-----A-----A-----	8988
B-BCS7J	.-----T-----T-----A-----A-----A-----A-----C-----A-----A-----A-----C-----A-----ACAGCAGCT-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----	8528
B-JREFL	G-----T-----T-----A-----A-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----CC-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8315
B-JRCSE	G-----T-----T-----A-----A-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----CC-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	9013
B-OYI	G-----T-----T-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----C-----C-----T-----C-----C-----G-----G-----A-----A-----A-----A-----	8550
B-CAM1	G-----G-----T-----A-----A-----A-----A-----A-----A-----ACAGCAGCT-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	9009
B-NY9	.-----T-----T-----A-----A-----A-----A-----C-----A-----A-----A-----C-----A-----ACAGCAGCT-----C-----C-----A-----T-----C-----C-----G-----G-----G-----A-----A-----	8933
B-PML43	.-----T-----T-----A-----A-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----	8986
B-MXB2	.-----TA-----T-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----CC-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8996
B-LA1	.-----C-----T-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8589
B-YU2	G-----T-----T-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8987
B-YU10	G-----T-----T-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8989
B-ACH320.2A.2.1	.-----T-----T-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	9005
B-ACH320.2A.1.2	.-----T-----T-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	9010
B-SP2	G-----T-----T-----C-----C-----C-----C-----C-----A-----ACAGCAGCT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	9013
B-D31	.-----G-----G-----A-----A-----A-----A-----C-----A-----ACAGCAGCT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8523
B-MANC	G-----G-----G-----A-----A-----A-----A-----C-----A-----ACAGCAGCT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8989
B-HAN	G-----T-----T-----C-----C-----C-----C-----C-----AA-----ACAGCAACT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8458
B-C18MBC	.-----T-----T-----C-----C-----C-----C-----C-----AA-----ACAGCAACT-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8721
C-ETH220	.-----G-----A-----A-----CG-----A-----A-----AC-----T-----G-----C-----T-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8396
C-92BR025.8	.-----G-----A-----A-----C-----CG-----A-----A-----CA-----AT-----T-----C-----G-----C-----G-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8355
D-842R085	G-----T-----T-----C-----C-----C-----C-----C-----A-----G-----G-----G-----G-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8544
D-NDK	.-----T-----T-----A-----A-----A-----A-----C-----A-----G-----G-----G-----G-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8503
D-Z226	.-----T-----T-----A-----A-----A-----A-----C-----A-----G-----G-----G-----G-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8992
D-ELI	.-----T-----T-----A-----A-----A-----A-----C-----A-----G-----G-----G-----G-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8535
D-94UG114.1	.-----T-----T-----C-----C-----C-----C-----C-----A-----G-----G-----G-----G-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8314
F-93BR020.1	T-----G-----G-----G-----G-----G-----G-----A-----A-----G-----GG-----G-----T-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8309
H-90CR056.1	.-----T-----T-----A-----A-----G-----G-----A-----A-----T-----AT-----G-----GC-----G-----G-----T-----AT-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8319
AC-ZAM184	.-----T-----T-----A-----A-----G-----G-----A-----A-----T-----T-----G-----G-----G-----G-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8345
AC-92RW009	.-----G-----A-----A-----A-----CG-----A-----A-----C-----A-----AC-----T-----G-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8352
ADZ-MAL	.-----A-----A-----AC-----A-----A-----A-----A-----GATGCAGTATCTCAA-----T-----A-----AT-----TG-----GC-----G-----C-----G-----C-----C-----C-----A-----A-----A-----A-----	8588
AE-93TH253	.-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----TC-----A-----AT-----G-----A-----T-----AT-----T-----C-----G-----G-----G-----A-----A-----A-----A-----	8998
AE-CM240	.-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----TC-----A-----AT-----T-----A-----T-----ATA-----G-----T-----C-----G-----G-----G-----A-----A-----A-----A-----	8564
AB-90CF402	TG-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----TC-----A-----AT-----G-----A-----C-----T-----ATA-----AAT-----C-----G-----G-----G-----A-----A-----A-----A-----	9098
AGI-2321	.GT-----A-----A-----A-----A-----T-----A-----T-----GG-----C-----C-----C-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8349
AG-92NG083.2	.-----A-----A-----A-----C-----C-----A-----A-----A-----T-----GG-----C-----C-----C-----C-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8310
AG-92NG003.1	.-----AC-----A-----A-----A-----A-----A-----A-----T-----A-----AT-----G-----G-----G-----G-----C-----C-----T-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8522
AG-IBRG	.-----A-----A-----A-----A-----A-----A-----C-----A-----A-----G-----G-----G-----G-----T-----C-----C-----C-----C-----G-----G-----G-----A-----A-----A-----A-----	8315
BP-93BR029.4	A-----TGT-----CCT-----A-----A-----T-----C-----CA-----G-----G-----A-----A-----GCTAGA-----G-----G-----AT-----TC-----C-----C-----C-----C-----C-----A-----CCCT-----CA-----TCC-----G-----C-----AGC-----CA-----A-----A-----	9106
O-MVP5160	A-----TGC-----CCT-----A-----A-----C-----AGA-----C-----CA-----G-----G-----A-----A-----A-----GCTAGA-----GG-----AC-----TC-----C-----C-----C-----C-----C-----A-----CCCT-----CA-----TCC-----G-----AGT-----CC-----A-----G-----A-----A-----	9083
O-AMT70	.-----T-----T-----A-----A-----A-----A-----T-----GAAG-----C-----A-----A-----G-----C-----T-----T-----T-----G-----C-----C-----C-----C-----C-----CAAC-----CTA-----CT-----C-----G-----G-----A-----ATG-----CA-----A-----A-----A-----A-----	9069
STVCF2AMT	.-----T-----T-----A-----A-----A-----A-----T-----GAAG-----C-----A-----A-----G-----C-----T-----T-----T-----G-----C-----C-----C-----C-----C-----CAAC-----CTA-----CT-----C-----G-----G-----A-----ATG-----CA-----A-----A-----A-----A-----	9069
STVCF2GAB	.-----T-----T-----A-----A-----A-----A-----T-----GAAG-----C-----A-----A-----G-----C-----T-----T-----T-----G-----C-----C-----C-----C-----C-----CAAC-----CTA-----CT-----C-----G-----G-----A-----ATG-----CA-----A-----A-----A-----A-----	9069
Nef	P A A R G V G A V S R D L A K H G A I T S S M E N M A D C V M L K A Q E D E E V G Nef	

		frameshift in B-WEAU, Nef CDS end		1 3' LTR U3 start	
		< --			
B-WEAU	TTTCCAGTTAGACCTCAGGTACCTTAAAGACCAATGACTACAGGAGGCGATGATCTTAGCCACTTTT.AAAAGAAAAAGGGGGGACTGGAAGGGGCTAATTTACTCCCAAAAAAGACAAGATATCTTGATCTGTGGGTCTACCACAC				9146
B-WEAU	F S S S T S G T F E T N D L O G S A S S P L I K R K G G T G R A M L L P K K T R Y P S S V G L P H				2901
B-WEAU	F P V R P Q V P L R P K T V K A A H D L S M P I K R K G G L E G L I Y S C K R O D T L D L W V Y H T				2779
B-WEAU	V P O L D L R Y L S D Q S L T R O R N I L A T F I K K R O D M K G S F T P K K D R I S L I C O B T T				2732
A-U455	C-C-C-G-A-----T-----TTT-----C-CT-----T-----T-----C-A-G-----A-----T-----				8597
A-92UG037.1	C-----C-A-A-----C-G-----TTTC-----CG-CT-----T-----T-----A-G-----A-----				8555
A-Q23	C-----C-G-A-A-----C-----G-TGTG-----C-CT-----T-----T-----T-G-----TAGG-----A-----T-----				8598
B-MN	C-----C-A-----A-----TTTA-----A-----T-----T-----T-----G-----C-----				9186
B-AD8	C-----C-----A-----AGTA-----T-----T-----CT-----G-----G-----T-----				9136
B-WR27	C-----C-----A-----TGTA-----T-----T-----TGTA-----T-----G-----G-----				8524
B-RF	C-----C-----G-T-----T-----TGTA-----T-----T-----G-G-T-----G-----T-----				8697
B-896	C-----C-----TGTA-----T-----T-----TGTA-----T-----G-C-----G-----				9138
B-BCS93	C-----C-----A-----TGTA-----A-----T-----T-----T-----G-----G-----AC-----				8678
B-JRFL	C-----C-----G-----TGTA-----T-----T-----C-----A-G-----C-----				8465
B-JRCBP	C-----C-----T-----TATA-----T-----T-----A-G-----A-G-----A-----				9161
B-OVI	C-----C-----G-----TTTA-----T-----T-----G-----G-----T-----T-----				8700
B-CAM1	C-----C-----G-----TTTA-----A-----T-----T-----A-----G-----G-----A-----				9159
B-NYS	C-----C-----A-CG-----TATG-----T-----G-----				9022
B-PNL43	C-----C-----TGTA-----T-----T-----C-----G-----A-----				9136
B-HXB2	C-----C-----TGTA-----T-----T-----C-----G-----A-----				9146
B-LAI	C-----C-----TGTA-----T-----T-----C-----CG-----A-----				8739
B-YU2	C-----C-----C-----TATG-----T-----T-----C-----C-----				9137
B-YU10	C-----C-----G-----TATG-----T-----T-----C-----C-----				9139
B-ACH320.2A.2.1	C-A-----G-----TTTA-----T-----T-----G-----A-----				9155
B-ACH320.2A.1.2	C-A-----G-----TTTA-----T-----T-----G-----G-----A-----				9160
B-8P2	C-----C-----G-----TTTA-----A-----T-----T-----G-----G-----A-----				9163
B-D31	C-A-----T-----TGTG-----T-----T-----G-C-----C-----C-----				8673
B-MANC	C-----C-----T-----G-----CTTG-----T-----G-----G-----G-----C-----C-----				9139
B-HAN	C-----C-----A-----G-----TTTA-----T-----T-----T-----C-----G-----G-----				8608
B-C18NHC	C-----C-----T-----TATA-----G-----T-----T-----C-----A-GG-----C-----				8801
C-BTH2220	C-----C-----T-----ATTC-----C-----T-----T-----A-G-----G-----G-----T-----T-----TA-----				8545
C-92BR025.8	C-----C-----T-----TAGTC-----C-----TT-----T-----G-----T-----G-----T-----TA-----				8505
D-842R085	C-----C-----T-----TGTA-----C-----TT-----T-----G-----A-----G-----G-----T-----TA-----				8694
D-NOK	C-----C-----A-A-----TGT-----T-----T-----GG-----A-G-----G-----A-----				8653
D-2226	C-----C-----A-C-----TTA-----T-----T-----				9081
D-BLI	C-----C-----C-----A-A-----T-TA-----C-----T-----T-----GG-----A-----G-----G-----T-----A-----				8685
D-94UG114.1	A-----C-----A-----A-----GTA-----C-----A-----T-----T-----G-----GG-----CG-----G-----T-----T-----				8464
F-93BR020.1	C-----C-----C-----T-----G-----TGTA-----C-----T-----T-----T-----G-----A-G-----G-----G-----				8459
H-90CR056.1	C-----C-----G-----T-----G-----TPT-----C-----T-----T-----T-----A-GC-----G-----C-----T-----TA-----				8469
AC-2AM184	C-----C-----AG-----A-----C-----G-----TGTG-----C-----T-----T-----T-----G-----G-----G-----A-----				8495
AC-92RW009	C-----C-----C-----G-----T-----A-----AGT-----C-----TT-----T-----C-----TA-G-----G-----C-----G-----T-----TA-----				8500
ADI-MAL	C-----CG-T-----T-----A-G-----TTT-----C-----TT-----T-----T-----G-----GG-----C-----A-----				8738
AE-93TH253	C-----C-----G-----G-----C-----T-----G-----TPT-----TT-----T-----T-----G-----A-G-----G-----CT-A-----T-----T-----				9148
AE-CN240	C-----C-----TG-----G-----C-----T-----G-----TPT-----TT-----T-----T-----A-G-----G-----G-----CT-A-----TA-----				8714
AE-90CF402	C-----C-----G-----G-----C-----T-----G-----TTT-----TT-----T-----T-----C-----A-G-----G-----G-----CT-A-----C-----A-----T-----				9248
AGI-2321					
AG-92NG083.2	C-----G-----C-----A-----G-----T-----T-----TPT-----C-----TT-----T-----T-----A-G-----G-----C-----A-----TA-----				8495
AG-92NG003.1	C-----C-----AA-----G-----T-----T-----CTT-----C-----TT-----T-----T-----TA-G-----C-----C-----A-----TA-----				8460
AG-18NG	C-----C-----G-----A-----G-----G-----T-----G-----TGTG-----C-----T-----T-----T-----A-G-----A-----TA-----				8672
BP-93BR029.4	C-----C-----C-----AG-----A-----C-----G-----T-----G-----T-----T-----T-----T-----A-G-----G-----G-----T-----TA-----				8465
O-MVP5180	C-----C-----A-----A-----G-----C-----G-----C-----TT-----A-----CTT-----C-----TT-----T-----T-----A-----T-----G-----GC-----A-----G-----C-----A-----T-----				9256
O-AMT70	C-----C-----AGC-----A-----G-----C-----G-----C-----T-----A-----G-----ATT-----C-----C-----TT-----T-----T-----A-----T-----T-----G-----TA-----				9233
SIVCP2AMT					
SIVCP2GAB	A-----C-----A-----A-----T-----AACG-----T-----A-----TTT-----TCA-----T-----T-----AGG-G-----G-----G-----C-----C-----T-----				9219
Nef	F P V R P Q V P L R P K T Y K A A H D L S F F I K R K G G L E G L I Y S C K R O D T L D L W V Y H T Nef				
		Nef end in NEAU only			

HIV1 Complete Genomes

B-WBAU	ACAAGGCTTCCTCCCTGATTGGCAGAACTACACACAGGGCCAGGSACTAGATACCCACTGTGCTTTGGATGCTGCTTCAGCTAGTACCBGTGGAACCA...GAGAGGTTAGAGAA...GCTAATGAAAGGAGAGAACACAGCTTGT	9290
B-WBAU	T R L L F S L A E L H T R A R D S I P T V L W M V L O A S T S G T S I E G R R I S S R R R E Q Q L V	2943
B-WBAU	Q G F F P D W O N Y T P G P G T R Y P L C F G W C F R L V P V E P E K V E E A N E G E N N S L L	2827
B-WBAU	H K A S S L I Q R T T M Q G Q G L D T H C A L D G A S S S V O N N I R R S K I L M K E R T T A C	2776
A-U455	A-----T-----A-TC-----AACA-----A-----S-----T-----T-----CTG-A-----G-----C-----G-----TC-AC-	8741
A-92UG037.1	A-----T-----T-----TC-----AACA-----A-----TGA-----TG-A-----G-----C-----G-----T-----C-----A-----	8699
A-Q23	AT-----T-----C-----G-----T-----AACA-----G-----T-----T-----TG-A-----GA-G-----C-----G-----C-----A-----	8742
B-MN	A-----T-----TC-----T-----AC-----A-----T-----G-----TG-A-----G-----C-----A-----T-----T-----C-----	9310
B-AD8	A-----GTC-----T-----AC-----G-----C-----T-----G-----C-----A-----G-----C-----A-----T-----C-----	9280
B-WR27	A-----A-----C-----T-----AC-----C-----T-----G-----TC-----G-----C-----C-----T-----T-----	8668
B-RF	A-----A-----C-----T-----AC-----T-----G-----T-----T-----G-----C-----C-----G-----C-----C-----	8841
B-696	A-----A-----A-----TC-----T-----AC-----A-----T-----G-----TG-A-----G-----G-----CA-CA-----G-----C-----	9276
B-BCSQ3	A-----A-----TC-----T-----AC-----T-----T-----T-----TG-A-----G-----C-----C-----A-----C-----	8822
B-JRFL	A-----A-----TC-----T-----AC-----T-----G-----T-----A-----G-----C-----C-----T-----T-----C-----	8609
B-JRCSF	A-----A-----G-----A-----GTC-----T-----AC-----T-----T-----T-----G-----C-----C-----T-----T-----	9307
B-OYI	A-----AT-----TC-----T-----A-----T-----A-----T-----CC-----G-----C-----C-----C-----C-----	8814
B-CAN1	A-----A-----TC-----T-----ACA-----T-----G-----T-----C-----C-----G-----C-----A-----A-----GC-----	9303
B-MY9	A-----A-----GTC-----T-----AC-----A-----T-----G-----T-----G-----G-----C-----A-----C-----C-----	9280
B-pNL43	A-----A-----GTC-----T-----AC-----A-----T-----G-----T-----A-----G-----C-----A-----C-----C-----	9290
B-MXB2	in frame TAG stop codon	
B-LAI	A-----A-----GTC-----T-----AC-----A-----T-----G-----T-----G-----G-----C-----A-----C-----C-----	8883
B-YU2	A-----A-----G-----AC-----G-----T-----G-----TG-A-----G-----C-----C-----C-----T-----A-----	9278
B-YU10	A-----A-----G-----AC-----G-----T-----G-----TG-A-----G-----C-----C-----C-----T-----T-----	9283
B-ACH320.2A.2.1	A-----A-----TC-----G-----T-----AC-----T-----G-----T-----G-----A-----A-----G-----C-----C-----	9299
B-ACH320.2A.1.2	A-----A-----TC-----T-----AC-----T-----G-----T-----G-----A-----A-----G-----C-----C-----	9304
B-SF2	A-----A-----T-----TC-----T-----AC-----T-----G-----T-----G-----A-----A-----G-----C-----C-----	9307
B-D31	A-----A-----C-----T-----AC-----T-----T-----T-----TA-G-----G-----C-----C-----G-----C-----	8817
B-MANC	A-----A-----G-----TC-----T-----GC-----T-----T-----T-----C-----G-----G-----C-----C-----	9283
B-HAN	A-----A-----GTC-----T-----AC-----T-----T-----T-----T-----G-----G-----A-----A-----G-----C-----	8734
B-C18HBC	A-----A-----A-----GTC-----T-----AC-----T-----T-----T-----GA-----A-----G-----C-----C-----A-----A-----	8854
C-ETH2320	A-----A-----A-----GTC-----T-----AC-----T-----T-----T-----AGTG-A-----G-----ATC-----T-----C-----	8690
C-92BR025.9	A-----A-----A-----GTC-----G-----T-----C-----AC-----G-----T-----C-----AG-G-----G-----C-----C-----AC-----	8649
D-64ZR085	A-----A-----A-----TC-----T-----T-----AAC-----G-----G-----G-----T-----T-----GT-----A-----G-----C-----C-----AG-----T-----C-----	8838
D-NDK	A-----A-----C-----TC-----T-----AC-----C-----T-----T-----C-----G-----G-----C-----C-----A-----G-----T-----	8797
D-2286	A-----A-----A-----TC-----T-----AAC-----A-----G-----T-----T-----C-----G-----AC-C-----C-----C-----	8829
D-94UG114.1	A-----A-----A-----T-----T-----AC-----G-----G-----T-----A-----G-----A-----G-----G-----AAC-C-----G-----T-----	8608
F-93BR020.1	A-----A-----TC-----T-----AC-----A-----G-----G-----G-----T-----C-----A-----G-----C-----C-----T-----C-----	8603
H-90CR056.1	A-----A-----C-----GAG-----T-----C-----AC-----G-----AA-----T-----C-----G-----C-----G-----C-----C-----	8613
AC-2AM184	A-----A-----A-----T-----C-----T-----AAC-----T-----A-----T-----T-----TCTG-----G-----C-----G-----G-----A-----	8639
AC-92RM009	A-----A-----A-----GTC-----T-----ACT-----T-----A-----G-----T-----C-----AG-G-----A-----G-----C-----C-----G-----T-----AC-----	8646
ADI-MAL	A-----A-----A-----T-----T-----AC-----C-----T-----T-----T-----A-----AGT-----C-----C-----C-----G-----T-----TC-----	8882
AE-93TH253	A-----A-----A-----TC-----T-----T-----T-----T-----A-----T-----C-----AGAG-A-----G-----AC-----CA-----A-----T-----C-----	9292
AE-CH240	A-----A-----A-----TC-----T-----T-----T-----T-----A-----T-----C-----AGAG-A-----G-----AC-----CA-----A-----T-----C-----	8858
AE-90CP402	A-----A-----A-----G-----GTC-----T-----T-----T-----T-----T-----AGAG-A-----G-----AC-----A-----A-----T-----C-----	9392
AGI-2321	A-----A-----A-----CT-----AC-----G-----A-----A-----A-----C-----CAG-A-----G-----C-----A-----G-----T-----TC-----A-----	8639
AG-92NG081.2	A-----A-----A-----T-----T-----ACA-----C-----G-----A-----A-----T-----CAG-A-----G-----C-----A-----T-----A-----	8604
AG-92NG003.1	A-----A-----A-----T-----T-----AC-----G-----A-----A-----T-----CAG-A-----G-----C-----A-----T-----A-----	8816
AG-IBNG	A-----A-----A-----C-----T-----AC-----A-----G-----T-----T-----C-----G-----A-----G-----C-----C-----T-----C-----	8609
BF-93BR029.4	T-----G-----A-----TGT-----G-----A-----AC-----T-----ACA-----TG-----T-----A-----G-----TC-----G-----AG-----C-----GAG-----CTG-----G-----AC-----AAT-----AG-----TGCT-----TC-----TC-----	9403
O-MYF5180	T-----G-----A-----A-----A-----C-----G-----T-----ACA-----TG-----T-----A-----G-----TC-----GA-----AG-----C-----AG-----CTA-----GA-----AC-----T-----T-----GGCT-----ATC-----CC-----	9380
O-RMT70		
SIUCF2AMP	A-----A-----A-----A-----A-----T-----C-----T-----A-----G-----CC-----AC-----GAG-----C-----C-----C-----C-----T-----T-----C-----	9363
SIUCF2GAB		
Hef	Q G F F P D W O N Y T P G P G T R Y P L C F G W C F R L V P V E P E K V E E A N E G E N N S L L Hef	

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

	NF-kappa-B II										NF-kappa-B I										SPI-III										SPI-II										SPI-I										
B-NEAU	CATC	GAGTT	TTCT	ACAA	GGGACTTTCCAC	TTGGGACTTTCCCG	GGAGGCGTGATC	G	GGCGGA	CGGGGAGTGGCGA	GCCCT	950																																							
B-NEAU	T	S	F	T	G	L	S					299																																							
B-NEAU	M	S	F	F	O	G	L	P	S			2891																																							
B-NEAU	F	T	V	S	F	G	T	P	H			2812																																							
A-U455	A-CA	-AG	-G	GACA	-	-	-AG	-T	-G	-T	-AG	896L																																							
A-92UG037.1	-CA	-GAG	-G	GAC	T	-	-G	-T	-G	-T	-AG	8910																																							
A-Q23	-CA	-AG	-GT	GAC	T	-	-G	-T	-G	-T	-AG	8953																																							
B-MN	-C	-C	A	-	-	-	-C	-A	-T	-G	-C	9519																																							
B-AD8	-	-	-	-	-	-	-G	-A	-T	-G	-C	9490																																							
B-WR27	-A	-C	-T	-	-	-	-G	-A	-T	-G	-C	8879																																							
B-RF	-	-	-	-	-	-	-G	-A	-T	-G	-C	9051																																							
B-896	A-CT	-C	-A	-	-	-	-G	-A	-T	-G	-C	9491																																							
B-BC933	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9032																																							
B-JRFL	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	8819																																							
B-JRCSP	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9516																																							
B-OYT	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9054																																							
B-CAM1	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9514																																							
B-NYS	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9490																																							
B-PM43	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9500																																							
B-MXB2	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9093																																							
B-LAI	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9488																																							
B-YU2	-C	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9493																																							
B-YU10	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9510																																							
B-ACH320.2A.2.1	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9515																																							
B-ACH320.2A.1.2	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9516																																							
B-SP2	-	-C	-C	-	-	-	-G	-A	-GA	-G	-C	9035																																							
B-DD1	A	-	-	-	-	-	-G	-A	-GA	-G	-C	9494																																							
B-MANC	-G	-C	-	-	-	-	-G	-A	-GA	-G	-C																																								
B-HAN	-C	-	-	-	-	-	-G	-A	-GA	-G	-C	8955																																							
B-C184BC	-	-	-	-	-	-	-G	-A	-GA	-G	-C	8988																																							
C-ETH2220	-CA	-	-	-	-	-	-G	-A	-GA	-G	-C	8900																																							
C-92ER025.8	T-CA	-AG	-	-	-	-	-G	-A	-GA	-G	-C	8970																																							
D-84ER085	-	-C	-	-	-	-	-G	-A	-GA	-G	-C	8974																																							
D-NDR	-	-C	-	-	-	-	-G	-A	-GA	-G	-C	9007																																							
D-Z226	-	-C	-	-	-	-	-G	-A	-GA	-G	-C	9039																																							
D-ZLI	-	-C	-	-	-	-	-G	-A	-GA	-G	-C	8817																																							
D-94OG114.1	-	-C	-	-	-	-	-G	-A	-GA	-G	-C	8833																																							
F-93NR020.1	A-CA	-AGA	-A	-	-	-	-G	-A	-GA	-G	-C	8818																																							
H-90CR056.1	-CA	-AGA	-	-	-	-	-G	-A	-GA	-G	-C	8863																																							
AC-ZAM184	A-CA	-AG	-G	-	-	-	-G	-A	-GA	-G	-C																																								
AC-92RM009	-CA	-A	-	-	-	-	-G	-A	-GA	-G	-C	8857																																							
ADI-MAL	-CA	-AG	-G	-	-	-	-G	-A	-GA	-G	-C	9094																																							
AE-93TN253	-AA	-AG	-ACT	-A	-	-	-G	-A	-GA	-G	-C	9502																																							
AE-CM240	-AG	-AG	-T	-A	-A	-	-G	-A	-GA	-G	-C	9067																																							
AE-9																																																			

HIV1 Complete Genomes

	3' LTR \ / 3' LTR	3' LTR \ / 3' LTR	3' LTR \ / 3' LTR
	TATA box	U3 and \ / R repeat start	Poly-A signal R repeat and \ / U5 start
B-WERU	CAGATGCTGCATATAAGCAGCTGCTTTTGGCTGTACTGGGCTCTCTCTGGTTAGACAGATCTGAGCTGGGAGCTCTCTGAGCTAGAGGAACCCACTGCTT	AAGCCTCAATAAAGCTTGCCCTTGA.GTG.CTTCAAGTAGTGTGTG	9647
B-WZAD	O N L W T S E C F L P V L G L S G G T R S E P O S S L A S S Q T H C I K P O S S L P S S I F K S C V		3041
B-WBAU	R C C I S A A A F C L Y W V S L V R P D I S L G A L W L A R E P T A S L N R A C L I V I S S S V C		2836
B-WBAU	S D A A Y K O L L F C T G S L M L D Q I G A W E L S G S L O G R P L L S A S I K L L A L S L O V V C		2881
A-U455			9107
A-920G037.1			8980
A-Q23			9099
B-MN			9666
B-AD8			9637
B-WR27			9011
B-RP			9128
B-896			9638
B-BCSG3			9168
B-JRPL			8896
B-JRCSP			9540
B-OVI			9190
B-CAN1			9591
B-NY5			
B-DNL43			9637
B-HXB2			9647
B-LAI			9229
B-YU2			9634
B-YU10			9629
B-ACH320.2A.2.1			9631
B-ACH320.2A.1.2			9636
B-SP7			9663
B-D31			9074
B-MANC			9641
B-HAM			9076
B-C18MBC			9135
C-ETH2220			9031
C-92BR025.8			8960
D-84ZR085			
D-MDK			9143
D-2226			
D-BL1			9176
D-94UG114.1			8952
F-93BR020.1			8968
H-90CR056.1			8993
AC-2MCL84			8990
AC-92BR009			8992
ADT-MAL			9229
AE-93TH253			9649
AE-CN240			9203
AE-90CF402			9772
AGT-2321			
AG-92BR0083.2			8986
AG-92N3003.1			8966
AG-IBMG			9173
BF-93BR029.4			8956
O-MVPS180			9762
O-AM770			9754
SIWCP2ANT			
SIWCP2GAB			9740

HIV1 Complete Genomes

Extensive secondary structure in this region		
See Rizvi et al J Virol 67: 2681-88 (1993)		
	3' LTR US end I	
B-NEAD	CCCGTCTG.TTGTGTGACTCTGCTAAGATCCCTCAGACCATTTTGTAGTCAGTGTG.GAAAAATCTCTAGCA	9719
B-NEAD	P_V_I_L_C_D_S_G_M_S_P_S_L_R_P_P_S_S_V_I_B_N_L_S_X	3061
B-NEAD	P_S_I_C_V_T_L_V_T_R_D_P_S_D_N_F_S_O_C_I_K_I_S_S_G	2959
B-NEAD	A_R_I_I_V_S_I_H_G_L_E_I_P_Q_T_I_L_V_S_V_I_K_S_L_A	2902
A-U455	-----A-T-----C-A--A-T--A,-----	9177
A-92UG037.1		
A-Q23	-----C-C--A-G-----,-----	
B-MN	-----A-----G-----T-C-----G-----	9160
B-AD8	-----C-----	9739
B-NR27		9709
B-RP		
B-896	-----G-----CC---T---A,-----	
B-BCSG3		9710
B-JRFL		
B-JRCSF		
B-OYT		
B-CAM1		
B-MY5		
B-PNL43	-----C-----,-----	9708
B-HXB2	-----C-----,-----	9719
B-LAT		
B-YU2	-----C-----,-----	9705
B-YUL0	-----C-----,-----	9710
B_ACH320.2A.2.1		
B_ACH320.2A.1.2		
B-SF2	-----C-----GA-----	9736
B-D31		
B-NANC	-----A-----	9659
B-RAN		
B-CLBNC	-----T-----C-----	-9207
C-ETH2220		
C-92BR029.8		
D-842R085		
D-NDR		
D-2226		
D-BLI		
D-94UG114.1		
F-93BR020.1		
H-90CR056.1		
AC-2AN184		
AG-92BR009		
ADI-MAL	---A---,GT-AG-----T-C-C--A-T-A-A,-----	9720
AE-93TH253		
AE-CM240	-----GT-AG-----C-C--A-T-A-A,-----	9842
AE-90CF402		
AGI-Z321		
AG-93MG093.2		
AG-92MG001.1	-----A-----	9200
AG-IBMG		
BF-93BR029.4		
O-MVP5180	-T-A---,CA,,,--C---GT-----	9793
O-ANT70		
SIVCP2ANT		
SIVCP2GAB	--TA---A,--CA-AC-----TT-AA---TAOTCAAG-,-----	9811

HIV2/SIV Complete Genomes

[illegible]

BEN	AGACCTCTT...CAATAAAGCTGCCAGT...YAGAAAGCAATTAAGTGTGTGTCCCATCTCTCTAGTCGCGCGCTGGTTCATCGGTGTTTCATCTG...ASTAACAAG...ACCCCTGGTCTGTAGGACCCCTTCT	820
BEN	R P L E N K A A S S Q V R C V F P S I L V A A M S F G V H L S N K T L V C S D P S	260
BEN	D I I T R I P I R S L S V C S H I S S P P G H S V F I V T P M S V R T L I	346
BEN	T S Q S C Q I S A S I V C V P S P S A L L V I R C S S I S Q I P E L L G P F	257
GH1	-----A-----	263
D194	-----A-----	264
UC2	-----A-----	819
ROD	-A-----C-----C-----C-----C-----TA-----	264
NIH2	-A-----C-----C-----C-----C-----TA-----	263
ISV	-----C-----TC-----	265
ST	-----C-----A-----A-----TC-----	266
CAN2	-----C-----A-----	830
MDS	-T-----C-----A-----	264
KR	-----C-----A-----G-----	799
UC1	-AC-----A-----G-----A-----C-----TA-----A-----AAT-----A-----AG-----TC-----	820
D205	-AC-----A-----A-----C-----C-----CA-----TA-----G-----C-----TC-----	817
RHO	-AC-----A-----TA-----AC-----C-----CA-----A-----GAA-----C-----TC-----	816
NIH251	-----T-----T-----T-----CC-----A-----TCG-----A-----TCGGT-----A-----G-----TC-----	771
NIH32H	-----T-----T-----T-----C-----A-----TCG-----A-----TCGGT-----A-----G-----TC-----	783
NIH1A11AA	-----C-----T-----T-----C-----A-----TCG-----A-----TCGGT-----A-----G-----TC-----	782
NIH142	-----T-----T-----T-----C-----A-----TCG-----A-----TCGGT-----A-----A-----ACCCCTGGTCTGTAGG-----TC-----	280
NIH239	-----C-----T-----T-----C-----A-----TCG-----A-----TCGAT-----A-----G-----TC-----	783
NIH2	-----T-----T-----T-----CC-----C-----A-----TCG-----A-----TCGAT-----A-----G-----TC-----	264
SNHFBJ14.441	-----T-----T-----T-----CC-----C-----A-----TCG-----A-----TCGAA-----CA-----G-----C-----TC-----	559
SNHFBJ14.15	-----T-----T-----T-----CC-----C-----A-----TCG-----A-----TCGAA-----CA-----G-----C-----TC-----	559
SNHFBJ6.6	-----T-----T-----T-----CC-----C-----A-----TCG-----A-----TCGAA-----CA-----G-----C-----TC-----	751
SNH14	-----A-----T-----C-----TCG-----A-----TCGAC-----ACA-----G-----C-----TC-----	780
SNH19	-----T-----T-----C-----TCG-----A-----TCGAA-----CA-----G-----S-----Y-----TC-----	266
STH	-T-----T-----A-----T-----CA-----A-----TCG-----A-----TCAA-----TC-----	447
5' LTR <-1		
BEN	CGCTTTGGGAATCCAAAGCAGGAA...AATCCCTAGCAGTTGGCGCCCGGAACAGGAGCTTGA...AGAGGACTGAGA...AGCCCTGGAACTGG...CTGAGTGAAGCCAGTAAGGGCGGCAGGACAAACACAGACCGAGTGTCTCT	958
BEN	R P Q M F R O E N P S Q V G A R T G T S R G L N S F G T R L S E Q E K G G R N M F R R S A F	304
BEN	A L G I O G R I P S R L A P R C O L I E D S I A L E L I S V K A V R A A G T N H D G V L L	286
BEN	S L M R S X A G S S L A O M R P N R D L I S T B I P M N S I E S P O S G R O S O T T T E C S	297
GH1	-----A-----AG-----A-----T-----	398
D194	-----A-----A-----	403
UC2	-----A-----A-----	957
ROD	T-----A-----G-----T-----A-----T-----T-----A-----	402
NIH2	T-----A-----G-----CGAA-----A-----T-----A-----G-----G-----GC-----	404
ISV	T-----A-----G-----G-----A-----T-----A-----G-----	403
ST	T-----A-----A-----A-----T-----A-----	404
CAN2	T-----A-----A-----A-----T-----A-----G-----A-----	948
MDS	T-----A-----G-----A-----A-----T-----T-----A-----A-----	402
KR	T-----G-----A-----G-----A-----A-----GA-----A-----	926
UC1	T-----A-----A-----A-----A-----A-----G-----GTCTT-----G-----A-----TC-----A-----	958
D205	T-----A-----G-----A-----G-----G-----A-----G-----GTCTT-----AA-----G-----A-----TC-----A-----	955
RHO	T-----A-----A-----A-----C-----G-----A-----G-----GTCTT-----G-----A-----TC-----	954
NIH251	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----	908
NIH32H	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----	921
NIH1A11AA	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----	920
NIH142	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----	418
NIH239	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----	921
NIH2	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----G-----	402
SNHFBJ14.441	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----	696
SNHFBJ14.15	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----	696
SNHFBJ6.6	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----	888
SNH14	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----A-----	917
SNH19	T-----A-----A-----G-----A-----A-----GA-----G-----G-----CT-----GTA-----C-----	403
STH	T-----CTA-----TG-----A-----A-----GA-----G-----G-----CT-----GTA-----	595

GEN	AGAAAGGCGCGGCGGAGTACCAAAAG...GCAGCCTGTGGAGCGGGAGTGAAGAGGCGCTCC...GGTGAAGG...TAAGTACCTACACCAAAACT...GTAGCCAGAAAA...GCGTGTATCC...TACCTT...TAGACA...	1084
BEN	R K A R A E V P R . A A C G A G V X E A S . G S R . S V E T P K T . V A R K . G I L S . Y L . S T .	343
BEN	E R R G P R Y O . I R R V E A B E K R P . G G . K V Y L H O K I . S P E E . A C V Y . I T . H R .	318
BEN	S X G A G R G T K . I G V N S G S E R G L S . I V R . S T Y . T X N . S Q K . L V I . F A . L D .	331
GH1	-----A-----G-----G-----G-----C-----A-----T-----T-----	525
D194	-----A-----G-----A-----G-----G-----G-----T-----T-----	828
UC2	-----A-----A-----A-----G-----G-----G-----C-----T-----T-----	1084
ROD	-----A-----GAG-----G-----G-----C-----	527
NIH2	-----A-----CG-----G-----A-G-G-----	528
ISY	-----A-----C-----A-G-G-----	528
ST	-----A-----G-----T-A-----AG-G-----G-----	529
CAN2	GT-----A-----C-----T-T-G-----T-----T-----	1091
NDS	-----A-----A-----G-----A-G-----T-----	528
KR	-----A-----A-----T-----	1052
UC1	GT-----A-----G-----A-----GA-----GG-A-GA-----G-----T-----T-----CTGA--T-GTC-G-TC--AAGA-AGG--	1084
D205	GT-----A-----G-----A-----GAQ-----G--AGT-----T-----T-----TGC--T-GTC--TC--AA--GGG--	1089
ZHO	GT-----AA-----G-----A-----GAG-----GG-A-A-A-----G-----T-----T-----CA--T-GTC--TC--AGGACAGG--	1061
NIH251	-T-----T-----G-----A-----A-G-G-----T-C-----G-AACACAA--GAA-A--T-----T-----AGGA-GG--GA-ATA-G--	1028
NIH32H	-T-----T-----G-----A-----A-GACACAA--GAA-A--T-----T-----AGGA-GG--GA-ATA-G--	1041
NIH1A11AA	-T-----T-----C-----A-----A-G-----T-C-----G-AACACAA--GAA-A--T-----T-----AGGA-GG--GA-ATA-G--	1039
NIH142	-T-----T-----G-----A-----A-GAGAG-----T-----T-C-----G-AACA--A--GGA-A--T-----T-----AGGA-GG--GA-ATA-G--	538
NIH239	-T-----T-----G-----A-----A-G-----T-C-----G-AACACAA--GAA-A--T-----T-----AGGA-GG--GG-ATA-G--	1040
NHE	-T-----T-----G-----A-----A-----T-C-----G-AACACAA--GAG-A--T-----T-----AGGA-GG--GA-ATA-G--	520
SHNFBJ14.441	-T-----G-----CGG-----T-C-----G-AACA-AA-G-GT-AT--AGGA-TGA-----CC-----T-TGAG-A-AGA--	812
SHNFBJ14.15	-T-----G-----CGG-----T-C-----G-AACA-AA-G-GT-AT--AGGA-TGA-----CC-----T-TGAG-A-AGA--	813
SHNFBJ6.6	-T-----G-----CGG-----T-C-----G-AACR-AA-G-GT-AT--AGGA-TGA-----CC-----T-TGAG-A-AGA--	1005
SHNH4	-T-----G-----CGG-----T-C-----G-AACA-GA--GT-AT--AGGA-TGA-----CC-----T-TGAGGA-AGA--	1033
SHNH9	-T-----G-----CGG-----T-C-----G-AACA-AA-G-GT-AT--AGGA-TGA-----CC-----T-TGAG-A-AGA--	519
STN	GT-----GA-----A-----GCG-----A-----T-C-A-----G-AACA--AGT-GC-AG--AGG-TGA-----C-----T-TGAG-A-AGA--	693
--> Gag and Gag-Pol cds start (frame 1)		
BEN	GGTAGAAGATTG...TGGGAGATGGGCGCGAGAACTCGTCTTGAGAGGGAAGAAAGCAGACGAATTAGAAAAAGTTAGGTTACGGCCCGCGCGAAGAAAGATACAGGTTAAACATAATTGTGTGGGAGGGAATGAATTGGATPAAA	1231
BEN	_R_R_L_M_E_N_G_A_R_N_S_V_L_R_G_X_X_A_D_E_L_E_K_V_R_L_R_P_G_Q_Q_K_R_V_R_L_K_H_I_V_W_A_A_N_E_L_D_K	392
BEN	_V_E_D_I_I_G_R_W_A_R_R_T_P_S_S_E_R_Q_K_K_Q_T_X_S_K_K_L_G_V_Q_P_A_E_R_K_S_T_Q_S_M_Y_L_C_G_O_R_M_M_M_T_N	363
BEN	_S_K_I_I_G_D_G_R_E_X_L_R_L_E_R_B_K_S_R_R_T_R_R_S_V_T_A_R_R_K_R_V_V_Q_V_K_T_Y_C_V_Q_S_E_S_T_Q_S	374
GH1	-----T-----GA-----A-----GA-----G-----A-----C-----G-----	672
D194	-----C-----G-----A-----A-----GA-----G-----A-----C-----G-----	675
UC2	-----T-----G-A-C-----C-----A-----C-----G-----	1231
ROD	-----T-----A-----A-----A-----C-----G-----	674
NIH2	-----A-----G-----T-----A-----G-----A-----A-----C-----G-----	675
ISY	-----A-----G-----T-----A-----G-----A-----A-----C-----G-----	675
ST	-----QA-----A-----A-----C-----G-----	676
CAN2	-----A-----A-----A-----C-----G-----	1240
NDS	-----G-----A-----A-----C-----G-----	675
KR	-----G-----T-----T-----A-----G-----A-----T-----AC-----G-----GC-----	1199
UC1	A--GGC-AA-G-----G-----TC-----G-----A-----T-----G-----TT-TC-----BA-A-----T-----C-----C-G-----	1212
D205	A--GGC-AA-G-----G-----C-TC-----G-----A-----T-----T-----C-----G-B-A-----T-----C-----A-C-G-----	1207
ERO	.CGGA-AA-G-----G-----TC-----G-----GA-----T-----G-----GA-----T-----C-----A-----A-----T-----A-G-G-----	1208
NIH251	.A-----G-----TC-----G-----A-----C-----A-----T-----T-----G-G-----G-A-A-----A-----A-----G-----	1169
NIH32H	.A-----G-----TC-----G-----A-----C-----A-----A-----T-----T-----G-G-----G-A-A-----A-----A-----G-----	1182
NIH1A11AA	.A-----G-----TC-----G-----T-----A-----C-----A-----A-----T-----T-----G-G-----G-A-A-----A-----A-----G-----	1180
NIH142	.A-----G-----TC-----G-----A-----AC-----A-----A-----T-----T-----G-G-----G-A-A-----A-----A-----G-----	679
NIH239	.A-----G-----T-----TC-----G-----A-----A-----A-----T-----T-----G-G-----G-A-A-----A-----A-----G-----	1181
NHE	.A-----G-----TC-----G-----T-----A-----C-----A-----G-----T-----G-G-----G-A-A-----A-----A-----G-----	661
SHNFBJ14.441	.G-----G-----TC-----G-----A-----A-----G-----G-----CA-C-G-G-----A-----A-----A-----C-----G-----	956
SHNFBJ14.15	.G-----G-----TC-----G-----T-----C-----A-----G-----A-C-G-G-----A-----A-----A-----C-----G-----	957
SHNFBJ6.6	.G-----G-----TC-----G-----T-----A-----G-----A-C-G-G-----A-----A-----A-----C-----G-----	1149
SHNH4	.G-----G-----TC-----GG-----T-----AA-----T-----T-----G-G-----G-A-A-----A-----C-----G-----	1177
SHNH9	.G-----G-----TC-----G-----T-----A-----A-----T-----C-----G-----A-----A-----A-----C-----G-----	683
STN	.AG-----G-----TC-----G-----T-----G-----G-----A-----A-----C-----C-----G-----	837
Gag and Gag-Pol P16 Matrix start M G A R N S V L R G K K A D E L E K V R L R P G G K K X Y R L K H I V W A A N E L D K		
Gag, Gag-Pol		

BBM	TTGGGATTGGCAGAGAGCCCTGTTGGGATCAAAAGAGGTTGCCAAGAGTTCTCAGAGTTTATAGTCCATTAGTACCAACAGGGTCAAGAAATTTAANAAGCCTTTTAAATACCGTCTGGGTCATTGGTGCTTCGACGCGAGAGASAAA	1381
BBN	PGLAESESLLESKRGCGQXILRLVLDPLVPTGSEBNLRLSLFNTVGVIMCLHAEERK	442
BBN	SDMQRAACWSSQKRVAKRPFSEFSLHSYOOOGOKT9RAFLIPSSASFGACTOKR	410
BBN	IRIQRERPVGVKRLRLPDDSDSFRSISSTNRVRKFRKXPF4YRLRLMLVLARRRR	423
GH1	-----A-----T-A-----A-----C-A-----	822
D194	-----A-----A-----A-----T-A-----A-----	825
UC2	-----A-----A-----A-----C-----A-----G-----T-----T-----T-----A-A-----	1381
ROD	-----A-----A-----A-----G-----T-----A-----T-C-----A-G-----G-----T-----T-----T-----A-A-----	824
NIEZ	-----A-----A-----A-----T-----C-----C-----G-----T-----T-----C-----A-A-----	825
ISY	-----AA-----A-----A-----TTC-----A-----A-----TAC-----	825
ST	-----G-----A-----T-C-----G-----T-----T-----TA-A-----	826
CAM2	-----A-----A-----GA-----T-A-----G-----T-----A-----A-----A-A-----	1390
MDS	-----A-----A-----A-----O-----T-----C-----C-----G-----T-----T-----TT-----A-A-----G-----	825
RA	-----A-----A-----A-----A-----T-C-----C-----G-----A-----T-----T-----A-A-----	1349
UC1	-----T-----A-----A-----A-----T-C-----A-----T-CG-----C-----C-----A-----T-----AC-----	1362
D205	-----T-----A-----A-----A-----T-----A-----T-A-----C-----C-----T-----TT-----C-----	1357
EHO	-----T-----A-----G-----GA-----A-----GG-----A-----G-----A-----T-----A-----C-----TT-----C-----	1358
MM251	-----T-----A-----A-----AAC-----A-----T-----A-----TTC-----C-----C-----G-----C-----C-----A-----T-----C-----A-----T-----	1319
MM33H	-----T-----A-----A-----AAC-----A-----T-----A-----TTC-----C-----C-----G-----C-----C-----A-----T-----C-----A-----T-----	1322
MM1A11AA	-----T-----A-----A-----AAC-----A-----T-----A-----TTC-----C-----C-----G-----C-----C-----A-----T-----C-----A-----T-----	1330
MM142	-----T-----A-----A-----AAC-----A-----T-----A-----TTC-----C-----C-----G-----C-----C-----A-----T-----C-----A-----T-----	829
MM239	-----T-----A-----A-----AAC-----A-----T-----A-----TTC-----C-----C-----G-----C-----C-----A-----T-----C-----A-----T-----	1331
MNE	-----T-----A-----A-----AAC-----A-----T-----A-----TTC-----C-----C-----G-----C-----C-----A-----T-----C-----A-----T-----	811
SHMPBJ14.441	-----T-----A-----A-----AAC-----A-----T-----A-----TTC-----C-----C-----G-----C-----C-----A-----C-----T-----C-----A-----T-----	1106
SHMPBJ14.15	-----T-----A-----A-----AAC-----A-----T-----A-----TTC-----C-----C-----G-----C-----C-----A-----C-----T-----C-----A-----T-----	1107
SHMPBJ6.6	-----T-----A-----A-----AAC-----A-----T-----A-----TTC-----C-----C-----G-----C-----C-----A-----C-----T-----C-----A-----T-----	1299
SHMH4	-----T-----A-----A-----TAC-----T-----A-----TTC-----C-----C-----G-----T-----C-----G-----A-----T-----C-----A-----T-----	1327
SHMH9	-----T-----A-----A-----AAC-----T-----A-----TTC-----C-----C-----G-----T-----C-----G-----A-----C-----T-----C-----A-----T-----	813
STM	-----T-----A-----A-----G-----T-----A-----A-----CG-----C-----A-----G-----T-----T-----T-----TA-----C-----	987
P16 Matrix	PGLAESESLLESKRGCGQXILRLVLDPLVPTGSEBNLRLSLFNTVGVIMCLHAEERK	Gag, Gag-Pol
BBN	GTGAAGATACTGAGGAGCAGAGAACTAGACACAGAGACATCTAGTGGCAGAACTGGCACTGCAGAGAAATGCCAANTACAGTAGACCAACAGAGACACCTAGTGGGAAAGA...CGAAACTAG...CCCCGTCACAGCGGGT	1525
BBN	V_K_D_T_B_E_A_K_K_L_A_Q_R_L_V_A_E_T_G_T_A_E_K_M_P_N_T_S_R_P_T_A_P_F_S_G_X_R...G_N_Y...P_V_O_O_A_G	490
BBN	S_K_I_L_R_X_Q_R_N_S_M_R_D_T_S_M_G_R_L_C_Q_I_Q_V_D_Q_Q_H_L_V_G_K_F_E_T_E_I_P_C_M_R_R_V	453
BBN	S_E_R_V_S_G_S_M_E_T_E_T_E_T_S_M_G_R_N_Y_N_C_R_B_N_A_V_X_T_N_S_V_T_S_M_E_K_F_E_T_E_I_P_C_M_R_R_V	466
GH1	-----A-----A-----T-----G-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	969
D194	-----A-----A-----A-----T-----G-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	970
UC2	-----A-----G-----A-----T-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	1525
ROD	-----A-----G-----AC-----A-----TG-----G-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	968
NIEZ	-----A-----G-----AC-----A-----T-----G-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	961
ISY	-----A-----A-----A-----G-----T-----GG-----A-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	969
ST	-----C-----A-----AC-----AAG-----T-----A-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	970
CAM2	-----A-----A-----A-----G-----A-----T-----C-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	1534
MDS	-----A-----A-----A-----GT-----C-----G-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	969
RA	-----A-----G-----AC-----A-----T-----A-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	1493
UC1	-----A-----A-----A-----GA-----C-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	1491
D205	-----A-----A-----A-----GA-----C-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	1486
EHO	-----A-----A-----A-----GA-----C-----A-----G-----C-----G-----A-----A-----CT-----C-----GTACAA-----GAC-----GC-----	1487
MM251	-----C-----C-----AC-----GA-----TG-----C-----TG-----A-----A-----A-----CT-----A-----A-----C-----G-----G-----T-----A-----A-----ATA-----	1463
MM32H	-----C-----C-----AC-----GA-----TG-----C-----TG-----A-----A-----A-----CT-----A-----A-----C-----G-----G-----T-----A-----A-----ATA-----	1476
MM1A11AA	-----C-----C-----AC-----GA-----TG-----C-----TG-----A-----A-----A-----CT-----A-----A-----C-----G-----G-----T-----A-----A-----ATA-----	1474
MM142	-----C-----C-----AC-----GA-----TG-----C-----TG-----A-----A-----A-----CT-----A-----A-----C-----G-----G-----T-----A-----A-----ATA-----	973
MM239	-----C-----C-----AC-----GA-----TG-----C-----TG-----A-----A-----A-----CT-----A-----A-----C-----G-----G-----T-----A-----A-----ATA-----	1475
MNE	-----C-----C-----AC-----GA-----TG-----C-----TG-----A-----A-----A-----CT-----A-----A-----C-----G-----G-----T-----A-----A-----ATA-----	955
SHMPBJ14.441	-----C-----C-----AC-----A-----TG-----TG-----A-----A-----A-----C-----GCA-----C-----G-----G-----T-----A-----A-----ATA-----	1250
SHMPBJ14.15	-----C-----A-----AC-----A-----TG-----TG-----A-----A-----A-----C-----OCA-----C-----G-----G-----T-----A-----A-----ATA-----	1251
SHMPBJ6.6	-----C-----A-----AC-----A-----TG-----TG-----A-----A-----A-----C-----OCA-----C-----G-----G-----T-----A-----A-----ATA-----	1443
SHMH4	-----C-----A-----AC-----A-----TG-----TG-----A-----A-----A-----C-----GCA-----C-----G-----G-----T-----A-----A-----ATA-----	1471
SHMH9	-----C-----A-----AC-----A-----TG-----TG-----A-----A-----A-----C-----OCA-----C-----G-----G-----T-----A-----A-----ATA-----	957
STM	-----C-----A-----C-----G-----G-----T-----A-----T-----G-----A-----C-----TGCC-----C-----G-----G-----T-----A-----A-----ATA-----	1131
P16 Matrix	V_K_D_T_B_E_A_K_K_L_A_Q_R_L_V_A_E_T_G_T_A_E_K_M_P_N_T_S_R_P_T_A_P_F_S_G_X_R...G_N_Y...P_V_O_O_A_G	141

Gag and Gag-Pol p16 Matrix end / \ p24 Capsid start

HIV2/SIV Complete Genomes

GEN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	
-----	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	--

BEN	ACAAGCAGCAGTAGAGACAGATCCAGTGGATGTATAGGCCACAAATCCCGTACCGGTAGGGACACATCTACAGAAAGATGATCCAAATAGGGCTCCAAAAGTGTCTCAGAAAGTACACCCCACTAACATCTTAGACATAAAACAGGGA	1972
BEN	T S T V D S O I O W M Y R P O N F V P V G N I Y R R W I Q I G L O K C V R R X Y N P T M I L D I K Q G	639
BEN	O A O S M N R S S G C I Q M K I P V R S G T S T E D G S J S O C K S V S S S T P O L T S S T S M R D	589
BEN	N K M E R S T D P V D V S A T K S R P Q R R R I O K M D P N R A A R V C O K V O P W S M L R R T G	605
GH1	-----A-----G-----G-----G-----G-----A-G-----G-T-----G-----	1416
D194	-----A-C-----G-----G-----G-----G-----A-G-A-----G-T-----TG-G-----	1417
UC2	-----G-----AG-----G-----G-----G-----G-----G-----G-----G-----G-----	1972
ROD	-----A-----T-----T-----A-----A-----T-----G-----AT-----G-----G-T-----G-C-----C-----G-----	1415
NINZ	-----A-G-----T-----A-----A-----C-----T-----G-----T-----G-----C-----T-----C-----T-----A-----	1410
ISY	-----G-----G-A-----A-G-----T-----A-----A-----T-----G-----G-----A-----G-----T-----T-----C-----TC-----A-----	1416
ST	-----A-G-----T-----T-----A-----A-G-----T-----G-----G-----A-G-----C-----G-T-----C-----C-----	1417
CAM2	-----A-----A-----T-----G-----A-----A-----T-----G-----G-----G-----G-----G-----G-----C-----C-----	1591
NDS	-----A-----TC-----G-----G-----T-----A-----A-G-A-----T-----G-----G-----G-----G-----T-----C-----C-----G-----	1416
KR	-----A-----T-----AG-----T-----A-----A-----T-----G-----G-----A-----G-----C-----G-----T-----C-----C-----G-----	1940
UC1	-----C-----G-----A-A-----C-----G-----T-----T-----T-----T-----T-----T-----G-----A-----C-----G-----T-----T-----C-----A-----G-----	1938
D205	-----C-----G-----A-A-----C-----G-----C-----T-----C-----A-----G-----A-----T-----T-----T-----T-----AT-----G-----A-----C-----T-----T-----C-----A-----G-----	1933
ERO	-----C-----A-----A-----C-----A-----C-----T-----C-----A-----G-----A-----T-----G-----T-----GT-----A-----G-----A-----C-----G-----T-----T-----T-----T-----AC-----G-----G-----A-----G-----	1934
HN251	-----T-----TT-----A-----C-----A-----A-----G-----C-----A-----A-----C-----T-----G-----C-----G-----T-----A-----A-----T-----T-----A-----TC-----TG-----A-----G-----	1907
HN32H	-----T-----TT-----A-----C-----A-----A-----G-----C-----A-----A-----C-----T-----G-----C-----G-----T-----A-----A-----T-----T-----A-----TC-----TG-----A-----G-----	1920
HN1A11AA	-----T-----TT-----A-----C-----A-----A-----G-----C-----A-----A-----C-----T-----G-----C-----G-----T-----A-----A-----T-----T-----A-----TC-----TG-----A-----G-----	1918
HN142	-----T-----T-----A-----A-----C-----A-----A-----G-----C-----A-----A-----C-----T-----G-----C-----G-----T-----A-----A-----T-----T-----A-----TC-----TG-----A-----G-----	1417
HN239	-----T-----TT-----A-----C-----A-----A-----G-----C-----A-----A-----C-----T-----G-----C-----G-----T-----A-----A-----T-----T-----A-----TC-----TG-----A-----G-----	1919
MNE	-----T-----T-----A-----A-----C-----A-----A-----G-----C-----A-----A-----C-----T-----G-----C-----G-----T-----A-----A-----T-----T-----A-----TC-----TG-----A-----G-----	1398
SHNFBJ14.441	-----T-----C-----A-----T-----A-----AG-----C-----TA-----A-----T-----G-----G-----C-----G-----A-----A-----T-----T-----T-----A-----T-----G-----TG-----G-----A-----	1697
SHNFBJ14.15	-----T-----C-----A-----T-----A-----AG-----C-----TA-----A-----T-----G-----G-----C-----G-----A-----A-----T-----T-----T-----A-----T-----G-----TG-----G-----A-----	1699
SHNFBJ6.6	-----T-----C-----A-----T-----A-----AG-----C-----TA-----A-----T-----G-----G-----C-----G-----A-----A-----T-----T-----T-----A-----T-----G-----TG-----G-----A-----	1890
SHNH4	-----T-----T-----A-----A-----C-----A-----A-----C-----A-----A-----C-----T-----G-----C-----G-----A-----A-----T-----T-----A-----T-----TG-----G-----A-----	1910
SHNH9	-----T-----S-----C-----A-----T-----A-----Y-----AG-----C-----TA-----A-----C-----R-----G-----T-----C-----G-----A-----C-----A-----T-----T-----T-----A-----T-----G-----TG-----G-----A-----	1404
STN	-----T-----T-----CC-----G-----A-----T-----A-----CC-----AG-----TA-----A-----A-----C-----T-----C-----A-----A-----T-----T-----T-----GT-----T-----G-----G-----A-----G-----	1561
Capaid	T S T V D S O I O W M Y R P O N F V P V G N I Y R R W I Q I G L O K C V R R X Y N P T M I L D I K Q G	Gag, Gag-Pol
BEN	CEAAAGAACCGGTTCCAAAGCTATGTAGACAGGTTCTACAAAGCTTGAGGGCAGAACAAACAGAACCCAGCTAAAAATTTGATGACCCAAACCGCTCTAATACAGAAATGCCAACCCAGAGCTGCAAGTTAGTACTAAAAAGGACTGGGG	2122
BEN	P K E P F O S V V D R F Y K S L R A R O T D P A V K N W M T O T L L I O N A M P D C K L V L X G L G	689
BEN	O K M R S F A N I T G S T K A S G O N K O P O O S K I G S F X R C S Y R M P F O T A S S Y S K D W G	632
BEN	T K R T V P K L C R O V L O K L E G R T M N R P S S K K L D O D P N A A N T E C O P R I Q V S T K R T G	655
GH1	-----G-----G-----G-----T-----C-----A-----A-----G-----G-----G-----G-----A-----GG-----T-----T-----G-----A-----	1566
D194	-----T-----G-----G-----A-----A-----C-----A-----A-----G-----G-----G-----G-----A-----GG-----T-----T-----G-----A-----	1567
UC2	-----T-----A-----G-----T-----A-----A-----G-----G-----G-----G-----A-----GG-----T-----T-----G-----A-----	2122
ROD	-----G-----G-----T-----A-----T-----A-----T-----G-----G-----A-----G-----A-----T-----T-----A-----G-----A-----	1565
NINZ	-----G-----A-----T-----A-----T-----A-----T-----G-----G-----A-----G-----A-----T-----T-----A-----G-----A-----	1560
ISY	-----GT-----G-----T-----A-----T-----A-----T-----G-----G-----T-----G-----G-----A-----T-----G-----T-----A-----G-----A-----	1563
ST	-----G-----G-----A-----G-----T-----A-----T-----G-----G-----T-----A-----G-----G-----A-----G-----G-----T-----A-----A-----	1567
CAM2	-----G-----GT-----A-----T-----A-----T-----A-----T-----G-----G-----A-----G-----G-----T-----A-----G-----T-----A-----	2131
NDS	-----G-----G-----T-----A-----T-----A-----T-----G-----G-----T-----G-----A-----A-----G-----A-----T-----A-----G-----T-----A-----	1966
KR	-----G-----G-----T-----A-----T-----A-----C-----A-----A-----G-----G-----T-----G-----A-----A-----GG-----T-----A-----A-----	2090
UC1	-----T-----T-----C-----T-----A-----AC-----G-----G-----G-----A-----A-----T-----T-----T-----T-----A-----T-----G-----CT-----A-----A-----	2088
D205	-----G-----G-----C-----T-----A-----AC-----G-----G-----G-----A-----A-----G-----T-----T-----T-----T-----G-----T-----G-----CT-----A-----A-----	2083
ERO	-----C-----C-----T-----A-----AC-----G-----G-----G-----A-----A-----G-----T-----T-----T-----T-----A-----G-----T-----G-----CT-----A-----A-----	2084
HN251	-----G-----A-----T-----G-----T-----A-----A-----TG-----G-----G-----T-----A-----G-----T-----A-----T-----T-----C-----G-----G-----G-----T-----	2057
HN32H	-----G-----A-----T-----G-----T-----A-----A-----TG-----G-----G-----T-----A-----G-----T-----A-----T-----T-----C-----G-----G-----G-----T-----	2070
HN1A11AA	-----G-----A-----T-----G-----T-----A-----A-----TG-----G-----G-----T-----A-----G-----T-----A-----T-----T-----C-----G-----G-----G-----T-----	2068
HN142	-----G-----A-----T-----G-----T-----A-----A-----TG-----G-----G-----T-----A-----G-----T-----A-----T-----T-----C-----G-----G-----G-----T-----	1557
HN239	-----G-----A-----T-----G-----T-----A-----A-----TG-----G-----G-----T-----A-----G-----T-----A-----T-----T-----C-----G-----G-----G-----T-----	2069
MNE	-----G-----A-----T-----G-----T-----A-----A-----TG-----G-----G-----T-----A-----G-----T-----A-----T-----T-----C-----G-----G-----G-----T-----	1549
SHNFBJ14.441	-----G-----A-----T-----T-----A-----G-----TC-----A-----A-----G-----G-----G-----T-----A-----G-----T-----A-----G-----T-----A-----G-----G-----C-----G-----T-----T-----	1847
SHNFBJ14.15	-----G-----A-----T-----T-----A-----G-----CA-----A-----T-----G-----G-----G-----T-----A-----G-----T-----A-----G-----T-----A-----G-----G-----C-----G-----T-----T-----	1848
SHNFBJ6.6	-----G-----A-----T-----T-----A-----G-----CA-----A-----T-----G-----G-----G-----T-----A-----G-----T-----A-----G-----T-----A-----G-----G-----C-----G-----T-----T-----	2040
SHNH4	-----G-----A-----T-----C-----T-----A-----TC-----A-----A-----G-----T-----C-----G-----G-----T-----A-----G-----T-----A-----T-----AC-----G-----C-----G-----T-----C-----	2058
SHNH9	-----GT-----A-----T-----G-----T-----A-----G-----TC-----A-----A-----G-----S-----G-----T-----T-----A-----G-----T-----A-----G-----G-----C-----G-----T-----Y-----	1554
STN	-----A-----T-----T-----G-----T-----A-----G-----T-----A-----G-----T-----G-----G-----A-----T-----A-----T-----T-----A-----G-----T-----G-----T-----C-----	1731
Capaid	P K R P P O S V V D R F Y K S L R A R O T D P A V K N W M T O P L L I O N A M P D C K L V L X G L G	Gag, Gag-Pol

HIV2/STV Complete Genomes

BEN	ATGAATCCCACTAGAGAGATGCTAACCGGCTGCCAGGGGCTAGGCGGACCAAGCCGGAAGCCAGGCTAATGGCTGAAGCCCTAAAGAGGCTATGGGACCAAGCCCTATCCCATTTGACAGCAGCCCAAGAA	2280
BEN	N M P T L E E N L T A C O Q V G G P G O X A R L N A E A L K E A N G P S P I P F A A A Q Q	735
BEN	S I P P S K R C S P P A R G S A D Q A R R P G S M L R P S K R L W D Q A L S H L Q Q P N I	671
BEN	D E S M P R R D A M R L P F G G R R T R P F E S Q A N G S S P X R G V G T X P V P I C S S P T W	699
GH1	-----G--T--T-----T-----A-----A-----G--G--T--AC--CC-----	1704
DL94	-----G--G--T-----A-----A-----A-----G--T--ACG--GC-----	1705
UC2	-----G--G--T-----A-----A-----A-----G--T--AAC--CC-----	2280
RQD	---C--T--T-----G-----T-----T--G-----T--AT-----A--G--G-----TC--A--TC-----C-----G--G-----	1703
NIH2	---T--T--T-----G--A--T-----T-----G-----T--AT-----A-----G--TC--C-----GC--C-----G--G-----	1698
ISY	---T--T--T-----T--A--T-----T-----G-----T--AT-----A-----T-----C--C-----CC-----C-----G-----	1701
ST	---A--T--T-----A-----T-----T-----T-----AT-----A-----T-----G--C--C-----CC-----C-----G-----	1705
CAN2	---T--T--T-----T-----T--A--A--T-----T-----A-----A-----T-----G--C-----CC-----C-----G--G-----	2269
ND8	---T--T--T-----T-----T--A--A--T-----T-----A-----A-----G--T-----G--C--AC--GC-----C-----G--G-----	1704
KR	---T--T--G-----G-----T-----T--AA--A--T-----T-----A-----G-----G--CC--A--C-----CC-----G-----	2228
UC1	---T--G--A--G-----G--A--A--A--C--G-----G--A-----G--T-----G--C--AC--GC-----C-----G--G-----	2238
D205	---T--G--A-----G-----A--A--A--C--G-----G--A-----C-----T-----CC--AAC--TCGA--C--A--G-----T--C--T-----AAAGCAGGAAAG	2233
EHO	---C--AT-----A-----A--A--A--C--G-----G--A-----G--TT-----A--CF--AC--TTC--A--C--AT--G-----C--C--T-----C--AGAGCAGG--AG	2231
NIH251	G-----A-----G--G--T--T--A--A--G--C--A-----G--T--AT-----A-----G-----CC--C--C-----GTG--A-----T-----GA--GAGGGGACCA	2204
NIH32H	G-----A-----G--G--T--T--A--A--G--C--A-----G--T--AT-----A-----G-----CC--C--C-----GTG--A-----T-----GAGGGGACCA	2217
NIH1A11AA	G-----A-----G--G--T--T--A--A--G--C--A-----G--T--AT-----A-----G-----CC--C--C-----GTG--A-----T-----GAGGGGACCA	2215
NIH142	G-----A-----G--G--T--T--A--A--G--C--A-----G--T--AT-----A-----G-----CC--C--C-----GTG--A-----T-----GAGGGGACCA	1710
NIH239	G-----A-----G--G--T--T--A--A--G--C--A-----G--T--AT-----A-----G-----CC--C--C-----GTG--A-----T-----GAGGGGACCA	2216
NIH6	-----A-----G--G--T--T--A--A--A--A--G--A--A--G--A--AT-----A-----G-----CC--T--C-----G--G--AC--T-----GAGGGGACCA	1696
SHNFBJ14.441	---A--TT-----A--T--G--A--T-----T--A--A--A--G--A--A--T--A-----A--G--AT-----T--AC--ACG--A--G--G--AC-----T-----GAAGGACAA	1994
SHNFBJ14.15	---A--TT-----A--T--G--A--T-----T--A--A--A--G--A--A--T--A-----A--G--AT-----T--AC--ACG--A--G--G--AC-----T-----GAAGGACAA	1995
SHNFBJ6.6	---A--TT-----A--T--G--A--T-----T--A--A--A--G--A--A--T--A-----A--G--AT-----T--AC--ACG--A--G--G--AC-----T-----GAAGGACAA	2187
SHNH4	---A--TT-----A--T--G--A--T-----T--A--A--A--G--A--A--T--A-----A--G--AT-----T--AC--ACG--A--G--G--AC-----T-----GAAGGACAA	2215
SHNH9	---A--TT-----A--T--G--A--T-----T--A--A--A--G--A--A--T--A-----A--G--AT-----T--AC--ACG--A--G--G--AC-----R-----T-----GAAGGACAA	1701
STW	---A--TT-----A--T--G--A--T-----T--A--A--A--G--A--A--T--A-----A--G--AT-----T--AC--ACG--A--G--G--AC-----C-----GACAGGA	1875
Capsid	K M P T L E E N L T A C O Q V G G P G O X A R L N A E A L K E A N G P S P I P F A A A Q Q	Gag, Gag-Pol

Gag and Gag-Pol p24 Capsid and / \ X peptide start

X peptide end / \ Gag Nucleocapsid start

\ Gag-Pol TF Start

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

See introduction for ribosomal slip references

Pol p15 TF and / Protease start

I-99
DEC 97

HIV2/SIV Complete Genomes

	c-- Gag cds end	
BEH	CTCTTTGGAAAGACCACTAGTCACAGCGGTACATCTGAGGATCAGCCGCTAGAGTCTTACTAGACACAGGGCTGATGTCATCAATAGTAGCAGGAATAGATTAGGGGACAAATTACCTCCAAATATAGTAGGGGGAATAGGGGGATTATTA	2797
BEH	L P Q K D Q S S Q R T S R I S R S K S Y S T Q Q L M T Q S S Q E S X S Q T I T L Q K S S G E S Q D L	904
BEH	S L E K T S S M S V M A G S A G R S L T R N R G S S L N S S R N R I R G Q L N S K M S R Q N R G I Y	841
BEH	S L M K R P V V T A V T E D Q P V E V L L D T G A D D S I V A G I E L G D N Y T P K I V G G I G G F	872
QHL	-----A-----TA-----T-----G-----C-----G-----GT-----G-----C-----	2241
D194	-----A-T-----A-----G-----G-----G-----G-----G-----C-----	2242
UC2	-----A-----T-----G-----A-----GT-----C-----A-----T-GC-----G-----C-----	2797
ROD	-----A-----T-----G-----A-----GT-----C-----A-----T-GC-----G-----C-----	2243
NIN2	-----AC-----T-----G-----A-----GT-----C-----A-----T-GC-----G-----C-----	2238
ISY	-----C-----A-----T-----G-----A-----GT-----C-----A-----T-GC-----G-----C-----	2238
ST	-----AC-TG-T-----GC-----A-----T-----G-----C-----G-----G-----G-----C-----	2242
CAN2	-----AC-T-----G-----A-----T-----GT-----C-----A-----T-GC-----G-----C-----	2806
MDS	-----C-----T-----AC-T-----GA-----A-----T-----G-----G-----T-----A-----T-GC-----G-----C-----	2241
KR	-----A-----TG-T-----G-----A-----G-----C-----G-----G-----AG-----T-G-----C-----	2765
UC1	-----G-----Q-----A-GT-----G-----A-----G-----A-----C-----T-----G-----C-----G-----C-----	2778
D205	-----G-----A-----A-GT-----G-----A-----A-----A-----T-----C-----G-----A-----G-----C-----	2773
EHO	-----G-----A-----AACT-----T-----G-----AT-----A-----A-----A-----G-----C-----C-----T-----G-----A-----	2768
MM251	-----GG-----T-----TC-T-----T-----A-----GA-----T-----A-----T-----G-----T-----G-----T-----C-----C-----	2690
MM22H	-----GG-----T-----TC-T-----T-----A-----GA-----T-----A-----T-----T-----A-----G-----T-----C-----C-----	2715
MM1A11AA	-----GG-----T-----TC-T-----T-----A-----GA-----T-----A-----G-----T-----T-----T-----A-----G-----T-----C-----C-----	2713
MM142	-----GG-----T-----TC-T-----T-----A-----GA-----T-----A-----T-----T-----A-----G-----T-----C-----C-----	2200
MM239	-----GG-----T-----TC-T-----T-----A-----GA-----T-----A-----G-----T-----T-----T-----A-----G-----T-----C-----C-----	2714
MMB	-----G-----T-----TC-T-----T-----A-----GA-----T-----C-----A-----T-----T-----T-----T-----G-----T-----C-----C-----	2182
SHMPBJ14.441	-----G-----A-----G-----C-----T-----A-----A-----C-----A-----T-----T-----T-----G-----T-----G-----T-----C-----T-----	2480
SHMPBJ14.15	-----G-----A-----G-----C-----T-----A-----A-----C-----A-----T-----T-----T-----G-----T-----G-----T-----C-----T-----	2481
SHMPBJ6.6	-----G-----A-----G-----C-----T-----A-----A-----C-----A-----T-----T-----T-----G-----T-----G-----T-----C-----T-----	2673
SHMH4	-----G-----T-----C-----T-----A-----A-----CA-----A-----T-----T-----T-----G-----T-----G-----T-----C-----T-----	2701
SHMH9	-----G-----T-----C-----T-----A-----A-----CA-----A-----T-----T-----T-----G-----T-----G-----T-----C-----T-----	2187
STH	-----G-----T-----C-----T-----T-----G-----T-----AC-T-----G-----C-----G-----G-----G-----GC-----TCAC-----A-----C-----G-----	2361
Gag p6	L F G K D Q S Gag p6 CDS end	
Protease	S L M K R P V V T A V T E D Q P V E V L L D T G A D D S I V A G I E L G D N Y T P K I V G G I G G F	Pol
BEH	TAAACACCAAGAGATACAAAATATAGAAATTAAGTACTAAATAAAAGAGTANGAGGCCACATTAATGACAGGAGATACCCCAATCAACATCTTTGGCAGGAATATTCTGACAGCCTTAGGCATGTCAATTAATACAGTTCACAGGA	2947
BEH	S L P K M P K M S K S K V S I K E S E P P S S Q E I P Q S T S L A E I F S Q P S A C H S I Y R L P R	944
BEH	K H Q R I Q K C R R S T K S K S K S H H N D R R Y P N Q M L M Q R K V S D S L R H V I X F T S C Q D	890
BEH	I N T K E Y T M V E I X V L N K R V R A T I M T G D T P I M I F Q R N T L F A L G M S L N L P V A R	922
GM1	-----GA-----T-----A-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2391
D194	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2392
UC2	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2947
ROD	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2393
NIN2	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2385
ISY	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2386
ST	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2392
CAN2	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2956
MDS	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2391
KR	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2915
UC1	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2928
D205	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2923
EHO	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2918
MM251	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2840
MM22H	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2865
MM1A11AA	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2863
MM142	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2350
MM239	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2864
MMB	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2332
SHMPBJ14.441	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2630
SHMPBJ14.15	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2631
SHMPBJ6.6	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2823
SHMH4	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2851
SHMH9	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2337
STH	-----T-----T-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----G-----	2511
Protease	I N T K E Y T M V E I X V L N K R V R A T I M T G D T P I M I F Q R N T L F A L G M S L N L P V A R	Pol

Pol Protease end / RT start

HTV2/STV Complete Genomes

BEN TTAGAGCTCAATAAAGTATACATTGAAGCCAGGGGAAGATGGACCAAGGCTGAACAATATGGCCCTTAACAAAGAGAGAAATAGAGAGCACTAAAGAGATCTGTGAAAAAATGGAAAAAGAGGGCCAGCTAGAGAGAGGCACTCCACATAATC
 BEN S S O S S M S O G G M D O G G S M D G P S O K R K I K K S K R S V K K M K K R A S S K R H L Q L T
 BEN R A M K S N T A R G R R N W T K S I T H A P N K R E N R S T K R D L I K N G K R G P A R G C T S K S S
 BEN I R P I R V V L K P G K D G P R L K O W P L T K E R I E A L K E I C E K M E K E G O L E E A P P T N
 GN1 -A-----G-----A-----G-----G-----C-----G-----A-----
 D194 -A-----A-----A-----A-----A-----A-----T-----GG-----A-----A-----
 UC2 -A-----A-----A-----C-----A-----A-----G-----T-----G-----A-----A-----
 ROD -A-----A-----TGC-A-----AA-----G-----T-----A-----A-----A-----G-----A-----
 NIH2 -A-----A-----TGC-A-----A-----A-----TT-----A-----A-----A-----A-----A-----
 ISY -A-----G-----A-----A-----G-----AA-AA-G-----T-----G-----A-----G-----A-----T-----A-----G-----
 ST -A-----A-----A-----TGC-----A-----G-----AA-----G-----T-----A-----A-----G-----G-----G-----A-----G-----A-----G-----
 CAM2 -A-----A-----A-----TGC-A-----A-----AT-----G-----T-----G-----A-----A-----A-----G-----A-----G-----A-----
 HD5 -A-----A-----A-----TGC-A-----A-----G-----AA-----G-----G-----TT-----G-----A-----A-----G-----A-----G-----A-----
 KR -C-----G-----T-----C-----A-----A-----A-----A-----G-----A-----G-----A-----G-----A-----G-----A-----G-----A-----T-----C-----
 UC1 -A-----G-----C-----A-----C-----A-----T-----A-----A-----C-----G-----T-----T-----C-----A-----G-----T-----G-----C-----C-----G-----A-----G-----A-----G-----C-----T-----
 D205 -A-----G-----TGGG-----A-----A-----T-----A-----G-----A-----A-----C-----G-----T-----T-----C-----GG-----A-----G-----CT-----C-----C-----A-----G-----A-----AT-----A-----C-----T-----
 ERG -A-----G-----CCAG-----A-----T-----AA-----G-----A-----AA-----C-----G-----T-----C-----C-----CT-----C-----C-----A-----A-----T-----G-----T-----G-----T-----T-----G-----G-----A-----T-----C-----G-----C-----
 NM251 -TG-----CG-----C-----A-----A-----G-----T-----AAT-----G-----G-----AT-----T-----A-----G-----TT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----G-----A-----T-----C-----G-----C-----
 NM32H -TG-----CG-----C-----A-----A-----G-----T-----AAT-----G-----G-----AT-----T-----A-----G-----TT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----G-----A-----T-----C-----G-----C-----
 NM1A11AA -TG-----CG-----C-----A-----A-----T-----AAT-----G-----G-----AT-----T-----A-----G-----TT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----G-----A-----T-----C-----G-----C-----
 NM142 -TG-----GTCOC-----C-----A-----A-----G-----AAT-----G-----G-----AT-----T-----A-----G-----TT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----G-----A-----T-----C-----G-----C-----
 NM239 -TG-----CG-----C-----A-----A-----G-----AAT-----G-----G-----AT-----T-----A-----G-----TT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----G-----A-----T-----C-----G-----C-----
 NME -TG-----C-----C-----A-----A-----A-----AAT-----GG-----G-----AT-----T-----A-----G-----TT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----G-----A-----T-----C-----G-----C-----
 SHMPBJ4.441 -T-----C-----A-----A-----A-----AAT-----A-----G-----G-----G-----T-----A-----G-----ATT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----A-----C-----C-----C-----
 SHMPBJ4.15 -T-----C-----A-----A-----A-----AAT-----A-----G-----G-----G-----T-----A-----G-----ATT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----A-----C-----C-----C-----
 SHMPBJ6.6 -T-----T-----C-----A-----A-----A-----AAT-----A-----G-----G-----G-----T-----G-----A-----G-----ATT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----A-----C-----C-----C-----
 SHMR4 -G-----T-----C-----A-----A-----A-----AAT-----A-----G-----G-----G-----T-----A-----G-----ATT-----T-----G-----A-----G-----G-----T-----T-----T-----G-----A-----C-----C-----C-----
 SHMR9 -G-----R-----T-----C-----A-----A-----A-----AAT-----A-----G-----G-----G-----P-----G-----A-----G-----ATT-----T-----G-----A-----G-----G-----T-----T-----G-----R-----A-----T-----C-----C-----
 STM -A-----G-----G-----V-----A-----A-----A-----AAT-----A-----G-----G-----AT-----T-----A-----G-----G-----G-----T-----G-----A-----A-----T-----A-----A-----
 Pol RT I R P I R V V L K P G K D G P R L K O W P L T K E R I E A L K E I C E K M E K E G O L E E A P P T N

GEN	CTTATAATACCCACATTTGCAATTAGAGAAAGGACAMBAACAAATGGAGGATGCTGATAGATTTTAGAGAACTAAATAGGTGACTCAAGATTTCACAGAAATTCAGCTAGGAATTCACACCCCGCAGGACTAGCCAAAGAA
HEX	L L I F P H L Q L L K R T R T N G G C G G I L E M S I R S L K I S Q K F S S R F H T P Q D S Q K R K
HEX	L S Y P H I C H L E K R Q Q G Q M G C A D D R F I R T M K I S G D S R F M R N S S A R S T P Q R T S Q K R K
HEX	P Y N F P T F A I T K R R D R N K H R M L I D F R R I N K V T Q D F T F Y O L G I F H P A G L A K K K
GH1	-----C-----CG-----T-----
D194	-----A-----A-----CG-----T-----
UC2	-----C-----A-----A-----CG-----A-----A-----GT-G-----G-----G-----
ROD	-----C-----A-----A-----C-----G-----C-----A-----G-----T-----A-----T-G-----G-----G-----
NIH2	-----C-----A-----A-----A-----C-----G-----G-----C-----A-----G-----T-----A-----T-G-----G-----G-----
ISV	-----C-----A-----A-----A-----C-----G-----C-----A-----A-----G-----G-----T-----A-----T-----G-----A-----G-----
ST	-----C-----A-----A-----A-----A-----C-----A-----C-----A-----C-----T-----A-----G-----G-----C-----
CAN3	-----C-----G-----A-----A-----A-----G-----C-----A-----A-----G-----C-----T-----A-----T-----G-----G-----
MD9	-----C-----G-----A-----T-----A-----A-----T-----C-----A-----G-----G-----T-----A-----G-----T-----G-----G-----
KR	-----A-----C-----C-----A-----C-----A-----A-----G-----GA-----T-----A-----A-----A-----C-----A-----C-----G-----T-----A-----T-----G-----A-----G-----
UC1	-----A-----C-----C-----A-----C-----C-----A-----A-----G-----GA-----T-----A-----A-----A-----C-----T-----C-----C-----G-----C-----A-----G-----T-----C-----GT-G-----AG-----A-----GG-----
D205	-----A-----C-----C-----A-----C-----C-----A-----G-----A-----T-----A-----A-----A-----C-----GT-----C-----A-----C-----G-----C-----G-----A-----G-----T-----C-----G-----AG-----A-----GG-----
KHO	-----A-----C-----T-----G-----C-----C-----A-----A-----G-----A-----A-----A-----A-----C-----C-----A-----C-----A-----T-----GG-----C-----G-----T-----T-----A-----G-----ATC-----
NH251	-----A-----C-----C-----C-----A-----A-----A-----T-----A-----A-----A-----G-----G-----C-----G-----C-----T-----G-----C-----AT-----A-----T-----A-----G-----G-----
NH32H	-----A-----C-----C-----T-----A-----A-----A-----T-----A-----A-----A-----G-----G-----C-----G-----C-----T-----G-----C-----AT-----A-----T-----A-----G-----G-----
NH1A11AA	-----A-----C-----C-----T-----A-----A-----A-----T-----A-----A-----A-----G-----G-----C-----G-----C-----T-----G-----C-----AT-----A-----T-----A-----G-----G-----
NH142	-----A-----C-----C-----T-----A-----A-----A-----T-----A-----A-----A-----G-----G-----C-----G-----C-----T-----G-----G-----C-----AT-----A-----T-----A-----G-----G-----
NH1239	-----A-----C-----C-----T-----A-----A-----A-----T-----A-----A-----A-----G-----G-----C-----G-----C-----T-----G-----G-----C-----AT-----A-----T-----A-----G-----G-----
NHE	-----A-----C-----C-----C-----A-----A-----A-----G-----A-----A-----A-----G-----G-----C-----G-----C-----T-----G-----C-----AT-----G-----A-----T-----A-----G-----G-----
SHMPBJ14.441	-----A-----C-----C-----T-----A-----A-----G-----A-----T-----A-----A-----A-----T-----C-----C-----T-----T-----G-----A-----T-----G-----A-----T-----A-----G-----GA-----G-----
SHMPBJ14.15	-----A-----C-----C-----T-----A-----A-----G-----A-----T-----T-----T-----T-----C-----C-----T-----T-----G-----A-----T-----G-----A-----T-----A-----G-----GA-----G-----
SHMPBJ6.6	-----A-----C-----C-----T-----A-----A-----G-----A-----T-----T-----A-----A-----C-----C-----T-----T-----G-----A-----T-----G-----A-----T-----A-----G-----GA-----G-----
SHMPH4	-----G-----C-----C-----T-----T-----A-----G-----A-----A-----T-----A-----A-----A-----T-----G-----C-----C-----T-----T-----G-----A-----T-----A-----T-----A-----G-----GA-----G-----
SHMPH9	-----A-----C-----C-----T-----A-----R-----GR-----A-----GGA-----T-----A-----A-----A-----T-----C-----C-----T-----T-----G-----A-----T-----G-----A-----T-----A-----G-----GA-----G-----
STM	-----C-----C-----C-----T-----T-----A-----A-----G-----A-----G-----C-----A-----A-----A-----T-----T-----A-----T-----A-----T-----A-----G-----G-----G-----
Pol RT	P Y N F P T F A I T K K K K D K N K H R M L I D F R R I N K V T Q D F T F Y O L G I F H P A G L A K K K

HIV2/SIV Complete Genomes

BSN	GAATCTCTATATTAGATGATAGGAGATGCTATTTTTCATACCACTACATGAAGATTITAGGAGATATCTCCATTTACCTACCAGCAGTAACCAATATGGAACCAAGAAAAAATATATATATAAAGTCTTGCCACAGAGATGGAAGG	3397
BSN	-S-L-Y-S-M-S-G-H-P-I-F-P-Y-H-Y-M-K-I-L-Q-S-I-L-H-L-F-V-Q-Q-S-T-I-W-N-O-E-R-D-I-V-I-K-S-C-H-K-D-Q-R	1076
BSN	D-L-Y-I-R-C-R-G-C-L-F-S-H-T-T-T-S-R-F-S-A-V-Y-C-L-Y-P-T-S-S-R-Q-Y-G-T-R-K-K-I-Y-E-S-S-L-A-T-R-M-E-G	1031
BSN	R-I-S-I-L-D-V-Q-D-A-Y-F-S-I-P-L-M-E-D-F-R-O-Y-T-A-F-T-L-P-A-V-N-N-M-E-P-G-K-R-Y-T-Y-K-V-L-P-O-G-W-K	1072
QH1	-A-G-C-----C-----G-----T-----GCA-----G-----G-----A-----	2841
DI94	-TA-G-C-----C-----A-----T-----GCA-G-----A-----G-----A-----	2842
UC2	-TA-G-C-----C-----T-----T-----G-A-----C-----G-----A-----	3397
ROD	-A-TA-G-C-----T-C-----G-C-----A-CA-----T-----G-----GCA-----C-----G-----	2843
NH2	-A-TA-G-C-----T-C-----G-C-----A-----C-T-----T-----GCA-----C-----A-----G-----	2835
ISY	-A-TA-G-G-----A-----C-----T-----G-----T-----GCA-----C-----A-----G-----	2838
ST	-A-TA-G-CC-----T-C-----G-----A-----T-----A-----GCT-----C-----CA-----G-----	2842
CAN2	-A-TA-G-C-----T-----T-----C-----A-----C-----T-----T-----GCA-----C-----A-G-G-----A-----	3406
NDS	-A-TA-G-C-----T-C-----G-----C-----A-----C-----T-----G-----GCA-----C-----G-----	2841
KR	-A-TA-G-C-----A-----C-----G-----A-----A-----T-----A-----G-----GCA-----G-----C-A-----G-----A-----	3365
UC1	-A-AA-AG-C-----G-A-----C-CAGT-C-----G-CC-A-C-C-A-----A-----C-T-G-----T-A-----GCA-G-----G-----C-T-----TC-A-----G-----	3378
D205	-A-AA-AG-A-----G-A-----C-CAGT-C-----G-CC-A-C-C-A-----A-----C-T-G-----T-A-----GCA-G-----G-----C-T-----TC-A-----G-----	3370
EMO	-A-AA-AG-C-----A-----C-CAGT-C-----G-CC-A-C-C-A-----A-----C-T-G-----T-A-----GCA-G-----G-----C-T-----TC-A-----G-----	3368
NH251	-TA-AG-C-G-----A-----T-----A-----C-----T-----G-----A-----C-----T-----T-----GCA-G-----C-----C-T-----G-TC-----T-G-----	3290
NH32H	-TA-AG-C-G-----A-----T-----A-----C-----T-----G-----A-----C-----T-----T-----GCA-G-----C-----C-T-----G-TC-----T-G-----	3315
NH1A11AA	-TA-AG-C-G-----A-----T-----A-----C-----T-----G-----A-----C-----T-----T-----GCA-G-----C-----C-T-----G-TC-----T-G-----	3313
NH142	-TA-AG-C-G-----A-----T-----A-----C-----T-----G-----A-----C-----T-----T-----GCA-G-----C-----C-T-----G-TC-----T-G-----	2800
NH239	-A-TA-AG-C-G-----A-----T-----A-----C-----T-----G-----A-----C-----T-----T-----GCA-G-----C-----C-T-----G-TC-----T-G-----	3314
NNE	-A-AG-C-G-----T-----C-----A-----C-----T-----G-----A-----C-----T-----T-----GCA-G-----C-----C-T-----G-TC-----T-G-----	2782
SHNFBJ14.441	-A-AG-G-G-----T-----A-----C-----T-----G-----A-----C-----A-----C-----T-----T-----GCA-G-----C-----C-G-A-A-T-----G-----A-----	3080
SHNFBJ14.15	-A-AG-G-G-----T-----A-----C-----T-----G-----A-----C-----A-----C-----T-----T-----GCA-G-----C-----C-G-A-A-T-----G-----A-----	3081
SHNFBJ6.6	-A-AG-G-G-----T-----A-----C-----T-----G-----A-----C-----A-----C-----T-----T-----GCA-G-----C-----C-G-A-A-T-----G-----A-----	3273
SHNH4	-A-AG-G-G-----T-----A-----C-----T-----G-----A-----C-----A-----C-----T-----T-----GCA-G-----C-----C-G-A-A-T-----G-----A-----	3301
SHNH9	-A-AG-G-G-----T-----A-----C-----T-----G-----A-----C-----A-----C-----T-----T-----GCA-G-----C-----C-G-A-A-T-----G-----A-----	2787
STN	-A-AA-I-L-D-V-Q-D-A-Y-F-S-I-P-L-M-E-D-F-R-O-Y-T-A-F-T-L-P-A-V-N-N-M-E-P-G-K-R-Y-T-Y-K-V-L-P-O-G-W-K	2951
Pol RT	R-I-S-I-L-D-V-Q-D-A-Y-F-S-I-P-L-M-E-D-F-R-O-Y-T-A-F-T-L-P-A-V-N-N-M-E-P-G-K-R-Y-T-Y-K-V-L-P-O-G-W-K	Pol
BSN	GATCACAGCAATTTTTCATACCAATGAGGAAAGTCTTASAACTTTTCAGAAAGCAACCCAGATGTCATTTCTATCCAGTACATGATGATATCTTAAATAGCTAGTACAGGACAGGTTTAGAGCATGACAAAGTGGTCTGCAAC	3547
BSN	-D-L-Q-Q-F-F-M-T-Q-S-G-K-S-S-M-L-S-E-K-Q-T-Q-K-S-F-E-S-S-T-N-M-I-S-S-S-L-V-T-G-Q-V-S-B-M-T-K-N-B-C-S	1121
BSN	-I-T-S-M-F-E-I-H-N-B-A-S-I-R-T-F-Q-K-S-K-F-R-C-H-S-M-F-V-H-G-S-Y-L-N-S-S-S-Q-D-R-F-R-A-S-Q-Q-S-Q-F-A-A	1077
BSN	-S-P-A-E-F-Q-Y-T-M-R-Q-V-L-E-P-F-R-K-A-N-P-D-V-I-L-I-Q-Y-M-D-D-I-L-I-A-S-D-R-T-G-L-E-M-D-K-V-V-L-O	1122
QH1	-C-----C-----T-----A-----C-----A-----G-----A-----G-----A-----	2991
DI94	-C-----C-----T-----A-----C-----A-----G-----A-----G-----A-----	2992
UC2	-C-----C-----T-----A-----C-----A-----G-----A-----G-----A-----	3547
ROD	-C-----C-----T-----A-----C-----A-----G-----A-----G-----A-----	2993
NH2	-G-----G-----A-----A-----GAG-----A-----T-----C-----C-----AC-----A-----	2985
ISY	-G-----G-----A-----A-----GAG-----A-----T-----C-----C-----AC-----A-----	2988
ST	-G-----G-----A-----A-----GAG-----A-----T-----C-----C-----AC-----A-----	2992
CAN2	-G-----G-----A-----A-----GAG-----A-----T-----C-----C-----AC-----A-----	3555
NDS	-G-----G-----A-----A-----GAG-----A-----T-----C-----C-----AC-----A-----	2991
KR	-G-----C-----C-----A-----G-----A-----G-----C-----AGC-----A-----A-----T-----C-----C-----A-----GT-----A-----C-----G-----GG-----A-----GTCC-----A-----	3515
UC1	-G-----C-----C-----A-----G-----A-----G-----C-----AGC-----A-----A-----T-----C-----C-----A-----GT-----A-----C-----G-----GG-----A-----GTCC-----A-----	3528
D205	-G-----C-----C-----A-----G-----A-----G-----C-----AGC-----A-----A-----T-----C-----C-----A-----GT-----A-----C-----G-----GG-----A-----GTCC-----A-----	3520
EMO	-C-----C-----C-----G-----GCAA-G-AC-----C-----AAT-----C-----A-----A-----TC-----C-----G-----A-----GC-----A-----C-----G-----GG-----A-----GTCT-----A-----	3518
NH251	-G-----C-----C-----C-----T-----A-----T-----GC-----C-----G-----G-----T-----G-----CCT-----AG-----T-----C-----A-----ACC-----G-----A-----GG-----A-----TT-----A-----	3440
NH32H	-G-----C-----C-----C-----T-----A-----T-----GC-----C-----G-----G-----T-----G-----CCT-----AG-----T-----C-----A-----ACC-----G-----A-----GG-----A-----TT-----A-----	3465
NH1A11AA	-G-----C-----C-----C-----T-----A-----T-----GC-----C-----G-----G-----T-----G-----CCT-----AG-----T-----C-----A-----ACC-----G-----A-----GG-----A-----TT-----A-----	3463
NH142	-G-----C-----C-----C-----T-----A-----T-----GC-----C-----G-----G-----T-----G-----CCT-----AG-----T-----C-----A-----ACC-----G-----A-----GG-----A-----TT-----A-----	2950
NH239	-G-----C-----C-----C-----T-----A-----T-----GC-----C-----G-----G-----T-----G-----CCT-----AG-----T-----C-----A-----ACC-----G-----A-----GG-----A-----TT-----A-----	3464
NNE	-G-----C-----C-----C-----T-----A-----T-----GC-----C-----G-----G-----T-----G-----CCT-----AG-----T-----C-----A-----ACC-----G-----A-----GG-----A-----TT-----A-----	2932
SHNFBJ14.441	-T-----T-----GC-----T-----AA-----T-----C-----T-----T-----G-----CC-----G-----A-----C-----T-----A-----A-----A-----GG-----A-----TT-----A-----	3230
SHNFBJ14.15	-T-----T-----GC-----T-----AA-----T-----C-----T-----T-----G-----CC-----G-----A-----C-----T-----A-----A-----A-----GG-----A-----TT-----A-----	3231
SHNFBJ6.6	-T-----T-----GC-----T-----AA-----T-----C-----T-----T-----G-----CC-----G-----A-----C-----T-----A-----A-----A-----GG-----A-----TT-----A-----	3423
SHNH4	-Q-----T-----Q-----T-----T-----AA-----T-----A-----G-----G-----C-----G-----C-----G-----A-----A-----A-----GG-----A-----TT-----A-----	3451
SHNH9	-T-----T-----RC-----T-----RR-----T-----A-----T-----A-----T-----G-----CC-----G-----A-----C-----T-----A-----A-----A-----GG-----A-----TT-----A-----	2937
STN	-T-----T-----T-----T-----AA-----T-----A-----G-----A-----G-----C-----C-----G-----A-----C-----T-----A-----A-----A-----GG-----A-----TT-----A-----	3111
Pol RT	-S-P-A-I-F-Q-V-T-M-R-Q-V-L-E-P-F-R-K-A-N-P-D-V-I-L-I-Q-Y-M-D-D-I-L-I-A-S-D-R-T-G-L-E-M-D-K-V-V-L-O	Pol

HIV2/SIV Complete Genomes

BEN	TAAAGAACTTC	3697
BEN	5 K N F S M A S G P L L Q M R S S K K T L R F N Q M A V N V G Q L R G S C R N Y R C P R K T V G Q S	1168
BEN	K R T S K M P R V F V S R S R V P K R P S I S M D G L S T M A N S M K A A E T T T A P F B R H N P S Q	1124
BEN	L K E L L N G L G F S T P D E K F Q K D P P F Q M N G C E L M P T K N K L Q L Q L P Q X D I N T V	1272
GH1	-----G-----C-----C-----G-----G-----A-----T-----A-----T-----A-----A-----C-----	3141
D194	-----G-----C-----C-----G-----G-----A-----T-----A-----T-----A-----T-----A-----A-----C-----	3142
UC2	-----C-----G-----C-----C-----G-----G-----A-----G-----G-----A-----A-----A-----C-----	3597
ROD	-----C-----G-----A-----C-----C-----A-----G-----A-----G-----A-----G-----A-----A-----G-----A-----A-----	3143
NIH2	-----A-----A-----C-----A-----A-----G-----A-----G-----A-----T-----A-----A-----A-----A-----A-----A-----	3135
ISV	-----G-----A-----C-----C-----A-----G-----A-----G-----A-----C-----A-----A-----A-----A-----A-----A-----	3138
ST	-----G-----A-----C-----C-----A-----G-----A-----G-----A-----T-----A-----A-----A-----A-----A-----A-----	3142
CAM2	-----G-----A-----C-----C-----A-----G-----A-----G-----A-----T-----A-----A-----A-----A-----A-----A-----	3705
NDS	-----G-----C-----A-----G-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3141
KR	-----T-----A-----C-----C-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3565
UC1	-----G-----A-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3678
D205	-----G-----A-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3670
END	-----G-----A-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3668
NIH2S1	-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3590
NIH2H	-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3615
NIH1A11AA	-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3613
NIH142	-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3100
NIH239	-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3614
NIH2	-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3082
SHMPBJ14.441	-----G-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3380
SHMPBJ14.15	-----G-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3381
SHMPBJ6.6	-----G-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3573
SHKH4	-----G-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3601
SHKH9	-----G-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3087
STM	-----G-----G-----C-----A-----A-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	3261
Pol RT	L K E L L N G L G F S T P D E K F Q K D P P F Q M N G C E L M P T K N K L Q L Q L P Q X D I N T V	Pol
BEN	ATGACATCCAAAGCTGCGGGAGCTTAAATGGGCGCCCAATCCTCTAGGSAATAAAACCAACACTTATSTAGCAATATAGAGGAAATGACACTCAGAGAAAGTGCAGTGGACACCACTAGCAGAAACAGAGCTAGAG	3847
BEN	M T S R S S W E S I Q R H K S I Q R S K P N T V V D S L B E X S H S O K K C S G O M S Q K Q S S K	1211
BEN	S H P K A S G S L K L G Q T M L F R M K M O T L M S T N S R K M D T H R R S A V D R T S R S R A R R	1171
BEN	N D I O K L V G V L N M A A O I V S G I K T N L L C R L I R G K M T L T E S V Q W T E L A B A E L E	1222
GH1	-----C-----G-----G-----A-----T-----G-----C-----G-----G-----G-----T-----G-----A-----G-----A-----G-----A-----	3291
D194	-----A-----A-----A-----T-----G-----C-----G-----G-----G-----G-----T-----G-----A-----G-----A-----G-----A-----	3292
UC2	-----A-----A-----A-----T-----G-----C-----G-----G-----G-----G-----T-----G-----A-----G-----A-----G-----A-----	3847
ROD	-----G-----T-----A-----A-----C-----C-----G-----G-----G-----G-----G-----T-----C-----A-----A-----A-----A-----	3293
NIH2	-----T-----C-----A-----A-----C-----C-----G-----G-----G-----G-----G-----T-----C-----A-----A-----A-----A-----	3285
ISV	-----G-----A-----T-----C-----C-----C-----C-----G-----G-----G-----G-----G-----T-----C-----A-----A-----A-----A-----	3288
ST	-----A-----G-----T-----C-----A-----C-----G-----G-----G-----G-----G-----T-----C-----A-----A-----A-----A-----	3292
CAM2	-----A-----A-----T-----C-----A-----C-----G-----G-----G-----G-----G-----T-----C-----A-----A-----A-----A-----	3856
NDS	-----A-----A-----T-----C-----A-----C-----G-----G-----G-----G-----G-----T-----C-----A-----A-----A-----A-----	3291
KR	-----A-----A-----T-----C-----A-----C-----G-----G-----G-----G-----G-----T-----C-----A-----A-----A-----A-----	3815
UC1	-----T-----A-----A-----A-----C-----A-----T-----C-----T-----G-----A-----G-----A-----A-----A-----A-----A-----	3826
D205	-----CA-----T-----A-----G-----A-----A-----C-----A-----T-----C-----T-----G-----A-----G-----A-----A-----A-----A-----	3820
END	-----T-----G-----T-----A-----A-----A-----A-----T-----C-----T-----C-----G-----G-----A-----A-----A-----A-----A-----	3818
NIH2S1	-----T-----A-----G-----T-----A-----A-----A-----A-----T-----C-----T-----C-----G-----G-----A-----A-----A-----A-----	3740
NIH2H	-----T-----A-----G-----T-----A-----A-----A-----A-----T-----C-----T-----C-----G-----G-----A-----A-----A-----A-----	3765
NIH1A11AA	-----T-----A-----G-----T-----A-----A-----A-----A-----T-----C-----T-----C-----G-----G-----A-----A-----A-----A-----	3763
NIH142	-----T-----A-----G-----T-----A-----A-----A-----A-----T-----C-----T-----C-----G-----G-----A-----A-----A-----A-----	3280
NIH239	-----T-----A-----G-----T-----A-----A-----A-----A-----T-----C-----T-----C-----G-----G-----A-----A-----A-----A-----	3764
NIH2	-----T-----A-----G-----T-----A-----A-----A-----A-----T-----C-----T-----C-----G-----G-----A-----A-----A-----A-----	3232
SHMPBJ14.441	-----T-----A-----A-----A-----G-----C-----A-----T-----C-----T-----C-----T-----C-----T-----A-----A-----A-----A-----	3530
SHMPBJ14.15	-----T-----A-----A-----A-----G-----C-----A-----T-----C-----T-----C-----T-----C-----T-----A-----A-----A-----A-----	3531
SHMPBJ6.6	-----T-----A-----A-----A-----G-----C-----A-----T-----C-----T-----C-----T-----C-----T-----A-----A-----A-----A-----	3723
SHKH4	-----T-----A-----A-----A-----G-----C-----A-----T-----C-----T-----C-----T-----C-----T-----A-----A-----A-----A-----	3751
SHKH9	-----T-----A-----A-----A-----G-----C-----A-----T-----C-----T-----C-----T-----C-----T-----A-----A-----A-----A-----	3217
STM	-----T-----A-----A-----A-----G-----C-----A-----T-----C-----T-----C-----T-----C-----T-----A-----A-----A-----A-----	3411
Pol RT	N D I O K L V G V L N M A A O I V S G I K T K H L C R L I R G K M T L T E S V Q W T E L A B A E L E	Pol

[illegible]

HIV2/STV Complete Genomes

GEN	ACTGGCAAGTGCATGGATCCGAGTGGGACTTTGGTATCTACCCACCACCTGGTCAAGGTTAACATTTAAACCTGATAGGAGATCTCTACCCAGGCGCAGAGACCTTCTACACAGATGGATCATSCAAATAGACAGTCAAAGAGGGGAAAG	4294
BEN	T G A K S H G S Q S Q T L Y L V F R H M G Q S R L T S L Y O A A O R F S T O M D H A I D S O K R E K	1347
BZN	L G A S H M D P R V G L C T I Y P T T G O V V N I S P F R S L T V A R D L L H R M I M O S T V K R G K S	1316
BEN	Y W Q V T W I P E W D F V S T P P L V R L T F R L V G D P I P G A E T F V Y T D G S C N R O S R E G K	1370
GH1	-----C-----A-----G-----T-----A-----T-----G-----A-----	3738
D194	-----A-----G-----C-----G-----T-----G-----A-----T-----A-----	3739
UC2	-----G-----A-----C-----G-----G-----G-----G-----A-----C-----G-----A-----	4294
ROD	-----C-----C-----G-----G-----G-----G-----T-----C-----G-----A-----	3740
NIH2	-----A-----C-----C-----C-----G-----G-----G-----G-----T-----A-----T-----C-----G-----A-----G-----	3732
ISY	-----T-----C-----C-----C-----G-----G-----G-----AA-----G-----T-----G-----A-----	3735
ST	-----C-----CA-----G-----A-----GT-----A-----GT-----G-----GAA-----C-----T-----A-----A-----C-----AG-----A-----G-----A-----	3739
CAM2	-----G-----C-----C-----C-----AC-----G-----T-----A-----G-----C-----T-----A-----A-----C-----G-----A-----A-----	4303
HDS	-----G-----C-----C-----GT-----AA-----C-----T-----A-----C-----C-----G-----A-----A-----	3738
KR	-----T-----C-----C-----T-----G-----AA-----C-----T-----A-----C-----C-----T-----G-----A-----	4262
UC1	-----A-----T-----A-----T-----C-----A-----T-----A-----A-----AC-----G-----C-----AC-----CAA-----C-----CC-----GA-----ANG-----TC-----A-----A-----C-----C-----GCC-----G-----G-----	4275
D205	-----A-----C-----C-----G-----T-----AA-----A-----AC-----G-----C-----AC-----CAA-----C-----CC-----GA-----ANG-----TC-----A-----A-----G-----C-----AOC-----G-----A-----	4267
EHO	-----A-----C-----A-----A-----T-----C-----A-----T-----AA-----A-----AC-----G-----C-----AC-----CAA-----C-----CC-----GA-----A-----T-----A-----T-----A-----C-----T-----C-----A-----GCC-----A-----G-----	4265
NI251	-----T-----G-----A-----C-----A-----G-----T-----A-----C-----A-----G-----T-----A-----A-----A-----GTC-----C-----T-----GAAG-----C-----GAG-----A-----A-----A-----AT-----T-----T-----A-----A-----G-----	4187
NI32H	-----T-----G-----A-----C-----A-----A-----T-----A-----C-----A-----A-----A-----A-----A-----GTC-----C-----T-----GAAG-----C-----AAG-----A-----A-----A-----AT-----T-----T-----A-----A-----G-----	4212
NI1A11AA	-----T-----G-----A-----C-----A-----A-----A-----T-----A-----C-----A-----A-----A-----G-----A-----GTC-----C-----T-----GAAG-----C-----GAG-----A-----A-----A-----AT-----T-----T-----A-----A-----G-----	4210
NI142	-----T-----G-----A-----C-----A-----G-----A-----T-----CA-----C-----A-----A-----T-----A-----A-----A-----GTC-----C-----T-----GAAG-----C-----GAG-----A-----A-----A-----AT-----TGT-----T-----G-----A-----A-----	3687
NI239	-----T-----G-----A-----C-----A-----G-----A-----T-----A-----C-----A-----A-----G-----A-----A-----A-----GTC-----C-----T-----GAAG-----C-----GAG-----A-----A-----A-----AT-----T-----T-----A-----A-----G-----	4211
NIIE	-----T-----G-----A-----C-----A-----GA-----A-----T-----A-----C-----A-----A-----A-----A-----A-----GTC-----C-----T-----G-----ANG-----C-----AAG-----A-----A-----A-----AT-----TGT-----T-----A-----A-----G-----	3679
SHNPEJ14.441	-----T-----A-----T-----A-----A-----G-----A-----A-----T-----CT-----A-----AC-----GTT-----C-----AA-----G-----AG-----A-----A-----A-----T-----TGT-----C-----T-----G-----G-----A-----G-----	3977
SHNPEJ14.15	-----T-----A-----C-----A-----A-----A-----G-----A-----A-----T-----CT-----A-----AC-----GTT-----C-----G-----A-----G-----AG-----A-----A-----A-----T-----TGT-----C-----T-----G-----G-----A-----G-----	3978
SHNPEJ6.6	-----T-----A-----C-----A-----A-----A-----G-----A-----A-----T-----CT-----A-----AC-----GTT-----C-----AA-----G-----AG-----A-----A-----A-----T-----TGT-----C-----T-----G-----G-----A-----G-----	4170
SHH4	-----T-----A-----C-----A-----T-----G-----A-----A-----T-----CT-----A-----A-----GTC-----C-----AA-----G-----AG-----A-----A-----A-----T-----TGT-----C-----T-----G-----G-----A-----	4184
SHH9	-----T-----G-----A-----C-----A-----A-----A-----G-----A-----A-----T-----CT-----A-----AC-----GTT-----C-----AA-----G-----AG-----A-----A-----A-----T-----TGT-----A-----C-----T-----G-----G-----A-----	3684
STN	-----T-----G-----A-----C-----A-----A-----T-----T-----A-----A-----GTC-----TT-----AA-----A-----CT-----GGA-----AA-----A-----T-----T-----TGT-----C-----T-----C-----G-----A-----A-----G-----	3858
Pol ST	Y W Q V T W I P E W D F V S T P P L V R L T F R L V G D P I P G A E T F V Y T D G S C N R O S R E G K	Pol

[illegible]

HIV2/SIV Complete Genomes

[illegible]

BEN	AAACAGAAAGGAGAGGCAATACATGGGCAAGTAAATGCAGAAATAGGCGTTTGGCAATGGAACTACACACACTTGAAGGAAAATCATTATAGTAGCAGTACATGTTGCAAGTGGATTTCATAGAGGAGGAGTCATCCACAGGAATCAG	4891
BEN	M R K E K P Y M G K S M O K S A F G K W T Y M T S X E K S L S S O Y M L O V D S S K O R S S M R N Q	1518
BEN	T E R R S H T M A S C R R R L A N G L R T L R R X N M Y S S S T C C K M Y H R S R S M P T G T R	1507
BEN	Q O K G E A I H G O V N A E I G V W O M D Y T H L B G R F T I V A V H V A S G F I R A E V I P O R S	1560
GM1	-----G-----G-----T-C-----G-----G-----G-----T-----	4335
D194	-----G-----G-----G-----G-----G-----G-----G-----G-----	4336
UC2	-----G-----G-----G-----G-----G-----G-----G-----G-----	4891
ROD	-----G-----T-----C-----AC-----G-----T-----G-----T-----T-----G-----	4337
NIH2	-----G-----G-----T-----C-----AC-----G-----T-----G-----T-----G-----T-----	4329
ISY	-----A-----G-----T-----C-----ACC-----G-----G-----T-----T-----G-----T-----	4332
ST	-----G-----G-----G-----T-----T-----AC-----G-----G-----T-----T-----	4336
CAM2	-----G-----G-----T-----G-----AC-----G-----T-----G-----G-----C-----T-----	4900
NDS	-----G-----G-----T-----CT-----AC-----G-----T-----G-----G-----T-----G-----A-----	4335
KR	-----G-----T-----G-----G-----C-----G-----AC-----G-----G-----G-----G-----C-----G-----T-----G-----T-----A-----G-----	4859
UC1	-----A-----G-----T-----A-----G-----G-----G-----ACCA-----G-----GT-----G-----GG-----A-----G-----C-----A-----C-----A-----A-----A-----	4872
D205	-----A-----G-----T-----T-----A-----G-----CC-----GACA-----G-----GT-----T-----T-----A-----G-----C-----A-----C-----G-----T-----G-----A-----A-----C-----A-----GA-----	4864
END	-----A-----G-----T-----T-----A-----G-----T-----C-----GACA-----G-----GT-----T-----G-----GG-----T-----A-----G-----T-----A-----C-----A-----A-----A-----A-----	4862
NIH2S1	-----T-----G-----T-----G-----T-----TC-----GAC-----G-----GT-----C-----TC-----G-----AG-----C-----T-----A-----T-----A-----A-----A-----GA-----	4784
NIH2H	-----T-----A-----T-----T-----G-----T-----TC-----GAC-----T-----GT-----C-----TC-----G-----A-----C-----T-----A-----T-----A-----A-----A-----GA-----	4609
NIH2A11AA	-----T-----A-----T-----T-----G-----T-----TC-----GAC-----T-----GT-----C-----TC-----G-----A-----C-----T-----A-----T-----A-----A-----A-----GA-----	4607
NIH22	-----T-----A-----T-----T-----G-----T-----CC-----GAC-----T-----GT-----C-----TC-----G-----AG-----C-----T-----A-----T-----A-----A-----A-----	4294
NIH239	-----T-----G-----T-----G-----C-----T-----TC-----GAC-----T-----GT-----C-----TC-----G-----A-----C-----T-----A-----T-----A-----A-----GA-----	4808
NIH2	-----T-----G-----T-----G-----C-----T-----TC-----GAC-----T-----GT-----C-----TC-----G-----A-----C-----T-----A-----T-----A-----A-----GA-----	4276
SHMPBJ14.441	-----C-----A-----C-----GAC-----G-----GT-----C-----G-----C-----A-----A-----G-----T-----G-----A-----T-----G-----A-----	4574
SHMPBJ14.15	-----C-----A-----C-----GAC-----G-----GT-----C-----G-----C-----A-----A-----T-----G-----T-----G-----A-----A-----G-----A-----	4575
SHMPBJ6.6	-----C-----A-----C-----GAC-----G-----GT-----C-----G-----C-----A-----A-----T-----G-----T-----G-----A-----A-----G-----A-----	4767
SHNH4	-----C-----A-----C-----GAC-----G-----GT-----G-----C-----C-----A-----A-----T-----G-----T-----G-----A-----G-----A-----	4796
SHNH9	-----C-----T-----Y-----R-----R-----C-----GAC-----G-----GT-----C-----G-----C-----A-----A-----T-----G-----GT-----G-----A-----G-----R-----A-----	4281
STN	-----C-----T-----G-----GT-----G-----GT-----TC-----C-----G-----A-----A-----T-----G-----T-----G-----A-----A-----A-----A-----	4455
Integrase	Q O K G E A I H G O V N A E I G V W O M D Y T H L B G R F T I V A V H V A S G F I R A E V I P O R S	Pol
BEN	GAAGGCAGACAGCACTCTTCTTATTAATACTGGCCAGTAGGTGGCCAAATAAGCACTTGCACACAGCAATAGGCCCACTTCACTTCACAGCAAGTGAAGATGCTGGCATGCTGGGTAGGTATAGAACAACTCTTGGAGTACCTTACA	5041
BEN	E G R O H S S V S M N P V G G Q S M A T A C T O T M A P T S L H R R S R W N K G G S V S M N P L B Y L T	1560
BEN	X A D S T L F I K T Q O S V A N M A T A H R O M P O L I F T G S E D G G M V G R Y R F I L M S T L Q	1554
BEN	R O T A L F L L R L A S R W F I T E L H T D M G P N F T S O E V K M V A M N V G T E O S F G V P Y	1618
GM1	-----G-----A-----A-----T-----T-----T-----G-----G-----A-----A-----G-----G-----S-----F-----G-----V-----P-----Y-----	4485
D194	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4486
UC2	-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	5041
ROD	-----A-----A-----G-----A-----A-----T-----T-----TG-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4487
NIH2	-----A-----G-----T-----C-----G-----A-----C-----T-----TG-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4479
ISY	-----A-----A-----C-----A-----T-----T-----TG-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4482
ST	-----A-----G-----C-----A-----T-----A-----TG-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4486
CAM2	-----A-----A-----T-----T-----C-----A-----A-----A-----TG-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	5050
NDS	-----A-----G-----T-----T-----A-----A-----T-----TG-----G-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	4485
KR	-----A-----A-----C-----A-----C-----T-----TG-----G-----C-----C-----A-----A-----A-----A-----A-----A-----A-----A-----A-----	5009
UC1	-----A-----A-----T-----G-----T-----C-----A-----T-----C-----A-----C-----G-----C-----A-----C-----A-----C-----A-----G-----A-----C-----G-----C-----T-----	5022
D205	-----A-----A-----T-----C-----GT-----C-----A-----T-----C-----A-----C-----A-----C-----TG-----C-----CAAGT-----A-----C-----A-----A-----A-----T-----C-----T-----	5014
END	-----A-----A-----T-----G-----G-----G-----C-----A-----T-----C-----A-----C-----TG-----G-----A-----T-----A-----CA-----C-----A-----G-----A-----A-----C-----G-----C-----T-----	5017
NIH2S1	-----A-----A-----A-----T-----G-----T-----AG-----C-----A-----TG-----T-----A-----TC-----A-----T-----TG-----T-----TG-----C-----G-----A-----A-----T-----C-----G-----G-----CA-----G-----A-----	4934
NIH2H	-----A-----A-----A-----T-----G-----T-----AG-----C-----A-----TG-----T-----A-----TC-----A-----T-----TG-----T-----TG-----C-----G-----A-----A-----T-----C-----G-----G-----CA-----G-----A-----	4959
NIH2A11AA	-----A-----A-----A-----T-----G-----T-----AG-----C-----A-----TG-----T-----A-----TC-----A-----T-----TG-----T-----TG-----G-----A-----A-----T-----C-----G-----G-----CA-----G-----A-----	4967
NIH22	-----A-----A-----A-----T-----G-----T-----A-----C-----A-----T-----T-----A-----TC-----A-----T-----TG-----T-----TG-----G-----A-----A-----T-----C-----G-----G-----CA-----G-----A-----	4444
NIH239	-----A-----A-----A-----T-----G-----T-----AG-----C-----A-----T-----T-----A-----TC-----A-----T-----TG-----T-----TG-----G-----A-----A-----T-----C-----G-----G-----CA-----G-----A-----	4958
NIH2	-----A-----A-----A-----T-----G-----T-----A-----C-----A-----T-----T-----A-----TC-----A-----T-----TG-----T-----TG-----G-----A-----A-----T-----C-----G-----G-----CA-----G-----A-----	4426
SHMPBJ14.441	-----A-----A-----G-----T-----GC-----T-----A-----C-----A-----C-----C-----A-----TC-----T-----T-----TG-----T-----TG-----A-----A-----A-----T-----C-----G-----T-----GA-----G-----T-----	4724
SHMPBJ14.15	-----A-----A-----G-----T-----GC-----T-----A-----C-----A-----C-----C-----A-----TC-----T-----T-----TG-----T-----TG-----A-----A-----A-----T-----C-----G-----T-----GA-----G-----T-----	4725
SHMPBJ6.6	-----A-----A-----G-----T-----GC-----T-----A-----C-----A-----C-----C-----A-----TC-----T-----T-----TG-----T-----TG-----A-----A-----A-----T-----C-----G-----T-----GA-----G-----T-----	4917
SHNH4	-----A-----A-----G-----T-----GC-----T-----A-----C-----A-----C-----C-----A-----TC-----T-----T-----TG-----T-----TG-----A-----A-----A-----T-----C-----G-----T-----GA-----G-----T-----	4946
SHNH9	-----A-----A-----G-----T-----GC-----T-----A-----C-----A-----C-----C-----A-----TC-----T-----T-----TG-----T-----TG-----A-----A-----A-----T-----C-----R-----T-----GA-----R-----G-----T-----	4431
STN	-----G-----A-----T-----G-----G-----GT-----A-----C-----A-----TG-----C-----A-----C-----G-----T-----G-----T-----C-----G-----G-----A-----C-----G-----A-----A-----A-----	4605
Integrase	R O T A L F L L R L A S R N P I T H L B G R F T I V A V H V A S G F I R O S P G V P Y	Pol

HIV2/SIV Complete Genomes

BEN	ACCCACAAAACCCAGGGAGTAGTAGAAGCAATGAAATCAACACCTAAGGAATCAGATAAGTAGAATTAGAGAAACAGGCAAAATACAATAGAAACATAGTACTGATGGCAGTTTCATTCATGATGAATTTTAAAGAAAGGGGAGGAAATAGGGGATA	5191
BEN	T H R A R E S I R O S I T T G R I R S V E L E N R Q I O S R O S Y S W O F I A S I L K E G E S G I	1600
BEN	P T K P G S S R S N E S F P K E S D F S N S R T G K I Y N R N S T D O S S L H E P S K K G R M A G Y	1603
BEN	N P Q S Q G V V E A M N H R L K N Q I S R I R E Q A N T I B T I V L M A V H C M N F K R R G G I O D	1668
GH1	-----T-----C-----A-----	4635
D194	-----T-----A-----	4636
UC2	-----T-----A-----C-----A-----T-----G-----G-----A-----	5191
ROD	-----T-----G-----A-----T-----A-----C-----A-----C-----A-----A-----G-----	4637
NIMZ	-----T-----A-----T-----T-----G-----A-----GAC-----G-----G-----T-----A-----	4629
ISY	-----T-----A-----G-----A-----GAA-----G-----G-----A-----	4632
ST	-----T-----A-----G-----A-----C-----G-----C-----G-----A-----G-----	4636
CAM2	-----T-----A-----G-----A-----C-----G-----C-----G-----A-----	5200
MD5	-----T-----A-----G-----C-----C-----C-----T-----G-----A-----	4635
KR	-----T-----A-----G-----T-----A-----C-----C-----G-----G-----T-----A-----	5159
UC1	-----T-----A-----C-----T-----A-----C-----A-----GA-----C-----G-----C-----GTAT-----G-----G-----T-----A-----C-----C-----	5172
D205	-----T-----A-----G-----C-----T-----G-----A-----A-----GAC-----C-----C-----A-----GTAT-----G-----G-----T-----A-----AC-----C-----	5164
EMO	-----T-----G-----T-----G-----C-----T-----G-----A-----GAC-----C-----T-----GTAT-----G-----G-----T-----A-----AC-----C-----	5162
MN251	-----T-----G-----T-----G-----G-----A-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----AC-----C-----	5084
MN32H	-----T-----G-----T-----G-----G-----A-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----AC-----C-----	5109
MN1A11AA	-----T-----G-----T-----G-----G-----A-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----CG-----C-----	5107
MN143	-----T-----G-----T-----G-----G-----A-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----	4594
MN239	-----T-----G-----T-----G-----G-----A-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----	5108
NNE	-----T-----G-----T-----G-----G-----A-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----	4576
SHNFBJ14.441	-----T-----G-----A-----G-----G-----G-----C-----T-----T-----A-----CC-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----	4874
SHNFBJ14.15	-----T-----G-----A-----G-----G-----G-----C-----T-----T-----A-----CC-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----	4875
SHNFBJ6.6	-----T-----G-----A-----G-----G-----G-----C-----T-----T-----A-----CC-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----	5067
SHN4	-----T-----G-----A-----G-----G-----C-----T-----T-----A-----CC-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----	5096
SHN9	-----T-----G-----A-----G-----G-----G-----C-----T-----T-----A-----CC-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----	4581
STH	-----T-----G-----A-----G-----G-----C-----T-----T-----A-----CC-----A-----GA-----C-----G-----A-----T-----G-----C-----T-----A-----	4755
Integrase	N P Q S Q G V V E A M N R H L K N Q I S R I R E Q A N T I B T I V L M A V H C M N F K R R G G I O D Pol	
BEN	TGACCCAGCAGAAAGACTAATCAACATGATTACACAGAAAGAAATACAAATTCCTCCAAAGAAAAATTTAAAAATTTCCAGGTCATTACAGAGAAAGGAGATCAGCTGTGGAAAGGACCTGGTGAATCTGTGGA	5341
BEN	S P Q Q K D S S T S L P Q M K K Y N S S R E K K I O I L K I S R I T R K A E I S C G K D L V N Y C G	1647
BEN	D P S R K T M Q H D V H R T R N T I P P K K K P K F S K F P G L L Q R R Q R S A V E R T M S T T V E	1691
BEN	M T P A E R L I N M I T T E Q E I Q F L Q R K M S M P A N F Q V Y Y R E G R D O L W K G P Q E L L M	1718
GH1	-----T-----C-----A-----	4785
D194	-----T-----C-----A-----	4786
UC2	-----T-----C-----A-----	5341
ROD	-----T-----T-----T-----T-----C-----G-----GCC-----A-----A-----G-----T-----G-----T-----A-----G-----	4787
NIMZ	-----T-----A-----T-----C-----GCC-----A-----A-----T-----G-----T-----A-----G-----G-----	4779
ISY	-----T-----G-----T-----C-----TG-----G-----GC-----A-----A-----T-----G-----T-----A-----A-----C-----G-----G-----	4782
ST	-----T-----T-----T-----G-----C-----TG-----G-----GC-----A-----AC-----T-----G-----T-----A-----G-----	4786
CAM2	-----T-----T-----T-----C-----G-----GCC-----A-----A-----T-----G-----T-----A-----G-----	5350
MD5	-----T-----T-----T-----C-----GCC-----A-----A-----T-----G-----T-----A-----G-----	4785
KR	-----T-----T-----T-----C-----GCC-----A-----A-----T-----G-----T-----A-----G-----	5309
UC1	-----T-----G-----G-----T-----A-----G-----T-----GC-----T-----A-----C-----A-----C-----G-----T-----	5322
D205	-----T-----G-----G-----T-----A-----G-----T-----GC-----T-----A-----C-----A-----C-----G-----T-----	5314
EMO	-----T-----A-----G-----A-----T-----G-----T-----GC-----T-----A-----C-----A-----C-----G-----T-----	5312
MN251	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	5234
MN32H	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	5259
MN1A11AA	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	5257
MN143	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	4744
MN239	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	5258
NNE	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	4726
SHNFBJ14.441	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	5024
SHNFBJ14.15	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	5025
SHNFBJ6.6	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	5217
SHN4	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	5244
SHN9	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	4731
STH	-----T-----T-----T-----C-----T-----T-----AA-----TC-----C-----A-----T-----G-----A-----C-----G-----T-----	4905
Integrase	N T P A E R L I N M I T T E Q E I Q F L Q R K M S M P A N F Q V Y Y R E G R D O L W K G P Q E L L M Pol	

[illegible]

HIV2/SIV Complete Genomes

	-- Pol CDS end (frame 3)																																																		
BEN	TGGACCTAACACGAGAAAAAGGATGGCTCTCTCTCTCAAGCAGTAAGATTAACTGGTATACAGAAAAATCTGACAGATGTTACCGCAGACTGCGCAGATCTTAATACATACGACTTATTCTCTTGGCTTACGGCAGGTGAAGTA																																																		5780
BEN	MNLTLTPEKGVLSGNAVRLTLNVTETKFWNTDVTDPDCAGADILIHSTYFSCFTAGGEV																																																		1792
BEN	GTPSHQKKKDGSPKSDSPGITOKESSGQNLPTOTVOTSSSYVILSLALROVK																																																		1788
BEN	LEPNTTRKRLNALLSCKSKILNLVYRKVLDRCYPRLCRHPMTSLPLLLLYGRSS																																																		1864
GN1	-----G-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5232
D194	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5233
UC2	-----G-----T-----C-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5780
ROD	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5234
NIRZ	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5226
ISY	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5229
ST	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5233
CAM2	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5797
NDS	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5232
KR	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5756
UC1	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5769
D205	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5761
BHO	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5759
MN251	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5691
MN32M	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5706
MN1A11AA	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5704
MN142	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5191
MN239	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5705
MNE	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5173
SNMPBJ14.442	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5471
SNMPBJ14.15	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5472
SNMPBJ5.6	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5664
SNM14	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5633
SNM19	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5178
STN	-----T-----G-----A-----G-----V-----L-----D-----R-----C-----Y-----P-----R-----L-----C-----R-----H-----P-----M-----T-----S-----R-----L-----P-----L-----L-----Y-----G-----R-----S-----S-----																																																		5352
Integrase	LEPNTTRKRLNALLSCKSKILNLVYRKVLDRCYPRLCRHPMTSLPLLLLYGRSS																																																		
Vif	MNLTLTPEKGVLSGNAVRLTLNVTETKFWNTDVTDPDCAGADILIHSTYFSCFTAGGEV																																																		Vif

[illegible]

I-111
DEC 97

HIV2/SIV Complete Genomes

	--...> Vpx CDS end	--> Vpr CDS start		
BEN	TCTAGTCTA	ATGACTGAAGCACCAACAGAGATTTC	CCCCAGAAAGATGGGACCCCAAGGAGGGACTTAGGGAGTGA	CTGGGTAAATGAAACTCTGGGGGAAATAAGGAAGAGCGCTTAAGACATTTTGATCC
BEN	S S L	M T E A P T E F P P E D G T F R R D L G S D W V I E T L R E I R E E A L R H P D P R L L I A		
BEN	L V	S L K H A Q Q S E F P Q K H G T P H Q G G T G V T G S K L S G M R K K K F S D I L I E A C S L L		
BEN	V S S	S S T M R V S F R R M D P T E Q L R E S L G N R N T S P G M X Q R S L K T P S S F L A N C		
GH1			AC G	AG T G A
D194			A GC	A C
UC2		TC G G T		AG G TGA TGT T A A G
UC2				
ROD	G	G	C C	G TG
NIH2	T		AC C	G G A
ISV	T		G	G G A
ST			C C	G G A
CAM2	GT		C C T	G A
HDS	T		C C	G G
KR	C		G	G A A
UC1	C C		G A	G CC
D205	C G C		G A	G CC
EHO	C		G A	G TC
MD251	A CA A		GAA AG	TC
MD32H	A CA A		GAA AG	TC
MD31A1A	A CA A		GAA AG	TC
HE142	A CA A		GAA AG	TC
MD239	A CA A		GAA AG	TC
MNE	A CA A		GAA AG	TC
SMNPBJ14.141	A CA AA		A AG	TC
SMNPBJ14.15	A CA AA		A AG	TC
SMNPBJ6.6	A CA AA		A AG	TC
SMNH4	A CA AA		A AG	TC
SMNH9	A CA AA		A AG	TC
STM	CT CA GA		AC CAG	TC
Vpx	L V S	Vpx CDS end		

	--> Tat CDS start																																																																																																				
BEN	CTTGGCTACTATATCATATAGACATGAGAGACACCCCTGAGAGGCCCCAGAGAGCTCTMTAAACCCCTACACACGAGCCCTCTTCTGTCCTTCAGAGCGGATGTAAACCCCTCAAGANTTGGCTAAACAAGGAGAAGAACTCCTTGGCCA																																																																																																				6526
BEN	L G Y V I H N R K G D T L E G A R E L I K T L Q R A L P V M F R A G C N R S R I G S T T R R R T P C P																																																																																																				2033
BEN	L A T I S I D M E T F L K A P E S L K P Y N E P E S C T S E R A D V T A Q E L A A K O G E E L L A Q																																																																																																				2048
BEN	S M L L Y P S S M R N P S R F Q R A H S L N P T T S G P L R A L Q S G M S P A K M W L K M K E K N S L P																																																																																																				2091
GH1	-A- -A- -G- -A- -C- -O- -T- -T- -S- -P- -L- -R- -A- -L- -Q- -S- -G- -M- -S- -P- -A- -K- -M- -W- -L- -K- -M- -K- -E- -K- -N- -S- -L- -P- -T-																																																																																																				5967
D194	-A- -A- -T- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T-																																																																																																				5968
UC2	-A- -A- -T- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T-																																																																																																				6465
ROD	-A- -A- -T- -C- -G- -T- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T-																																																																																																				5969
MIH2	-AGA- -T- -C- -G- -G- -T- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T-																																																																																																				5961
ISY	-A- -A- -T- -C- -G- -G- -T- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T-																																																																																																				5964
ST	-A- -A- -T- -G- -G- -T- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T-																																																																																																				5968
CAN2	-GGAC- -T- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T-																																																																																																				6532
HDS	-AGAC- -T- -C- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T-																																																																																																				5967
KR	-A- -A- -T- -G- -G- -T- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T- -A- -A- -G- -G- -T-																																																																																																				6468
UC1	-TA- -T- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6507
D205	-TA- -T- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6499
ERD	-A- -T- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6497
MC251	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6402
NH32H	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6427
NH41A12A	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6425
NH4142	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				5912
NH239	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6426
NHE	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				5894
SNPNB714.441	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6193
SNPNB714.15	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6194
SNPNB76.6	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6286
SNM84	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6391
SNM19	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				5900
STN	-TA- -TC- -T- -T- -C- -G- -G- -T- -C- -A- -AG- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A- -A- -G- -T- -C- -G- -G- -A-																																																																																																				6074
Vpr	L G Y V I H N R K G D T L E G A R E L I K T L Q R A L P V M F R A G C N R S R I G S T T R R R T P C P																																																																																																				Vpr
	Tat CDS Start M E T P L L E G A P E S L K P Y N E P E S C T S E R A D V T A Q E L A A K O G E E L L A Q																																																																																																				Tat42

HIV2/SIV Complete Genomes

	Vpr CDS and <--	--> Rev CDS start	
SEN	GCTGCAACGACCCCTAGAGCGATGCTATTAACAAGTCTATTGTAAGCGATGACAGTTTCOCATTGCCAGCTGTGTTCTCGAAAAAGGGGCTCGGATATCAATATAGAGCGAAAGGGGACAGACGANGAGGACTCCAGGAAAACTAAGACTCC		6676
SEN	A A P T P R A M H L L S A M Q P L F A R N I S A G O T S R D S R E N S D S		2079
SEN	L H R F L E F C T N K C Y C K R C S F M C Q L C F S K K G L G I S Y E R K Q R R R R T P R N T K T F		2069
SEN	S C T D P F S M A L T N A I V S D A V S I A S C V S R K R G S Y H M S E R A D E R Q L O G K L R L		2140
GM1	-----G-----T-----TC-----AG-----C-----T-----A-----G-----GG-----CA-----AA-----G-----A-----G-----A-----		6117
DI94	-----G-----T-----TC-----AG-----A-----T-----A-----G-----GG-----C-----C-----A-----A-----G-----A-----		6118
UC2	A-----G-----CC-----A-----AAG-----T-----A-----G-----GG-----T-----A-----G-----A-----G-----A-----		6615
ROD	AT-----AA-----A-----TC-----A-----T-----A-----TCTA-----C-----A-----G-----GT-----A-----A-----G-----A-----		6139
MINZ	AT-----G-----A-----TCC-----C-----AA-----T-----C-----A-----G-----T-----TAC-----G-----GG-----C-----A-----G-----A-----		6121
ISY	AT-----AA-----A-----T-----T-----G-----T-----C-----A-----T-----T-----C-----G-----GT-----C-----G-----G-----A-----G-----A-----		6114
ST	AT-----A-----G-----GA-----T-----C-----AAG-----T-----C-----A-----A-----T-----TA-----C-----G-----GG-----A-----A-----G-----A-----		6118
CAN2	AT-----A-----AGA-----TGA-----TC-----T-----C-----AT-----A-----A-----TCTA-----G-----GT-----C-----A-----G-----A-----G-----A-----		6682
NDS	AT-----A-----AG-----A-----T-----T-----C-----A-----A-----A-----T-----TA-----C-----G-----GG-----A-----A-----G-----A-----		6117
KR	AT-----A-----G-----A-----CC-----T-----AA-----TT-----A-----A-----T-----TA-----A-----G-----GT-----A-----G-----A-----G-----A-----		6638
UC1	AT-----C-----T-----AG-----GA-----T-----C-----AAA-----T-----A-----T-----CTT-----C-----T-----G-----GG-----C-----ACTC-----A-----A-----A-----T-----T-----AA-----GG-----GT-----A-----		6651
D205	A-----AT-----C-----TA-----G-----GA-----T-----C-----C-----AA-----T-----C-----A-----T-----TCTT-----T-----T-----G-----GT-----C-----CTC-----GA-----G-----AT-----G-----AA-----G-----G-----A-----		6643
ENO	A-----AT-----C-----T-----TA-----G-----G-----T-----C-----C-----AAA-----TTCC-----A-----T-----C-----TCTT-----C-----G-----G-----GT-----A-----CTC-----A-----AGA-----T-----T-----AA-----GG-----G-----A-----		6641
ME251	A-----AT-----C-----T-----G-----G-----TA-----C-----AAG-----TT-----C-----A-----T-----TCTT-----AT-----G-----G-----GT-----A-----AGTCAC-----A-----AGA-----G-----AA-----GG-----G-----AA-----		6552
MD32H	A-----AT-----C-----T-----G-----G-----TA-----C-----AAG-----TT-----C-----A-----T-----TCTT-----CT-----G-----G-----GT-----A-----AGTCAC-----A-----AGA-----A-----G-----AA-----GG-----G-----AA-----		6577
MM1A11RA	AT-----C-----T-----AG-----TA-----C-----AAG-----TT-----C-----A-----T-----TCTT-----A-----CT-----G-----G-----GT-----A-----AGTCAC-----A-----AGA-----A-----G-----AA-----GG-----G-----AA-----		6575
MM142	A-----AT-----C-----T-----AG-----TA-----C-----C-----AAG-----TT-----C-----A-----T-----TCTT-----CT-----G-----G-----AGT-----A-----AGTCAC-----A-----GA-----A-----G-----AA-----GG-----G-----AA-----		6082
MD239	AT-----C-----T-----AG-----TA-----C-----AAG-----TT-----C-----A-----T-----TCTT-----A-----CT-----G-----G-----GT-----A-----TCAC-----A-----AGA-----A-----G-----AA-----GG-----G-----AA-----		6576
MNE	A-----AT-----C-----T-----AG-----TA-----C-----C-----AAG-----TT-----C-----A-----T-----TCTT-----CT-----G-----G-----GT-----A-----AGTCAC-----A-----A-----A-----G-----AA-----GG-----G-----AA-----		6044
SNMFBJ14.441	A-----AT-----C-----T-----G-----G-----TTA-----G-----C-----AA-----G-----TT-----C-----AT-----T-----AT-----TCTT-----CT-----G-----G-----GC-----AGC-----AC-----A-----A-----G-----A-----G-----G-----AA-----		6337
SNMFBJ14.15	A-----AT-----C-----T-----G-----G-----TTA-----G-----C-----AA-----G-----TT-----C-----AT-----T-----AT-----TCTT-----CT-----G-----G-----GC-----AGC-----AC-----G-----A-----G-----A-----G-----G-----AA-----		6338
SNMFBJ6.6	A-----AT-----C-----T-----G-----G-----TTA-----G-----C-----AA-----G-----TT-----C-----AT-----T-----AT-----TCTT-----CT-----G-----G-----GC-----AGC-----AC-----A-----A-----G-----A-----G-----G-----AA-----		6530
SNMH4	-----T-----G-----G-----TA-----G-----C-----A-----G-----TT-----C-----AT-----T-----AT-----TCTT-----CT-----G-----G-----GC-----AGC-----A-----C-----A-----A-----G-----A-----G-----G-----AA-----		6523
SNMH9	A-----AT-----C-----AG-----G-----G-----TTA-----G-----Y-----C-----AA-----TT-----C-----AT-----T-----AT-----TCTT-----CT-----G-----G-----GC-----AGC-----AC-----A-----A-----G-----A-----G-----G-----AA-----		6044
STM	A-----AT-----C-----A-----GAG-----TGG-----G-----T-----C-----AAA-----TT-----C-----AT-----C-----TGTT-----C-----A-----T-----T-----A-----C-----TCAA-----A-----A-----GT-----AAG-----AA-----GG-----TA-----		6218
Vpr	A A P T P R A M H L L S Vpr CDS and		
Tat	L H R F L E F C T N K C Y C K R C S F M C Q L C F S K K G L G I S Y E R K Q R R R R T P R N T K T F		Tat
Rev	-----M-----S-----E-----R-----A-----D-----E-----R-----Q-----L-----O-----G-----K-----L-----R-----L-----		Rev

[illegible]

7 V3 loop start

[illegible]

Δ VJ loop end

[illegible]

BBN	ATAGATATAAAATCGGACTCATACCAATTTACCTTTAGTGCAGAAAGTGGGACGACTGACGATTAGAACTGGGAGACTACAAATTAAATAGAAATACACCAATTTGGCTTTCGCACCTACAGATCAAGAGAGGTTACTCTCAACT	8215
95H	I D K H R T H T H T P S A E H V A E L Y R L E L G D Y K L I B I T T P I G P A P T D Q R R Y S S T	2592
BBN	# # # T T C L T G L T L P L V Q K H Q H C L S D G N H R S B K T F H O L A S A P L Q T R E G C T P O F	2529
BBN	# # R S R S R S D S Y V Q H V L S C R S G R T V P I R T G R L Q T H R H H T V H L R T Y R S E K V L L H	2590
GB1	GT-----A-T-AC-----G-A-T-----G-----A-----T-T-----G-----CAAGS-A-A-T-----GG	7638
D194	TCAG-TGGAAAT-----G-----T-----G-----G-----CCG-----A-AG-----A-A-T-----GG	7626
UC2	-C-G-G-A-A-G-----G-----T-----G-----G-----CG-G-A-A-A-T-----GG-C	8169
ROD	TGGC-----CAAT-A-C-A-----G-----A-A-G-GT-----T-T-----CG-----A-AG-A-A-A-----TG	7661
NH26	CGCGGAS-A-A-A-T-----G-C-----A-----G-T-G-T-----G-C-----TCAGTA-AG-A-----TG	7626
ISY	T-G-G-TGGAGAT-AC-GG-A-T-----G-----A-----G-T-G-T-----G-G-----C-----CAS-A-A-A-----TG	7635
BT	GG-GGAGH-AC-G-A-T-----G-----A-----T-----G-T-----G-----G-----T-----CCAGTA-A-A-----TG	7645
CAN2	G-GGG-AATCAGACA-A-T-----T-----G-T-----G-T-----GG-----TCA-A-AG-A-----C-TG	8227
HDS	G-C-GGAT-A-G-A-T-----T-----A-----T-----G-----CCAG-A-A-C-A-----TG	7692
KR	TGTT-GAT-AC-G-A-G-----G-----A-----G-T-----G-----TCAG-A-A-T-----TG	8159
UC1	GGGTATTAT-GGC-GAT-CA-G-T-----T-----A-G-----G-----G-----G-----T-----GAT-A-A-T-----G	8202
D205	TC-G-CAG-G-ACC-----T-----T-TG-G-----T-----TA-G-----G-----T-----G-----AG-ATA-A-A-----GTG	8191
XBO	TGG-T-G-AA-AC-T-----Y-T-----TG-G-----T-----TA-G-----G-----G-----G-----AG-ATA-A-A-----GTG	8177
NH251	TGG-CTG-G-A-AC-A-T-GT-C-A-G-----G-----T-----G-GT-----T-T-----G-----G-C-T-A-G-----G-C-C-----GT-AG-----A-TA-TGG	8139
NH32B	TGG-CTG-G-A-AC-A-T-T-C-A-G-----G-----T-----G-T-----T-T-----G-----G-C-T-----G-C-C-----GT-AG-----A-TA-TGG	8155
NH1A11AA	TGG-CTG-G-A-AC-A-T-T-C-A-G-----G-----T-----T-G-T-----T-T-----G-----G-C-T-----G-C-C-----A-GT-AG-----A-TA-TGG	8153
NH142	TGG-CTG-G-A-AC-A-T-GT-C-A-G-----G-----T-----T-G-T-----T-T-----G-----G-C-T-----G-C-C-----A-GT-AG-----A-TA-TGG	7652
NH239	TGG-TTG-G-A-AC-A-T-T-C-A-G-----G-----T-----T-G-T-----T-T-----G-----G-C-T-----G-C-C-----GT-AG-----A-TA-TGG	8157
NH2	TGG-TTG-G-A-AC-A-T-T-C-A-G-----G-----T-----T-G-T-----T-T-----G-----G-C-T-----G-C-C-----GT-AG-----A-TA-TGG	7621
SHMPBJ74.441	TGG-TC-G-C-A-G-G-----T-C-A-G-----G-----T-----G-GT-----T-T-----G-----G-T-T-----C-----C-C-----AG-GTA-A-----A-A-----	7948
SHMPBJ74.15	TGG-TC-G-C-A-G-G-----T-C-A-G-----G-----T-----T-G-GT-----T-----G-----G-T-T-----C-----C-C-----AG-GTA-A-----A-A-----	7949
SHMPBJ6.6	TGG-TC-G-C-A-G-G-----T-C-A-G-----G-----T-----T-G-GT-----T-----G-----G-T-T-----C-----C-C-----AG-GTA-A-----A-A-----	8141
SHM94	TGG-TC-A-C-A-G-G-----T-C-A-G-----G-----T-----T-G-GT-----T-----G-----G-T-T-----G-C-C-----AG-GTA-A-----A-A-----	8122
SHMH9	TGG-TC-G-C-A-A-G-----T-C-A-A-G-----G-----T-----T-G-GT-----T-----G-----G-T-T-----C-C-----AG-GTA-A-----A-A-----	7648
STN	TGG-CT-CAAT-A-G-G-----T-C-TGCA-----G-----T-----G-T-----T-----T-----G-C-A-GA-GT-AG-A-A-A-T-G	7911
Env op41	I D K H R Y H T H T P S A E H V A E L Y R L E L G D Y K L I B I T T P I G P A P T D Q R R Y S S T	Env

[illegible]

Strain	Sequence	Position
BEN	TTGGAAGTACCAAGACAGACAGAAATGTTGGGACTGACCGTCTGGGGAACGAAAAACCTCCAGGCAAGTACTGCTATCGAGAAATACCTAAAGCATCAGGCGACGCTAAATTCATGCGGATGTCGGTTTAGACAGGCTCGCCAC	8512
BEN	L D V V K R Q Q E M L R L T V W G T K H L Q A R V T A I E K Y L X H Q A O I N S W G C A F R O V C H	2691
BEN	N T S T S D R H K R C C D T P G S E R Y T S A O E S L T S B I R H S D T H G G D V F L D R S A A T	2620
BEN	V G A S Q E F T A M V A T D R L G M E A P F G R S H C T R E V P K A S G A A K F N G M C V S T G L F	2695
GH1	-----	7932
DI94	-----	7923
UC2	-----	8466
ROD	-----	7958
NH38	-----	7923
ISY	-----	7932
ST	-----	7947
CAH2	-----	8524
ND6	-----	7969
KR	-----	8456
UC1	-----	8499
D205	-----	8468
EHO	-----	8474
NH251	-----	8439
NH328	-----	8455
NH1A11AA	-----	8453
NH142	-----	7952
NH239	-----	8457
NH2	-----	7931
SNHFBJ14.441	-----	8248
SNHFBJ14.15	-----	8249
SNHFBJ6.6	-----	8441
SNH4	-----	8422
SNH9	-----	7946
STH	-----	8108
Env gp41	L D V V K R Q Q E M L R L T V W G T K H L Q A R V T A I E K Y L X H Q A O I N S W G C A F R O V C H	Env
BEN	ACTACTGTACCGGGGTAATGACTCTTTATCGGCTGACTGGAAMATATGACATGGCAGGAGTGGGAGAAACAAAGTCCGCTACCTAGAGGCAATATACCTCAAAGTTTAAAGAGAGCCCAATTCACACAGAGAGAGATATGTATGAA	8562
BEN	T T V V P H V R D S L S P D W K H N T W Q E W B K Q V R V L S A N I S Q S L E A Q I O Q E K H N Y E	2741
BEN	L T V V R G D N T L V R L T Q K T T H O R S O R R B K O R B A V L E A N I S Q S L E A Q I O Q E K H N Y E	2666
BEN	H V C T V G R S L F I A S L E E Y D H A G V G E T S F I P R G R V Q S R F P A R S P H S T R R E V V S	2742
GH1	-----	8092
DI94	-----	8073
UC2	-----	8616
ROD	-----	8108
NH38	-----	8073
ISY	-----	8082
ST	-----	8092
CAH2	-----	8674
ND6	-----	8139
KR	-----	8606
UC1	-----	8649
D205	-----	8638
EHO	-----	8624
NH251	-----	8589
NH328	-----	8605
NH1A11AA	-----	8603
NH142	-----	8102
NH239	-----	8607
NH2	-----	8081
SNHFBJ14.441	-----	8398
SNHFBJ14.15	-----	8399
SNHFBJ6.6	-----	8591
SNH4	-----	8572
SNH9	-----	8096
STH	-----	8258
Env gp41	T T V V P H V R D S L S P D W K H N T W Q E W B K Q V R V L S A N I S Q S L E A Q I O Q E K H N Y S	Env

HIV2/SIV Complete Genomes

BEN	TTACAAAAATTAATAGCTGGCATATCTCTGGCAATCTGCTTGACTTAACCTCTGGGTCAGTATATTCATATGGAGTCATAGTACAGGGGAATAATAGCTTTAAGAAATAGCAATCTATCTAGTCAATTGTTAAGTAGATTAGA	8
BEN	L O K L I N S W D I L G N W P D L T S N V K Y I O V G V H I V V G I T A L R I A J Y V V O L L S R P R	7
BEN	V K N L S I A G S T L T A T G L T S P G V S I F N M E C I S M E S L L E S O S M C N C S V D L S	2
BEN	I T K I K S L Q Y S W D I V S L N L L G O V Y S I W S A Y S S G M N S P K M S N I C S A I V K S I S	2
GH1	-----GC-----G-CT-----T-----A-----TT-----A-G-G-----T-----A-A-----C-----	8
D194	C-----C-----G-CT-----T-G-----A-A-----TT-----A-----G-----C-A-A-----C-----	8
UC2	-----C-----G-CT-----T-----A-A-----G-TT-----A-----G-----A-----R-----C-----	8
ROD	C-----C-----T-----T-----T-----A-----A-----G-----A-----A-----A-----A-----GC-----	8
NIH2	C-----C-----G-T-AC-----T-C-----C-----G-----TT-G-----A-G-----G-----T-----A-A-----A-GA-----C-----	8
ISY	C-G-----GC-----G-T-----T-----A-----A-----CATG-----A-----G-----C-C-----T-----A-----A-A-C-----C-----	8
ST	C-----C-----G-T-----T-----T-----A-A-----G-----TT-----A-----A-----T-----T-----A-----A-A-----C-----	8
CAN2	C-----C-A-----G-T-AC-----T-----A-----GC-----TT-----A-----A-----T-----T-----A-----A-C-----GC-----	8
MDS	C-----C-----C-----T-----T-----T-----A-----TT-----T-C-----A-----T-----T-----A-----A-A-C-----C-----	8
KR	C-----C-----G-T-----T-----T-----A-----A-----T-----G-----A-A-----T-----T-----A-----A-----AC-----C-----	8
UC1	-----G-----G-T-----T-----T-----C-----A-GGCA-----C-CAGG-TA-----C-AT-G-----CA-----T-----TG-----T-----A-----CA-TA-----GA-----C-GCA-----C-----G	8
D205	-----G-A-----G-T-----T-----T-----C-----C-----A-A-A-----C-T-TA-----C-AT-----CA-----GT-----G-----TC-----T-----B-A-----CA-T-----GA-----C-GCA-----C-----G	8
EHO	-----G-----CAA-----T-CA-----T-----T-----A-GGCA-----C-CAGG-TA-----T-AT-----A-A-----G-----TA-----T-----A-----CA-TA-----GA-----C-GCA-----C-----G	8
NH251	-----G-G-----G-GT-----T-----T-----C-TG-T-----T-----A-A-----A-AT-G-----T-----A-----G-----CTG-----TG-----A-----A-A-----C-GC-----AG-----A-G	8
NH32H	-----G-G-----G-GT-----T-----T-----C-TG-T-----T-----A-A-----A-----TT-----T-----A-----G-----CTG-----TG-----A-----A-A-----C-GC-----AG-----A-G	8
NH1A11AA	-----G-G-----G-GT-----T-----T-----C-TG-T-----T-----A-A-----A-----TT-----T-----A-----G-----CTG-----TG-----A-----A-A-----C-GC-----AG-----A-G	8
NH143	-----G-G-----G-GT-----T-----T-----C-T-T-----T-----A-A-----A-----A-TT-----A-T-----A-----G-----CTG-----TG-----A-----A-A-----C-GC-----G-A-----G	8
NH239	-----G-G-----G-GT-----T-----T-----C-TG-T-----T-----A-A-----A-----T-----T-----T-----A-----G-----CTG-----TG-----A-----A-A-----C-GC-----AG-----A-G	8
NME	-----G-G-----G-GT-----T-----T-----C-TG-T-----T-----A-A-G-----A-----C-----TT-----T-----A-----G-----CTG-----TG-----A-----A-A-----C-GC-----AG-----A-G	8
SHMPBJ14.411	-----G-----T-----T-----T-----C-T-T-----T-----A-A-A-----A-----T-A-TG-----TC-A-----G-----GA-----TG-----A-----A-----GA-----GC-----G-A-----	8
SHMPBJ14.15	-----G-----T-----T-----T-----C-T-T-----T-----A-A-A-----A-----T-A-TG-----TC-A-----G-----GA-----TG-----A-----A-----GA-----GC-----G-A-----	8
SHMPBJ16.6	-----G-----T-----T-----T-----C-T-T-----T-----A-A-A-----A-----T-A-TG-----TC-A-----G-----GA-----TG-----A-----A-----GA-----GC-----G-A-----	8
SHSH4	-----G-----CT-----T-----T-----C-T-T-----T-----A-A-GA-----A-----T-A-T-----TC-A-----G-----GA-----TG-----A-----A-----GC-----G-A-----	8
SHSH9	-----G-----SK-----H-T-Y-----T-K-----C-T-T-YT-----A-A-A-----A-----T-A-TG-----TC-A-----G-----GA-----TG-----A-----A-----GA-----GC-----G-A-----	8
STM	-----C-----G-AT-----T-----T-----C-T-----T-----A-GA-----C-A-----AT-----T-----A-----T-----G-----ATG-----G-----A-----A-----A-----GC-----GC-----G	8
Env op41	L O K L I N S W D I L G N W P D L T S N V K Y I O V G V H I V V G I T A L R I A J Y V V O L L S R P R	En

[illegible]

HIV2/STV Complete Genomes

Buy CDs and e-

[illegible]

→ 3' LTR start

BEN	TGATGATGATGACCT.....AATAGGAGT...TCCTGTATC...ACCAAGAGTACCRCGGAGAGAAATGACCTATAAATTGGCAATAGATATGTCACATTTTATAAAGAAGAAAGGGGACTGCAGAGGGATGTTTACAGTAGGAG	9
BEN	-S-0-S-F-P-----M-R-E-----S-C-V-----T-K-S-T-E-R-N-D-L-S-I-G-M-R-Y-V-T-F-Y-K-R-K-R-Q-T-A-R-D-V-L-Q-S-E	2
BEN	D-D-D-D-D-----I-G-I-----P-V-----P-R-V-P-R-R-E-N-T-Y-K-L-A-I-D-M-S-H-F-I-R-K-K-G-G-L-O-Q-G-M-F-Y-S-R-R	3
BEN	L-M-N-N-T-----L-S-E-----F-L-L-F-----D-B-V-H-G-E-K-S-F-I-N-W-O-S-I-C-M-I-L-S-K-K-R-G-D-C-R-O-C-F-T-V-G	8
GHI	-AG-----G-G-----T-----G-----G-TA--C-----G-----GA--GG-----G-T-----GA	1
D194	-----G-G-----T-----G-----G-TA--CC-----C-G-----G-----A-----A-----GA	8
UC2	-----R-----A-----G-C-----T-----T-----TA--CC-----C-G-----G-----A-----A-----GA	9
R00	A-----A-----G-A-----T-----A-----TA--CC-----AC--G-----A-----AC--GG-----G-----GAA-	6
NIE2	A-----A-----G-T-----C-----T-----T-----TA--CC-----A-TC-----G-----T-----G-----C-----T-----CA	6
ZSY	A-----A-----G-G-----C-----T-----AC-----T-----T-----TA--C-----A-----G-A-----G-----A-----GA	6
ST	A-----A-----G-G-----C-----C-----T-----T-----TA--A-----A-----GG-----G-----G-----G-----C-----A-----QAT	5
CAN2	A-----A-----G-G-----G-----CC-----T-----TA--AC-----A-----G-----C-----G-----C-----GAA-	9
MDS	A-----A-----G-----TA-----C-----C-----AC-----TA--C-----A-----G-----G-----G-----GAA-	8
KR	A-----A-----A-----G-----C-----C-----T-----T-----TA--G-----A-----G-----G-----G-----G-----GA	9
UC1	A-----AG-----TGAGGATATGA-G-G-G-----CTA--A-G-----C-ATAGG-----T-----TC-----A-C-GA-----C-----T-----T-----AC-----T-----GA	9
D205	A-----AAG-----TGAGGAGATGA-G-G-----C-----CATA--A-G-ACC-----G-----T-----CC-----A-C-C-A-G-----C-----T-----C-----G-----G-----T-----UA	9
EHO	A-----G-----TGAGGACAATGA-G-G-----C-G-A-A-G-----CG-G-C-----T-----GCC-----A-TC-----C-A-G-----C-----T-----T-----G-----A-----G-----T-----C-----T-----GA	9
NH251	G-A-----T-----GG-G-----AT-A-G-G-----G-----A-----T-----C-----TA--C-----T-----C-----G-----G-----T-----A-----OCA	9
NH37E	G-A-----T-----GG-G-----A-A-G-T-----G-----C-----T-----C-----TA--AC-----GT-C-----G-----G-----T-----A-----CCA	9
NH41AAA	G-A-----T-----GG-G-----AT-A-G-G-----G-----A-----T-----C-----TA--AC-----GT-C-----G-----T-----G-----T-----A-----T-----GCA	9
NH142	G-A-----T-----GG-GA-----A-A-GA-----GG-C-----T-----C-----TA--AC-----GT-C-----T-----G-----G-----T-----A-----GCA	8
NH239	GT-A-----T-----GG-G-----AT-A-G-G-----G-----A-----T-----C-----TA--AC-----GT-C-----G-----G-----T-----A-----GCA	9
NHE	A-A-----A-----T-----GG-G-----A-A-G-G-----G-----C-----T-----CPFA--AT-A-GT-C-----G-----T-----T-----A-----GAX	8
SMMFBJ14.4d1	-C-C-C-----T-----G-----TTG-----A-G-G-----C-----C-----CGTA--ATC-----A-C-----C-----T-----G-----T-----A-----QAT	9
SMMFBJ14.15	-C-C-C-----T-----G-----TTG-----A-G-G-----C-----C-----CGTA--ATC-----A-C-----C-----T-----G-----T-----A-----GAT	9
SMMFBJ6.6	-C-C-C-----T-----G-----TTG-----A-G-G-----C-----C-----CGTA--ATC-----A-C-----C-----T-----G-----T-----A-----GAT	9
SMMH4	--A-----T-----T-----CT-A-GCA-----C-----C-----CTTA--G-CC-----A-C-----C-----T-----G-----T-----A-----GA	9
SMMH9	--A-----T-----G-----TTG-----A-G-G-----C-----C-----CGTA--ATC-----A-C-----C-----T-----G-----T-----A-----TT	8
STM	G-A-----A-----T-----G-----AG-A-ACH-----T-----C-----TY-G-----T-----C-----G-----CC-----A-CC-----G-----T-----A-----GR	9
Ref	D-D-D-D-D-----I-G-I-----P-V-----P-R-V-P-R-R-E-N-T-Y-K-L-A-I-D-M-S-H-F-I-R-K-K-G-G-L-O-Q-G-M-F-Y-S-R-R	9

[illegible]

		Nef CDS end <--			
BEN	GTCTGGGCAAGTCAGGATGCCAGAGAAAGTATGGAAGGCAAACTGAAGCAAGAGGATACCATATATGTA...	ATAACAGGAACAAACCTACTTGGTCAAGGCAAGAGTACCTAC...	TAAAGAACAGCTGAG	9947	
BEN	VMAQVRIARSRMGRTESKRQDTPISS...	ITGTTITLGGQGRRLIS...	SRTE	3149	
BSN	FQHFSGLFRRFWRAKLRKATPKYS...	SQBQPYLVKAGSSV...	FKQLR	3069	
BSN	SLGTSQDCQAKNGRQNSKQEGYHIV...	INANNHTWSSRQBVAS...	IANSS	2138	
GH1	-----G-----	-----G-----	CTAG	9363	
D194	A-----G-----	-----G-----	CTAG	9357	
UC2	A-----G-----	-----G-----	-----T-----	9901	
ROD	A-----CC-----G-----G-----G-----A-----T-----T-----AG-----G-----G-----A-----	-----G-----G-----G-----A-----G-----T-----G-----A-----	AC-----	9383	
NIH2	A-----T-----T-----C-----G-----G-----G-----G-----A-----T-----T-----A-----	-----G-----G-----G-----A-----G-----T-----G-----G-----	CTGAG-----	9141	
ISY	A-----A-----GA-----AG-----T-----C-----	-----T-----T-----A-----	CTGA-----	9345	
ST	-----T-----CC-----G-----T-----G-----G-----G-----T-----CT-----	-----A-----G-----T-----G-----A-----A-----	C-----GA-----	9386	
CAN2	A-----CC-----A-----AG-----T-----G-----G-----	-----T-----T-----GGG-----A-----C-----G-----	CTGA-----	9961	
WDS	A-----C-----CC-----A-----AG-----G-----G-----G-----G-----A-----C-----T-----ACT-----	-----G-----G-----G-----G-----G-----T-----G-----G-----A-----	CTGA-----	9412	
KR	A-----T-----T-----CC-----A-----AG-----G-----G-----	-----T-----T-----AG-----G-----T-----T-----GAA-----	-----A-----A-----	9893	
UC1	-----T-----TC-----A-----G-----G-----A-----T-----G-----A-----	-----CA-----CTAGG-----A-----GAC-----G-----CAT-----AA-----C-----GACACTGC-----AC-----AAGAGGGA-----CT-----C-----	-----C-----	9863	
D205	-----C-----T-----C-----GC-----A-----G-----G-----G-----T-----G-----A-----	-----CA-----TGGG-----A-----GAC-----G-----CAT-----AA-----C-----GACACTGC-----AC-----AAGAGGGA-----CT-----A-----C-----	-----C-----	9867	
BHO	-----T-----TC-----A-----G-----G-----G-----T-----G-----A-----	-----CA-----GTA-----G-----AT-----A-----CAC-----AAG-----C-----ACACTGC-----ATAGAGAGGGA-----CT-----T-----	-----T-----	9839	
MM251	-----AAG-----CC-----T-----G-----GGTT-----GAAG-----GG-----A-----CC-----CC-----T-----TTA-----CA-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	GGGA-----CTC-----	9850	
MM32H	-----AAG-----CC-----T-----G-----GGTT-----GAAG-----GG-----A-----CC-----CC-----T-----TTA-----CA-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	AGGA-----CTC-----	9857	
MM1A11AA	-----AAG-----CC-----T-----G-----GGTT-----GAAG-----GG-----A-----CC-----CC-----T-----TTA-----CA-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	AGGA-----CT-----A-----	9855	
MM142	-----AAG-----CC-----T-----G-----GGTT-----AAG-----GG-----AGCC-----CC-----T-----TTG-----A-----G-----C-----G-----	-----C-----G-----A-----C-----G-----A-----	AGGA-----CT-----C-----	9361	
MM239	-----AAG-----CC-----T-----G-----GGTT-----GAAG-----GG-----A-----CC-----CC-----T-----TTA-----CA-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	AGGA-----CTC-----A-----	9859	
MOE	-----AAG-----CC-----T-----G-----GGTT-----GAAG-----GG-----A-----CC-----CC-----CTP-----A-----A-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	GGGA-----CT-----	9343	
SHMPBJ14.441	-----TAGTC-----C-----T-----A-----G-----GGT-----C-----AG-----GG-----A-----CC-----CC-----T-----TTA-----A-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	AG-----A-----CA-----A-----	9654	
SHMPBJ14.15	-----TAGTC-----C-----T-----A-----G-----GGT-----C-----AG-----GG-----A-----CC-----CC-----T-----TTA-----A-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	AG-----A-----CA-----A-----	9655	
SHMPBJ6.6	-----TAGTC-----C-----T-----A-----G-----GGT-----C-----AG-----GG-----A-----CC-----CC-----T-----TTA-----A-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	AG-----A-----CA-----A-----	9647	
SHH4	-----TAGTC-----C-----T-----A-----G-----GGT-----C-----AG-----GG-----A-----CC-----CC-----T-----TTA-----A-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	AGGA-----CA-----	9833	
SHH9	-----TAGTC-----C-----T-----A-----G-----GGT-----C-----AG-----GG-----A-----CC-----CC-----T-----TTA-----A-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	AG-----A-----CA-----AG-----	9352	
STH	-----TCTC-----C-----T-----A-----G-----GGTTG-----AG-----GG-----A-----CC-----CC-----T-----TTA-----GA-----G-----C-----G-----A-----	-----C-----G-----A-----C-----G-----A-----	AGGA-----CA-----	9523	
Nef	F G H K S G L P R K E W R A K L K A R Q I P V B B	Nef CDS end			

		NF-kappa-B			
BEN	GCTGCA GGGACTTTCCAG	AAAGGGCTGTACCAACGGGAG GGACATGGG	AGGAGCTGGTGGGAAACGCC TCATACTTACTGTATATAATGTACCCGC	TCTGTGCTTGT	10055
BEN	A A Q T F Q	K G L S P R E G B G	R S W M G T P	S Y L L Y K C T R	P L H C
BEN	L I G L S	F A G C N Q G	D M F	G A G G E R	H T Y C I M V P
BEN	C D F P I	Q A V T K G	T M	B L V G N A	I L P V S M Y P
GH1	-----GCTGCA-----	-----	-----	-----	-----
D194	A-----	-----	-----	-----	-----
UC2	A-----	-----	-----	-----	-----
ROD	A-----	-----	-----	-----	-----
NIH2	A-----	-----	-----	-----	-----
ISY	A-----	-----	-----	-----	-----
ST	A-----	-----	-----	-----	-----
CAN2	A-----	-----	-----	-----	-----
WDS	A-----	-----	-----	-----	-----
KR	A-----	-----	-----	-----	-----
UC1	A-----	-----	-----	-----	-----
D205	A-----	-----	-----	-----	-----
BHO	A-----	-----	-----	-----	-----
MM251	ATA-----	-----	-----	-----	-----
MM32H	A-----	-----	-----	-----	-----
MM1A11AA	A-----	-----	-----	-----	-----
MM142	A-----	-----	-----	-----	-----
MM239	A-----	-----	-----	-----	-----
MOE	A-----	-----	-----	-----	-----
SHMPBJ14.441	A-----	-----	-----	-----	-----
SHMPBJ14.15	A-----	-----	-----	-----	-----
SHMPBJ6.6	A-----	-----	-----	-----	-----
SHH4	A-----	-----	-----	-----	-----
SHH9	A-----	-----	-----	-----	-----
STH	A-----	-----	-----	-----	-----

HIV2/SIV Complete Genomes

	3' LTR end U3 <-	
BEN	...ATTGAGTTCCTCTGCGGAGAGGCTGGCAGATCGAGCCCTGAGAGGTTCTCTCCAGCACTAGCAGGTAGAGCCTGGGTGTTCCTGCTGACTCTCACCAGTACTTGGCCGGTACTGGGCA...GAGGGCTCCACGCTTGCCTGCTTA	10199
BEN	...I Q S L C G E A G R S S P F R F S P A L A G R A W V F P A G L S P V L G R V M A D G S T L A C L	3232
BEN	...F S S A R L A D R A L R G S L D N S Q V E P G C S L L D S H Q Y L A G T G R L A P R L L A	3163
BEN	...F S S V A L R R G W Q I E P S E V L S S T S R S S L G V P C N T L T S T H P V L Q S R L H A C L L	3211
GH1		9480
D194		9469
UC2	...T-C-.....C-.....C-.....A-.....G-C-.....C-.....	10153
ROD	..AC-G-.....T-.....G-.....G-.....A-.....C-.....G-.....C-.....	9637
N1H2	C-.....T-.....G-.....A-.....G-.....CG-.....	9393
ISY	T-.....G-.....G-.....G-.....CG-.....	9597
ST	T-.....G-.....A-.....G-.....C-.....	9638
CAN2	..AC-.....T-.....G-.....A-.....G-.....C-.....	10213
KD9	..AC-.....	9525
KR	G-.....A-.....	10134
UC1	G-.....A-.....C-.....G-.....	10114
D205		9978
ENQ	G-.....A-.....C-.....G-.....	10089
MM251	T-.....G-.....A-.....C-.....A-.....G-.....GAGT-.....	10090
MM32H	T-.....G-.....A-.....C-.....G-.....GAGT-.....	10118
MM1A11AA	..G-.....T-.....G-.....A-.....C-.....G-.....GAGT-A-.....	10115
MM142	C-T-.....G-.....A-.....C-.....G-.....GAGT-.....	9612
MM239	T-.....G-.....A-.....C-.....G-.....GAGT-A-.....	10120
MM2	T-.....G-.....A-.....C-.....G-.....GAGT-.....	9594
SMMPBJ14.441	A-.....T-.....G-.....A-.....C-.....G-.....GAGT-.....	9927
SMMPBJ14.19	A-.....T-.....G-.....A-.....C-.....G-.....GAGT-.....	9928
SMMPBJ6.6	A-.....T-.....G-.....A-.....C-.....G-.....GAGT-.....	10120
SMNH4	A-.....T-.....G-.....A-.....C-.....G-.....GAGT-.....	10082
SMNH9	A-.....T-.....G-.....A-.....C-.....G-.....GAGT-.....	9603
STM	A-.....T-.....G-.....A-.....C-.....A-.....G-.....T-.....GAGT-.....	9774
BEN	RAGACCTCTCAATAAAG...CTGCCA...GTAGAAGCAAGTTAAGTGTGTTCOCATCTCTCCTAGTGGCCGCTTGGT...CAT...TCGGTGTTCATC...TGAGTAACAAGACCC...TGGTCTGTAGGACCTTCTCGCTTTG	10331
BEN	...K T S S I X L P V R S K L S V C S H L S S S P F G H S V F I S V T R P M S V R T L L A L	3274
BEN	...R P L Q G F C F L E A S S V C V P T S P S R L V F R C S F S S O D I G L L G P F S L W	3178
BEN	K D L F N F S A S K D V A C V F P S L L V A A W S G V H S S N K T S V C S D P S R F	3247
UC2	...C-T-.....A-.....C-.....T-.....T-.....C-.....	10284
ROD	..A-..C-T-.....	9671
N1H2	C-.....A-.....	9431
ISY	..AGA-CTC-T-.....	9636
ST	..AGA-CTC-T-.....	9672
CAN2	C-T-.....C-.....T-.....A-.....T-.....	10344
KR	C-T-.....	10172
UC1	..AC-..C-T-.....A-.....G-.....C-.....T-.....AA-C-TA-T-.....AA-TA-.....T-.....TCT-.....	10243
ENQ	..AC-..C-T-.....A-.....G-.....C-.....T-.....AA-C-TA-T-.....AA-TA-.....T-.....TCT-.....	10214
MM251	T-.....T-.....T-CC-.....C-.....AC-.....AC-CA-T-.....G-.....T-.....C-TCT-.....	10221
MM32H	T-.....T-.....T-CC-.....C-.....AC-.....AC-GTA-.....A-.....G-.....T-.....TCT-.....	10249
MM1A11AA	C-.....T-.....T-CC-.....C-.....AC-.....AC-GTA-.....A-.....G-.....T-.....TCT-.....	10246
MM142	T-.....T-.....T-.....C-.....AC-.....AC-ATA-.....A-.....G-.....T-.....TCT-.....	9646
MM239	C-.....T-.....T-CC-.....C-.....AC-.....AC-ATA-.....A-.....G-.....T-.....TCT-.....	10251
MM2	T-.....T-.....T-CC-.....C-.....AC-.....AC-ATA-.....A-.....G-.....T-.....TCT-.....	9638
SMMPBJ14.441	A-.....G-.....T-.....CC-.....C-.....AC-.....AC-GA-.....CA-.....G-.....C-.....T-.....TCT-.....	9996
SMMPBJ14.15	T-.....T-.....CC-.....C-.....AC-.....AC-GA-.....CA-.....G-.....C-.....T-.....TCT-.....	9996
SMMPBJ6.6	T-.....T-.....CC-.....C-.....AC-.....AC-GA-.....CA-.....G-.....C-.....T-.....TCT-.....	10246
SMNH4	A-.....T-.....CA-.....C-.....AC-.....AC-GA-.....ACA-.....G-.....T-.....TCT-.....	10213
SMNH9	G-.....T-.....T-.....	9630
STM	T-T-T-.....A-.....T-.....T-CA-.....G-.....C-.....AC-.....AC-.....T-.....TCT-.....	9892

HIV2/SIV Complete Genomes

```

BEN      0GAATCCAAAGGCAGGAAATCCCTAGCA
BEN      -I-Q-S-R-R-I-F-S-
BEN      -E-S-K-A-G-K-S-L-A-
BEN      -N-P-R-Q-E-N-P-S-X-
UC2      -----A-G-----
CAM2     ---A-----
UC1      ---A-----
BRD      ---A-----
MN251    A--A--G-A-----
MN32H    ---A--G-A-----
MN1A11AA---A--G-A-----
MN239    ---A--G-A-----
SMNH4    -A--A--G-----

```

```

10359
3294
3187
3256
10312
10494
10271
10351
10277
10277
10274
10279
10241

```

Amino Acid Alignments

Construction of the Protein Alignments

The number of full-length gene sequences is still growing rapidly for all genes. The envelope master alignment now contains almost 450 sequences, more than half of which are full length. For the purposes of the printed alignments, we have had to limit the number of sequences dramatically, although the full set of sequences is still available through our anonymous ftp site (<ftp://ftp-t10.lanl.gov/pub/aids-db>). Here we list the criteria we have followed to make the selection.

First, we have decided not to overrepresent the B clade sequences in the printed alignment. Almost half of the envelope sequences are still subtype B, though the contribution of other subtypes is increasing. We have decided to limit the number of sequences to 20 per subtype. Only subtype B frequently reached this limit, so a further selection had to be made to determine which 20 to include. For subtype B, we have tried to include as many 'classical' sequences as possible. Extensive follow-up work has been done based on lab strains such as HXB2, MN, SF2, and JR-CSF/JR_FL, so these strains are included in the alignments. Furthermore, we have tried to represent sequences from diverse geographical origins, so as to represent a broad spectrum of variants. In the case of subtype B, this means that we have included Thai and Brazilian variants along with the 'Western' strains. We have not specifically attempted to include whole genomes in the alignments, because a separate section of this compendium is devoted to these.

Sequences from isolates that are known recombinants have been included in some cases when it has been established that they are not recombinant in the env gene, and there were only a small number of sequences (notably, subtypes E, F, and G). These mosaic sequences are indicated by a prefix of two letters indicating the subtypes that have contributed to their genome. An exception is the sequences of subtype E; all subtype E isolates found so far are recombinant and show the same mosaic pattern, so the prefix has been omitted. All subtype G sequences found so far are AG recombinants too, but since there exist many different mosaic patterns for subtype G, the prefix in this case has been maintained.

For subtypes other than B, there are less than 20 full-length sequences, and all available full-length sequences are included. In some cases (notably subtype H), we have used all available sequences, even if they are not full-length, because of the paucity of information on this subtype.

Recombinant (other than A/G and A/E) and unclassified sequences have been used in the alignment only when the number of sequences for a protein was very small and the sequence was known to be non-recombinant in that protein. Recombinants other than AE recombinants in subtype E are indicated with a prefix before the sequence name (for example, AD_MAL).

For reasons of standardization and synchronization with the HIV Immunology Database, the reference strain (shown on top of the alignments) is now WEAU instead of the overall consensus. WEAU is a primary isolate that has been dually sequenced and very carefully annotated, and it eliminates the disadvantages of working with an overall consensus (mainly, the existence of undefined positions).

Explanation of Symbols in Alignments	
Symbol	Meaning
Alignment symbols	
? in consensus	no majority-rule consensus could be determined at this position
x	nucleotide missing from codon
#	frameshift, or codon contains N or illegal character
\$	stop codon
Annotation symbols	
-	domain boundaries
/	protein start point
\	protein end point
\ /	splice site or exon join
->	start of overlapping coding region
<-	end of overlapping coding region
*	cysteine
^^^ [NCS, NCT]	glycosylation site
^^^ [NCS, NCT]	glycosylation site with cysteine
CD4	residue critical for CD4 binding
cds	coding sequence (indicates regions where two proteins overlap; the overlapping proteins use two different reading frames)
MHR	major homology region
nls	nuclear localization signal
phos site	phosphorylation site
Zn-motif	Zinc finger abinding motif

Sources of Annotation in the Alignments

Protein	Annotation	Reference
Gag	phos site Ser (111)	Yu, <i>J Biol Chem</i> 270:4792 (1995)
Gag	MHR, (284-302)	Otteken, <i>J Virol</i> 70:3407 (1996)
Gag	CyPa (205-241)	Braaten, <i>J Virol</i> 70:4220 (1996)
Gag	vpr packaging domain LKSLFG (489-494)	Lu, <i>J Virol</i> 69:6873 (1995) Kondo, <i>J Virol</i> 70:159 (1996)
Nef	myristylation (1-7)	Huang, <i>J Virol</i> 69:93 (1995)
Nef	(PxxP)3 (67-76)	Huang, <i>J Virol</i> 69:93 (1995)
Nef	PKC (75-80)	Huang, <i>J Virol</i> 69:93 (1995)
Nef	polypurine tract (89-97)	Huang, <i>J Virol</i> 69:93 (1995)
Nef	Beta turn (128-131)	Huang, <i>J Virol</i> 69:93 (1995)
Nef	PxxP (145-148)	Huang, <i>J Virol</i> 69:93 (1995)
Vpr	alpha helix (16-34)	Cornelissen, <i>ARHR</i> 13:247 (1997)
Vpr	H(S/F)RIG motifs (71-82)	Macreadie, <i>PNAS USA</i> 92:2770 (1995)
Vpu	all annotations	Cornelissen, <i>ARHR</i> 13:247 (1997)
Vpr	LR domain (60-82)	Wang, <i>Gene</i> 178:7 (1996)

PART II Amino Acid Alignments

Introduction	II-1
Contents (1997)	II-4
A. HIV-1 Amino Acid Alignments	
Gag	II-A-1
Pol	II-A-16
Vif	II-A-32
Vpr	II-A-36
Tat	II-A-39
Rev	II-A-42
Vpu	II-A-45
Env	II-A-48
Nef	II-A-68
B. HIV-2/SIV Amino Acid Alignments	
Gag	II-B-1
Pol	II-B-8
Vif	II-B-19
Vpr	II-B-22
Tat	II-B-24
Rev	II-B-27
Vpx	II-B-30
Env	II-B-32
Nef	II-B-50
C. SIV AGM and SIV SYK Amino Acid Alignments	
Gag	II-C-1
Pol	II-C-5
Vif	II-C-11
Vpr	II-C-13
Tat	II-C-15
Rev	II-C-17
Env	II-C-19
Nef	II-C-25

Protein Alignment Summary Table (Gag)

Name	Accession No.	Region	First Author	Reference
CONSENSUS-A				
U455	M62320	complete genome	Oram, J.D.	ARHR 6 (9), 1073-1078 (1990)
92UG037	U51190	complete genome	Gao, E.	J. Virol. 70 (3), 1651-1657 (1996)
BZ126B	L22083	gag	Louwagie, J.	ARHR 10 (5), 561-567 (1994)
VI59	L11795	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
K112	L11768	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
K88	L11773	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
VI32	L11788	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
VI415	L11791	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
CI4	L11757	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
LBV23	L11777	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
TN2431	L03702	gag	McCutchan, F.E.	ARHR 8 (11), 1887-1895 (1992)
TN245	L11762	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
CM240	U54771	complete genome	Carr, J.K.	J. Virol. 70 (9), 5935-5943 (1996)
LBV2310	L11779	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
CI51	L11758	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
IC144	L11767	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
DJ258	L11763	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
CM238	L11760	source gag	Louwagie, J.J.	AIDS 7 (6), 769-780 (1993)
UG266	L11798	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
Q23	AF004885	complete genome	Poss, M., R.A.	Unpublished
CONSENSUS-B				
HXB2	K03455 M38432	complete genome	Rosen, C.A.	Cell 41, 813-823 (1985)
SF2	K02007	complete genome	Sanchez-Pescador, R.	(in) Weiss, R.L., Teich, N., Varmus, H. and Coffin, J. (Eds.), <i>RNA Tumor Viruses</i> , Second Edition, 2, Vol. 2:1124-1141, Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
TB131	L03696	gag	McCutchan, F.E.	ARHR 8 (11), 1887-1895 (1992)
LAI	K02013	complete genome	Wain-Hobson, S.	Cell 40, 9-17 (1985)
MN	M17449	complete genome	Gurgo, C.	Virology 164, 531-536 (1988)

HIV1 GAG

JH31	M21137	5' LTR, gag, pol	Komiyama, N.	ARHR 5, 411 (1989)
JRCSP	M38429	complete genome	Koyanagi, S.	Science 236, 819 (1987)
JRFL	U63632	complete genome	O'Brien, W.A.	Nature 348, 69 (1990)
NL43	M19921	complete genome	Adachi, A.	J. Virol. 59, 284-291 (1986)
CDC451	M13136	5'-LTR, gag	Desai, S.M.	PNAS 83, 8380-8384 (1986)
HAN	U43141	complete genome	Sauermann, U.	ARHR 6, 813-823 (1990)
CAM1	D10112	complete genome	McIntosh, A.A.	Unpublished (1991)
RF	M17451	complete genome	Starcich, B.R.	Cell 45, 637-648 (1986)
D31	U43096	complete genome	Kreutz, R.	ARHR 8, 1619 (1992)
BCSG3C	L02317	complete	Ghosh, S.K.	Virology 194, 858-864 (1993)
P896	U39362	complete genome	Collman, R.	J. Virol. 66, 7517 (1992)
gag46	U29413	gag	Yoshimura, F.K.	J. Virol. 70 (12), 8879-8887 (1996)
MANC	U23487	complete genome	Zhu, T.	Nature 374, 503-504 (1995)
gag314	U29404	gag	Yoshimura, F.K.	Unpublished (1995)
WEAU160	U21135	complete genome	Ghosh, S.K.	Unpublished (1995)
4-7	U29414	gag	Yoshimura, F.K.	Unpublished (1995)
CONSENSUS-C				
ETH2220	U46016	complete genome	Salminen, M.O.	ARHR 12(14), 1329-1339 (1996)
92BR025	U52953	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1667 (1996)
UG268	L11799	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
SM145	L11803	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
ZAM18	L03705	gag	McCutchan, F.	JAIDS 5, 441-449 (1992)
ZAM19	L03706	gag	McCutchan, F.E.	JAIDS 5, 441-449 (1992)
ZAM20	L03707	gag	McCutchan, F.	JAIDS 5, 441-449 (1992)
DJ259	L11764	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
VI313	L11787	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
NOF	U07016	gag	Becker, M.L.	AIDS 7, 23 (1993)
HIVKON	U07238	gag	Becker, M.L.	ARHR 11 (10), 1265-1267 (1995)
CONSENSUS-D				
84ZR085	U88822	complete genome	Gao, F.	Unpublished
94UG114.Ig	U88824	complete genome	Gao, F.	Unpublished
ELI	K03454	complete genome	Alizon, M.	Cell 46, 63-74 (1986)

HIV1 GAG

Z226	M22639	complete genome	Theodore, T.	Unpublished (1988)
NDK	M27323	complete genome	Spire, B.	Gene 81, 275-284 (1989)
VI205	L11785	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
G109	L11765	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
K31	L11771	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
UG274	L11801	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
UG270	L11800	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
SE365	L11797	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
VI203	L11784	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
CONSENSUS-F				
93BR020.17	U27401	env, partial nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
VI174	L11782	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
VI69	L11796	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
BZ162	L11751	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
VI325	L11789	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
BZ163B	L22086	gag	Louwagie, J.	ARHR 10 (5), 561-567 (1994)
CONSENSUS-G				
JV10832	U88826	complete genome	Gao, F.	Unpublished
92NG003.1	U88825	complete genome	Gao, F.	Unpublished (1997)
LBV217	L11778	gag, LBV21-7	Louwagie, J.J.	AIDS 7, 769-780 (1993)
JP88	U13210	gag	Abimiku, A.G.	ARHR 10 (11), 1581-1583 (1994)
JV831	U13212	gag	Abimiku, A.G.	ARHR 10 (11), 1581-1583 (1994)
SE6165	L40752	source gag p17	Leitner, T.	Virology 209 (1), 136-146 (1995)
CONSENSUS-H				
VI525	L11792	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
VI557	L11793	gag	Louwagie, J.J.	AIDS 7, 769-780 (1993)
U40561	AF005496	complete genome	Gao, F.	Unpublished
CONSENSUS-J				
SE7022	L41179	gag p17	Leitner, T.K.	ARHR 11 (8), 995-997 (1995)
SE7887	L41178	gag p17	Leitner, T.K.	ARHR 11 (8), 995-997 (1995)
CONSENSUS-O				
ANT70C	L20587	complete genome	Vanden Haesevelde, M.	J. Virol. 68 (3), 1586-1596 (1994)

HIV1 GAG

MVP5180	L20571	complete genome	Gurtler, L.	J. Virol. 68, 1581-1585 (1994)
CONSENSUS-CPZ				
CPZGAB	X52154	complete genome	Huet, T.	Nature 345 (6273), 356-359 (1990)
CPZANT	U42720	gag, pol, vif, vpr, tat, rev, vpu, partial env	Vanden Haesevelde.C.	Virology 221, 346 (1996)

This page intentionally left blank.

Accession	Protein	Sequence	Position
WEAUL60	MGARASVLSGGELDRWSEKIRLRPGGNKKYKYLKHIWASRELRBFVNPGLLETSGECRQITLQQLQPSLOTGSEELRSLYNTVAVLYCVHQKIEYKDTKEALDKIEEBGNKSRKK	1-100	122
CONSENSUS-A	-----K-A-----K-R?-L-----L-S-----A-Q-ME-SA-K-T-K-F-----T-R-D-----I-Q-----T-----7-	1-100	120
U455	-----KK-S-----R-L-----K-TL-----A-Q-----A-T-----F-R-D-----N-M-N-OR-----T-----N-	1-100	122
92U3027	-----K-A-----K-R-L-----L-S-----T-Q-ME-SA-R-T-----F-R-----I-K-Q-----T-----	1-100	122
B2126B	-----K-A-----K-Q-RN-L-----K-L-S-----A-L-----A-P-T-I-----Y-R-----H-Q-----T-H-----	1-100	123
VI59	I-----E-K-AS-----K-Q-R-L-S-K-----L-S-----T-Q-IE-A-K-T-F-A-----R-DI-----M-Q-----T-----N-	1-100	120
K112	-----K-A-----K-RI-L-----K-L-S-----T-Q-ME-SA-K-T-----F-R-----I-Q-----T-----	1-100	122
K88	-----K-A-----K-RN-L-----D-L-S-----A-Q-ME-SA-K-T-----F-R-----I-Q-----T-----	1-100	122
VI32	-----K-A-----K-RN-L-----D-L-S-----K-Q-ME-ESA-K-T-IK-----F-R-D-----IR-Q-----T-----	1-100	122
VI415	-----K-A-----K-RN-L-----D-L-S-----A-Q-E-A-K-T-IK-----F-R-D-----IK-N-Q-----T-----G-	1-100	123
C14	-----K-A-----K-RN-LI-Q-----L-S-----T-Q-MA-SA-K-T-----F-R-----I-QQ-----T-----	1-100	120
LBV23	-----K-A-E-----K-RN-L-----L-A-----Q-M-TA-T-K-N-----F-R-D-----R-I-K-HQ-----T-----	1-100	122
AE_TM243	-----K-A-----K-RN-L-----L-A-----Q-IE-ST-K-K-F-T-W-----R-R-----V-QQ-----T-----G-	1-100	122
AE_TM245	-----K-A-----R-M-L-----L-A-----RQ-IR-ST-K-K-F-VT-W-----R-----V-QQ-----T-----G-	1-100	122
AE_CM240	-----K-A-----R-R-L-----L-9F-A-----Q-IE-ST-R-L-K-F-T-W-----R-----V-QQ-----T-----Q-	1-100	122
LBV2310	-----K-S-----K-R-L-----L-A-----Q-LNE-SA-R-K-F-I-T-W-----R-D-----L-I-Q-----T-----A-	1-100	122
CI51	-----I-K-S-----R-Q-R-L-----L-S-----A-Q-IE-SA-K-F-T-W-----R-DI-----I-S-Q-----T-----A-	1-100	120
IC144	-----K-A-----K-R-L-----L-F-A-----Q-LNE-ST-K-VK-F-I-T-I-R-DI-----I-Q-----T-----V-	1-100	122
DJ258	-----K-S-----K-R-L-----L-A-----Q-LNE-SA-G-E-K-K-I-W-----R-DI-----AV-Q-----T-----A-	1-100	122
CM238	-----T-K-A-----R-I-L-----L-V-Q-IE-ST-K-K-F-T-W-----R-----V-QQ-----T-----G-	1-100	132
UG266	-----V-E-K-AS-----E-K-L-S-----K-L-S-----T-QR-E-A-A-----R-I-----H-Q-----T-----N-	1-100	120
Q23	-----KF-A-----K-RN-LI-----D-L-S-----AD-Q-NBK-A-K-T-IK-F-----T-R-----IK-Q-----T-----	1-100	122
CONSENSUS-B	-----K-----T-----R-----?-----T-----	1-100	121
HXB2R	-----K-----K-----T-----R-----I-----T-----	1-100	122
SF2	-----K-----T-----R-----D-----E-----T-----AAG-----	1-100	124
TB132	-----Q-K-----K-R-----I-----E-----T-----N-----	1-100	122
LAI	-----K-----T-----R-----I-----T-----	1-100	122
MY	-----N-----K-----V-----T-----K-----T-----I-----E-----T-----	1-100	122
JH31	-----K-----S-----K-F-----T-----R-----I-----E-----T-----M-----	1-100	132
JRCSE	-----K-K-----K-R-----S-----L-A-K-----T-----R-----E-----A-----E-----	1-100	130
JRFL	-----K-K-----K-Q-----K-----I-----R-----D-----T-----R-----	1-100	122
NE43	-----K-----K-Q-R-----K-----K-----T-----R-----R-----T-----	1-100	132
CDC41	-----K-----K-Q-----K-----K-----T-----R-----R-----T-----	1-100	132
MAN	-----K-----K-Q-----M-----T-----V-----N-----E-----A-----	1-100	122
CAN1	-----K-----K-----D-----E-----T-----G-----	1-100	122
RF	-----K-K-----R-K-----S-A-----A-K-A-T-----M-R-----T-----	1-100	132
D31	-----K-R-----K-F-----T-----R-----F-R-----A-----	1-100	122
BCSQ3C	-----K-K-----X-L-----A-----K-I-----M-D-----E-N-----T-----	1-100	122
P896	-----R-----S-----S-----K-----T-----R-----T-----	1-100	122
GAG46	-----E-K-R-----V-----T-----R-----E-----T-----O-----	1-100	122
MANC	-----X-----X-----T-L-----V-----K-----G-----P-----T-----	1-100	132
GAG314	-----R-Q-----D-----A-----K-----T-----D-----V-----R-----T-----	1-100	122
4-7	-----E-K-R-----K-----V-----T-----R-----E-----T-----G-----	1-100	122
CONSENSUS-C	I-R-X-A-----K-R-M?-L-----L-----R-IK-A-T-?-?-T-?-?-R-?-?-T-?-?-A-2	1-100	116
ETH2220	I-R-EK-A-----X-K-H-M-L-N-----K-L-D-D-A-R-IK-A-T-K-F-----T-----I-----E-Q-----T-----G-A-R	1-100	123
92BR025	I-R-X-A-R-K-K-----K-H-M-L-----L-D-----K-MK-A-TX-I-H-----T-----E-D-R-----K-QQ-----T-----E-A-K	1-100	123
UG268	R-K-T-----K-----K-C-M-L-----G-L-S-----K-MK-A-TD-----T-----KG-R-----IQQ-----T-----ET-A-R	1-100	120
SN145	I-R-E-A-R-----K-H-MI-L-----L-----K-MK-A-T-----F-----T-----E-R-----Q-----T-----K-A-	1-100	122
ZAM19	I-K-R-T-----K-H-MI-L-----L-----K-IK-A-T-K-F-VT-W-----ED-T-R-----RL-----QQ-TK-----T-----K-A-	1-100	124
ZAM19	I-X-K-A-----K-H-MI-L-----L-----K-IK-A-T-K-H-A-----KX-T-R-----X-R-----CQ-----T-----RE-A-	1-100	122
ZAM20	I-R-K-R-----K-H-MI-L-----L-----AD-R-IR-N-A-----T-----AG-R-----E-----CQ-----T-----RT-A-	1-100	122
DJ259	I-R-EK-A-----K-R-R-M-L-----L-S-----K-IR-R-R-----T-----AR-IQ-----CQ-----T-----X-ST-A-K	1-100	123
VI313	I-R-K-T-R-K-K-----K-H-M-L-----L-D-----Q-E-IKK-A-TK-I-W-----T-----E-IR-----K-QQ-----T-----A-N	1-100	123
MCP	I-R-EK-A-----K-R-M-----L-----L-----K-MK-A-T-K-F-----T-----E-R-----TVQ-----T-----K-A-	1-100	122
KON	I-R-K-K-----K-H-M-L-----L-----A-K-IK-A-T-----F-----T-----KN-R-----CQ-----T-----KE-AD	1-100	123

HIV1 GAG

		Major homology region														p24 1/ p2																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		[- (MHR) -]																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
WEAU160	DQIAMTH..	NPP	IPV	GEI	YKRN	I	L	N	K	I	V	M	S	P	S	I	L	D	I	R	C	P	K	E	P	R	D	V	D	R	F	Y	K	T	L	A	D	Q	T	Q	E	V	K	N	M	T	E	T	L	L	V	Q	N	A	N	P	D	C	K	T	L	K	A	L	G	P	G	A	T	L	E	M	T	A	C	Q	G	Y	G	G	P	G	H	K	A	R	V	L	A	E	A	N	S	Q	V		370																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
CONSENSUS-A	---	G	---	S	---	D	---	---	---	---	F	---	S	---	R	---	?	---	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	

HIV1 GAG

		Major homology region										p24 1/ p2		
		(- (MHR) -)												
W2A0160	EQIAMWTH...	NPPIPVGEIYKRWIILGLMKIVRNYS	PVSILDIIRGQPKPEFRDYVDFYNTLRAEQATQEVKMMMTETLLVQWNPDCRTILKALGPQATLEIRNTACQGVGGPGHKARVLAEAMSOV.											370
CONSENSUS-D		--G--S--					S-D				E	S	A	367
B42R085		--G--S--					S-D				E	S	A	370
94UG114.1		--G--S--					S-D				E	S	A	370
ELI		--S--				V	S-D				Q	E	A	370
Z226		--S--					S-Q				Q	E	A	371
NDR		--E--					S-P				Q	E	A	367
VI205		--S--				KK	S-D				Q	KE	E	371
G109		--A--				K	S-D				A	E	A	373
K31		--G--					S-D				E		A	370
UG274		--G--S--					S-D			R	E	S	A	370
UG270		--VG--S--					S-D				A	E	A	370
SE365		--G--								N	E	E	A	370
VI203		--G--S--									E	S	A	370
CONSENSUS-F		--Q--S--				V	F	G	D		E		A	363
93BR020.1		--Q--G--				V	F	G	D		E	S	A	366
VI174		--Q--				V	F	G	D		E		A	365
VI69		--Q--S--				D	F	E	G	D	I	N	E	366
BZ162		--Q--S--				V	F	G	D		E		A	365
VI325		--T--S--				V	F	G	D		S	E	I	366
BZ163B		--Q--S--				V	F	G	D		Q			338
CONSENSUS-G		--R--S--					F		D		R	E	S	362
AG_92NG003		--R--S--					F		D		R	E	S	369
AG_93NG003		--T--S--					F		D		R	A	E	370
AG_LBV217		--R--S--					F		D		E	S	A	368
AG_JP08		--R--S--												198
AG_JV031		--R--S--												199
CONSENSUS-H		--G--				A	D		K	F	D	D	R	367
VI525		--G--				A	D		K	F	D	D	N	370
VI557		--E--					D		K	F	A	D	R	371
90CR056		--G--				A	D		K	F	D	D	R	370
CONSENSUS-O		--?--T--R--				??	??	D	RK	V	M	K	Q	332
ANT70		--H--T--R--				P	Q	D	RK	V	M	K	Q	370
MVP5180		--I--T--R--				GA	S	D	RK	V	M	K	Q	370
CONSENSUS-CPZ		--??-???				???	???	D	R	V	M	K	Q	249
CPZGAS		--G--T--A--						D	R	V	M	K	Q	379
CPZANT		--NQ--STPQQ--GGV--				D		M	V	X	E	K	I	390

HIV1 GAG

	p2 \/ p7	1- Zn-motif -1	1- Zn-motif -1	pol cds start ->	p7 \/ p6	vpr binding	
NEAU140	..TN.T.AT.NRMORONE	..RSPRKTIKCFNCOKEGHTIARNCTAPRKRCQWCKGQEGHMKDC7	..E.RQANFLGRINSSQMG	..RPGNFPQSRL	..SPTAPPEE	..SP	..RPRESTTTPS
CONSENSUS-A	H-??...	I-...GQKR?	L-...K-...	?	...	...	A-???GMS-I-7SP7
U455	Q-S-...I-...	G-R-...L-R-...	K-...P-M-...	...	...	...	A-I-GMS-KN-SP
92UGD37	H-...I-...	KQQKR-...L-R-...	R-...P-S-...	...	...	...	AAEI-GM-IV-SP
B2126B	N-ASP-...I-...	QKR-...L-...	K-...P-H-...	...	...	...	A-E-SA
VI59	Q-S-...I-...	KQQKR-...L-...	S-RK-...P-...	...	...	...	A-IL-I-K-M-SP
K112	H-...I-...	GQKR-...L-...	R-...K-...	...	...	...	A-I-GMS-MA-SP
K88	H-...I-...	GQKR-...L-...	K-...G-...	...	...	...	A-IC-GMS-IA-SP
VI32	Q-A-...I-...	GGS-R-...L-...	K-...N-...	...	...	...	A-IC-GMS-IA-SP
VI415	H-...I-...	KQO-R-...L-...	K-...G-...	...	...	...	A-IV-GMS-I-SP
CI4	H-...I-...	K-...GQKR-...L-...	K-...P-S-...	...	...	...	A-LL-GMS-I-S-
LBV23	H-...I-...	GQKR-...L-...	K-...P-S-...	...	...	...	A-HIFGMS-I-SP
AE_TN243	H-A-...I-...	KQQKR-...L-...	R-...K-...	...	...	...	A-DN-GMS-8U
AE_TN245	H-A-...I-...	KQQKR-...L-...	K-...P-M-...	...	...	...	A-NM-GMS-I-SL
AE_CM240	K-AT-...I-...	KQQKR-...L-...	Q-...K-...	...	...	...	A-NM-GMS-I-GESE
LBV2310	Q-...I-...	GQ-T-...L-...	K-...R-...	...	...	...	A-GMS-I-S-
CI51	Q-S-...I-...	GQ-T-...L-...	K-...R-...	...	...	...	A-GMS-I-S-
IC144	Q-S-...I-...	GQ-T-...L-...	K-...R-...	...	...	...	A-GMS-K-S-
DJ250	Q-S-...I-...	D-GQ-T-...L-...	K-...P-T-...	...	...	...	LA-GMS-I-SP
CM238	Q-AS-...I-...	KQQKR-...L-...	R-...K-...	...	...	...	A-NM-GMS-I-SL
UG266	Q-S-...I-...	KQQKR-...L-...	R-K-...S-KK-...	...	...	...	A-IL-NM-M-SP
Q23	K-...I-...	KQQKR-...L-...	K-...P-R-...	...	...	...	A-TC-GMS-V-SP
CONSENSUS-B	...S-?I-...	...NQ-?V-...	...K-...?...	...	...	...	...G-?...
MXB2R	...S-...I-...	...NQ-IV-...	...K-...T-...	...	...	...	...SGV-...
SP2	...P-N-I-...	...NQ-V-...	...K-...R-...	...	...	...	...O-K-...
TB132	...S-...I-...	...NQ-IV-...	...K-...R-...	...	...	...	...A-...
LAI	...S-...I-...	...NQ-IV-...	...K-...P-Y-...	...	...	...	...SGV-...
MM	...S-...I-...	...NQ-I-...	...K-...R-...	...	...	...	...C-...YQ
JH31	...S-T-I-...	...NQ-I-...	...K-...N-...	...	...	...	...G-...
JRCSP	...P-...I-...	...NQ-IV-...	...K-...E-...	...	...	...	...G-A-Q
JRFL	...P-...I-...	...NQ-IV-...	...K-...P-Y-...	...	...	...	...K-...G-A-...
NL43	...P-...I-...	...NQ-V-...	...K-...P-H-...	...	...	...	...G-...
CD41	...S-...I-...	...NQ-V-...	...K-...P-H-...	...	...	...	...G-...
NAM	...G-S-A-I-...	...NQ-V-...	...K-...P-H-...	...	...	...	...G-A-A-...
CAN1	...S-...I-...	...NQ-V-...	...K-...P-H-...	...	...	...	...G-K-...
RF	...S-...I-...	...NQ-IV-...	...K-...N-G-...	...	...	...	...G-...
D31	...S-...V-...	...NQ-V-...	...K-...P-Y-...	...	...	...	...G-A-F
BCSG3C	...S-...I-...	...NQ-V-...	...K-...P-H-...	...	...	...	...G-...
FB96	...S-...I-...	...NQ-V-...	...K-...P-H-...	...	...	...	...G-...
GAG46	...S-N-I-...	...NQ-V-...	...K-...P-H-...	...	...	...	...G-M-P
MANC	...S-...I-...	...NQ-V-...	...K-...P-H-...	...	...	...	...G-A-...
GAO314	...S-ATT-...	...NQ-I-...	...K-...I-...	...	...	...	...G-...
4-7	...S-...I-...	...NQ-IV-...	...K-...P-K-...	...	...	...	...G-A-SP
CONSENSUS-C	H-...I-...	...KQ-KR?	L-...K-...	...	...	...	...P-?...
ETH2230	H-T-...I-...	...KQ-KRA-...	L-...K-...	...	...	...	...PEPTAPP-ES-LRP-...
92BR025	H-...I-...	...S-C-KQ-KR-...	L-...K-...	...	...	...	...P-NR-LL-N-T-...
UG260	H-I-...I-...	...KQ-KRI-...	L-...K-...	...	...	...	...P-NR-LL-N-T-...
SM145	H-I-...I-...	...KQSKRIV-...	L-...K-...	...	...	...	...P-N-PEPTAPP-ES-FRVP-...
ZAH18	H-...I-...	...KQ-KASV-...	L-...K-...	...	...	...	...P-K-PEPTAP-...
ZAH19	S-X-...I-...	...KQ-KIV-...	L-...K-...	...	...	...	...P-H-K-X-XXA-...
ZAH20	S-A-...I-...	...KQ-KR-V-...	L-...K-...	...	...	...	...P-N-PEPTAPP-ES-FR-...
DJ259	H-...I-...	...KQ-...	L-...K-...	...	...	...	...P-N-PEPTAPP-ES-FR-...
VI313	H-I-...I-...	...KQ-KRI-...	L-...K-...	...	...	...	...P-Y-LL-N-P-...

HIV GAG

	p3 \/ p7	1- Zn-motif -1	pol cds start ->	1- Zn-motif -1	p7 \/ p6	vpr binding				
NEAU160	..TN.T.AT.NMMQRENF	..ASPERTIKCF	..ACGREGHIA	..MCRAPRKGG	..CWNKGGQEGNQHKDCP	..E.RQANFLGKIWSSQK	..G.RPGNFPQSR	..EPTAPFEE	..SF.RPRETTTTS	473
CONSENSUS-D	..SNTA.I	..KG	..I	..K	..K	..F-H	..L-P	..??	..A-..G-G-I	470
842R083	..S.AS-A.I	..KS	..KOT-IV	..K	..K	..F-H	..L-P	..A	..G-G-I	472
94UG114.1	..SNTA.I	..KG-K-I	..T-K	..R	..R	..P-HN	..L-P	..PA	..A-I..GLQ-I	475
ELI	..SVT-A	..KG	..I	..K	..R-K-L	..R-P-H	..L-P	..A	..G-G-I	473
Z226	..SA-A.V	..KG	..K	..R	..K-L	..P-H	..L-P	..A	..G-G-I	474
NDK	..G.SATA.V	..KG	..S	..T-K	..R	..P-H	..L-P	..A	..G-G-I	470
VI205	..S.SATA.VI	..KG-K-I	..K	..K	..R-I	..P-L	..L-P	..AK	..G-G-I	474
G109	..S.SH-A.I	..K	..KG	..VV	..R	..K	..R	..P-N	..L-P	476
K31	..GHTA.I	..KGO	..I	..L-K	..R	..RL-P-R	..G-L-P	..A	..G-G-I	472
UG274	..SNTA.I	..KG	..I	..K	..K	..P-H	..L-P	..A	..G-G-I	473
UG270	..S.AN-A.I	..A	..KG	..KRN	..L-K	..P-H	..L-P	..A	..G-G-I	472
SE365	..S.SA-A.I	..KGO	..IV	..K	..K	..P-S	..L-P	..A	..G-G-I	473
VI203	..SVA.I	..KG	..KRI	..L-K	..K	..P-H	..L-P	..A	..G-G-I	473
CONSENSUS-F	..??-A.I	..KS	..KGO	..RIV	..K	..P-N	..L-P	..A	..G-G-I	464
93BR020.1	..S.A.I	..KS	..KGO	..RIV	..K	..P-N	..L-P	..S-A	..G-G-I	467
VI174	..S.A.I	..KS	..KGO	..RIV	..K	..P-N	..L-P	..A	..G-G-I	468
VI89	..S.A.I	..KS	..KGO	..RIV	..K	..P-N	..L-P	..S-A	..G-G-I	467
BZ162	..S.A.I	..KS	..KGO	..RIV	..K	..P-H	..L-N-P	..A	..G-G-I	466
VI325	..S.A.I	..KS	..KGO	..RIV	..K	..P-N	..L-P	..A	..G-G-I	467
CONSENSUS-G	..SG?-A-A.I	..K?	..KG	..R?	..L	..P-N	..L-N-P	..A	..G-G-I	463
AG_92NG083	..SGAAA-A.I	..KS	..KG	..RI	..L	..P-N	..L-N-T	..A	..G-G-I	473
AG_93NG003	..G-S-A.I	..KN	..KG	..RG	..L	..P-N	..L-N-P	..A	..G-G-I	473
AG_LBV217	..SG-ATA.I	..K	..KG	..N	..D	..P-H	..L-N-P	..S-A	..G-G-I	471
CONSENSUS-H	..S-A.I	..K	..KG	..RIV	..K	..P-S	..L-P	..A	..G-G-I	469
VI525	..S-A.I	..K	..KG	..RIV	..K	..P-S	..L-P	..A	..G-G-I	473
VI557	..S-A.I	..K	..KG	..RIV	..K	..P-S	..L-P	..A	..G-G-I	472
90CR056	..S-A.I	..K	..KG	..RIV	..K	..P-S	..L-P	..A	..G-G-I	473
CONSENSUS-O	QDLKGGYTA.VF	..QN	..P?RK?	..K	..R	..Y-PGGT	..YV-???	..S-M	..YV-PQ	422
AMT70	QDLKGGYTA.VF	..QN	..P?RK?	..K	..R	..Y-PGGT	..YV-RPA	..S-M	..YV-PQ	472
MVP5180	QDLKGGYTA.VF	..QN	..P?RK?	..K	..R	..Y-PGGT	..YV-RQV	..S-M	..YV-PQ	472
CONSENSUS-CP2	?-?Q-???	..VF?	..?G??G?	..K	..R	..Y-PGGT	..YV-???	..S-M	..YV-PQ	319
CP2GAB	..NQ.GR-D.VFP	..K	..QG	..AG	..K	..G	..V	..S-M	..YV-PQ	482
CP2ANT	N.AQ..GTA.VFL	..QN	..GKKRL	..K	..R	..Y-PGGT	..YV-KEEVV	..S-M	..YV-PQ	498

HIV1 GAG

	vpr binding -	p6 terminus /
WEAU360	...QKQEPID...KELY..PLTSLKSLFGMDPSSQ	500
CONSENSUS-A	??P---QK---R-QTP--V-----L--	490
U455	...A---LK---R-QT--V-----L--	493
92UG037	...P---QN---RDQNP-SV-----LL--	499
B2126B	...P---LK---R--	478
V159	...P---QT-----SI----	477
K112	...P---QK---R-QTP-SI----	489
K86	...P---QR---R-QTP--V----	489
V132	...P---QK---R-QBQ--V----	490
VI415	...P---Q-----R-QAP--V----	486
CT4	...P---Q-----R-QP--SI----	484
LBV23	...P-K-QK---R-QPL-SV----	490
AE_TN243	...L---QK---DPP--V-----	495
AE_TN245	...L---QK---HPP--V-----	489
AE_CNC40	TSLP---QK---HPP--V-----L--	503
LBV2310	...P---R---G--P-----	486
CT51	...LQ---R---G--P--A----	484
IC144	...PQ---R---P-----	495
DJ258	...P---RE---QG--P-----	486
CM238	...L---QK---HPP--V-----	489
UG266	...P---RQR---QT--I----	480
Q23	...L---QK---R-QAQ--V-----LL--	498
CONSENSUS-B	??P---A---R-----	497
HXB2R	-----R-----	500
SF2	-----R-----	502
TB132	-----T-----A-----	497
LAI	-----R-----	512
WI	KQEK---T---D---A---L--	506
JH31	-----R---A---R-----P	500
JRCSP	KQE-----R-----	504
JRFL	-----M---R-----	498
ML43	-----A---R---S-----	500
CDC41	-----R---A---R-----	500
KAN	-----A---S-----	499
CAM3	-----A---R-----	500
RF	-----K---A-----	501
D31	-----A---R-----	500
BCSQJC	-----RE---M---A---R-----	500
P896	-----A---R-----	500
GAG46	-----R-----	500
MANC	-----A---RP-----	500
GAG314	-----Q--R---A---R-----	501
4-7	...L-----R-----L--	500
CONSENSUS-C	...P---K-????-???	483
ETH2220	...P---LK---R---A-----HLL-	504
92BR025	...R---T-----S--L--T	496
UG268	...P---K---R---I----	489
EM145	...P---LKE-----	495
ZAM18	...EN---K---R-----	492
ZAM19	...EN---K---R-----	428
ZAM20	...P---R---SPTGE--A----	496
DJ259	...---LK-KEPY-----A-R-	488
VI313	...P---K-----	484

HIV1 GAG

	vpr binding / (- - -) /	p6 terminus
MEAU160	...QKQEPID...KELY.PLTSLKSLFGWDFSSQ	500
CONSENSUS-D	...-QK-???-...-A-...-L-	497
B42R085	...-QK-...KD-...-A-...-L-	501
94UG114.1	...-QK-...-...-...-L-	502
ELI	...-QK-...-...-...-L-	500
2226	...-QK-...-...-S-A-...-L-	501
NDK	...-QK-...-...-A-...-L-	497
VI205	...-QK-...-H-...-...-L-	492
G109	...-QK-REKD-...-R-	498
K31	...-QK-...-H-...-A-...-L-	490
UG274	...-QK-...-PC-...-A-...-L-	491
UG370	...-QK-...-SAA-...-L-	490
SE365	...-QK-...-A-...-L-	491
VI203	...A-...KE-...-VA-...-L-	490
CONSENSUS-F	...P-...QK-...EG-...P-A-...-S-	491
93BR020.1	...P-...QK-...EQ-...P-A-...-S-	494
VI174	...P-...QKG-...EQ-...P-A-...-L-	487
VI69	...P-...QK-...EQ-...SP-A-...-L-	486
B2162	...P-...QKE-...EQ-...P-A-...-L-	485
VI325	...PR-...TK-...-QGS-...-L-	486
CONSENSUS-G	...P-...PE-...-...-?-...-S-	487
AG_52NG083	...P-...KE-...-...-S-...-S-	499
AG_93NG003	...L-...RE-...-SP-...-S-...-S-	499
AG_LBV217	...P-...-...-...-...-...-L-	487
CONSENSUS-H	...P-...??-...?-?-P-A-...-S-LL-	492
VI525	...P-...R-...-...P-A-...-L-	490
VI557	...P-...RK-...A-...S-...-L-	489
90CR056	...P-...QLK-...DKE-...P-A-R-...-S-LL-	500
CONSENSUS-O	...EN-7QKG-...???-...-FA-...-T-Q	444
AMT70	...EN-QKG-...GMN-...-FA-...-T-Q	498
MVP5180	...EN-SQKG-...DQE-...-FA-...-T-Q	498
CONSENSUS-CPZ	...?-????-...?-?P-??-...-?-??-	533
CP20AB	...E-K-GSS-...S-...P-...-S-...-L-	508
CP2ANT	...-GLKGE-...E-...P-SY-...-E-Q	532

HIV1 POL

Protein Alignment Summary Table (Pol)

Name	Accession No.	Region	First Author	Reference
CONSENSUS-A				
92UG037	U51190	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
U455	M62320	complete genome	Oram, J.D.	ARHR 6 (9), 1073-1078 (1990)
Q23	AF004885	complete genome	Poss, M.,R.A.	Unpublished
92UG059	AF009403	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92UG031	AF009398	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92RW008	AF009407	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92RW016	AF009409	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92UG035	AF009399	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH003	AF009382	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH021	AF009391	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH007	AF009384	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH009	AF009385	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH011	AF009386	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH015	AF009387	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH019	AF009389	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH005	AF009383	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH020	AF009390	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH022	AF009392	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92TH024	AF009394	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
CONSENSUS-B				
SF2	K02007	complete genome	Sanchez-Pascador, R.	(in) Weiss, R.L., Teich, N., Varmus, H. and Coffin, J. (Eds.), <i>RNA Tumor Viruses</i> , Second edition, 2, Vol. 2:1124-1141, Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
LAI	K02013	complete genome	Wain-Hobson, S.	Cell 40, 9-17 (1985)
HXB2	K03455	complete genome	Rosen, C.A.	Cell 41, 813-823 (1985)
NL43	M19921	complete genome	Adachi, A.	J. Virol. 59, 284-291 (1986)
MN	M17449	complete genome	Gurgo, C.	Virology 164, 531-536 (1988)
JRCFS	M38429	complete genome	Koyanagi, S.	Science 236, 819 (1987)

JRFL	U63632	complete genome	O'Brien, W.A.	Nature 348, 69 (1990)
P896	U39362	complete genome	Collman, R.	J. Virol. 66, 7517 (1992)
WEAU160	U21135	complete genome	Ghosh, S.K.	Unpublished (1995)
92BR017	AF009371	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR018	AF009372	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR020	AF009374	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
L.L0	AF027718	pol	Wong, J.K.	Science 278 (5341), 1291-1295 (1997)
92BR004	AF009370	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR028	AF009380	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR003	AF009369	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR021	AF009375	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR024	AF009377	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR030	AF009381	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR026	AF009379	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR019	AF009373	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
AP6.4	AF011754	pol	Tachedjian, G.	Unpublished
CONSENSUS-C				
92BR025	U52953	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1667 (1996)
ETH2220	U46016	complete genome	Salminen, M.O.	ARHR 12(14), 1329-1339 (1996)
92RW026	AF009410	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92BR023	AF009376	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
CONSENSUS-D				
ELI	K03454	complete genome	Alizon, M.	Cell 46, 63-74 (1986)
Z2Z6	M22639	complete genome	Theodore, T.	Unpublished (1988)
NDK	M27323	complete genome	Spire, B.	Gene 81, 275-284 (1989)
94UG1141	U88824	complete genome	Gao, F.	Unpublished
92UG001	AF009395	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92UG038	AF009400	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92UG021	AF009396	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92UG024	AF009397	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92UG046	AF009401	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92UG067	AF009405	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)

HIV1 POL

92UG065	AF009404	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92UG070	AF009406	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
92UG053	AF009402	pol	Cornelissen, M.T.	J. Virol. 71, 6348-6358 (1997)
CONSENSUS-F				
93BR020	AF005494	complete genome	Gao, F.	Unpublished
CONSENSUS-G				
92NG083	U88826	complete genome	Gao, F.	Unpublished (1997)
92NG003	U88825	complete genome	Gao, F.	Unpublished (1997)
CONSENSUS-H				
U40561	AF005496	complete genome	Gao, F.	Unpublished
CONSENSUS-O				
ANI70C	L20587	complete genome	Vanden Haesevelde, M.	J. Virol. 68 (3), 1586-1596 (1994)
MVP5180	L20571	complete genome	Gurtler, L.	J. Virol. 68, 1581-1585 (1994)
CONSENSUS-CPZ				
CPZGAB	X52154	complete genome	Huet, T.	Nature 345 (6273), 356-359 (1990)
CPZANT	U42720	gag, pol, vif, vpr, tat, rev, vpu, partial env	Vanden Haesevelde, C.	Virology 221, 346 (1996)

This page intentionally left blank.

HIV1 POL

```

/ protease <- gag cds end
WEAU160  PPPE.DLVFPKKGKAREF.....SSE..QFRINSPTRR.ELQVQGRDNNSLE...BAQANR..QG...AVSFNFPQITLWQRFPLVTIKIEQQLKEALLDTGADDTVLEDMNLPGKWKPKMIGG 105
CONSENSUS-A ---.N-A-QQ?E--?-.---?---S??-ND?--?-.?--P?..?--?--E---?..??-S-----V-G---??-----I-----?--- 88
92UG037  ---.N-A-QQRE--K-.....---SSRD-WDE---P6.....E---P..EPT-S-----V-G---K-----I----- 106
U455     ---.N-A-QQ-S-.....-A---S-.N-WDG-K-D..-P...C-T-E.....TD-S-----V-G-I-----I-----I--- 104
Q23      ---.N-A-Q--E-----..-G--S-S-.D-NDG--S..-P...S---E---T..GPTLS-----VR-G-----I----- 105
92UG059  .....V---MT-----EI----- 45
92UG031  .....V-G-----ID----- 45
92RW008  .....VR-G-----P-----R----- 45
92RW016  .....V-G-----I----- 45
92UG035  .....G-----I----- 45
92TH003  .....V-G-----I----- 45
92TH021  .....V-Q-----I----- 45
92TH007  .....V-G-----I----- 45
92TH009  .....V-G-----I----- 45
92TH011  .....V-G-----I----- 45
92TH015  .....V-G-----I-T----- 45
92TH019  .....V-G-----I----- 45
92TH005  .....V-Q-----I----- 45
92TH020  .....V-G-----I----- 45
92TH022  .....G-----I----- 45
92TH024  .....V-G-----I----- 45
CONSENSUS-B ---a-q---222222222222---s-----w-----2222--d-??-...t--s-----e-----r----- 105
SF2      ---A-LQ-----SSQTRANSPTI---A-----W-GE-----D.....T-----R-G-----E-S-----R----- 103
LAI      ---A-LQ-----SSQTRANSPTI---A-----W-GE-----D.....T-----R-G-----E-S-----R----- 117
HXB2R    ---A-LQ-----SSQTRANSPTI---A-----W-GE-----D.....T-----R-G-----E-S-----R----- 105
NL43     ---A-Q-----A-----W-GE-----D.....T-----R-G-----E-S-----R----- 105
MN       ---A-LQ-----N-A-----W-GE-----D.....T-----R-G-----E-S-----R----- 108
JRC8F    ---A-LQ-----?-----A-----W-GE-----D.....T-----R-G-----E-S-----R----- 109
JRFL     ---A-LQ-----A-----K-----W-GE-----D.....T-----R-G-----E-S-----R----- 105
P896     ---.N-A-Q-----A-----W-GE-----D.....T-----R-G-----E-S-----R----- 105
92BR017  .....R-G-----E-S----- 45
92BR018  .....R-G-----E-S----- 45
92BR020  .....S-G-----E-S-----R----- 49
L.LO     .....R-G-----E-S-----R----- 44
92BR004  .....R-G-----E-S-----R----- 45
92BR028  .....R-G-----E-S-----R----- 45
92BR003  .....R-G-----E-S-----R----- 45
92BR021  .....A-G-----I-----EI-----R----- 44
92BR024  .....G-----I----- 45
92BR030  .....VG-----R----- 45
92BR026  .....G-----S-----R----- 45
BF_93BR019 .....G-----E-S-----R----- 45
AP6.4    ..... 0

```


		/ protease <- gag cds end	
WEAU160	FFRE.DLVFFPKGKAREF.....SEE.QTRTNSPTRR.ELQVQGRDNNLSL...EAGANR..QG...AVSFNFPQITLWQRLPLVTIKIEGDLKEALLDTGADDTVLEDNNLPQKMKPICHIGG		105
CONSENSUS-C	---?A-Q?-???	---?A-Q?-???	81
92BR025	---N-A-Q-E-KS.....-N-A-----W-----DD-----TA.L-----N-VG-----EIK--H-----		104
ETM2220	---T-A-QQ-----ESEZTRANSP..TR-S-----A-----T-----E-S-----TF-----E-----L-----G-----E-----		112
92RW026		I-G-----E-----	44
92BR023		V-VG--R-----EIK--N-----	45
CONSENSUS-D	---?A-Q-Q-?.....??-A--S-?-R-W-?-P-----T-E-----T-----E-----		100
ELI	---N-A-Q-Q-L.....-PK--A--S--R-W--P-----KT-E-----T-----A-G-----E-----		104
2226	---A-Q-Q-L.....-A--S--R-W--P-----T-E-----T-----C-----G-----E-----		104
NDK	---A-Q-Q-G.....-A--S--R-W-G--P-----T-E-----T-----S-----G-----E-----		104
94UG114.1	---N-A-QW-----P-QTPS-A--S-RO-RIR-Q--TS-----T-E-----T-----L-----V-V-G-----E-----		106
92UG001		G-----I-----	45
92UG038		E-----R-----E-----	45
92UG021		V-G-----E-----	45
92UG024		V-G-----EID-----	45
92UG046		V-G-----E-----	45
92UG067		I-----G-----EIK-----	45
92UG065		G-----E-----	45
92UG070		F-V-G-----E-C-----	45
92UG053		G-----E-----	45
CONSENSUS-F	---N-A-QQ-E-KL.....HP---A-AV--AS.R--R-G--PI-----E--R-TVP--LS-----RVG-----V-----		105
93BR020.1	---N-A-QQ-E-KL.....HP---A-AV--AS.R--R-G--PI-----E--R-TVP--LS-----RVG-----V-----		105
CONSENSUS-G	---N-A-QQ-E-??.....-?-?-A-?-?-R-RTRG-S.EPP-----?G,?-...??-?-?-----V?-G--I-----?I-----		89
AG_92NG083	---N-A-QQ-E-KL.....-P--D-A--S.R--RIRG-S.P-P-----KG..E--I-L-----V-G--I-----GI-----		104
CONSENSUS-H	---N-A-QQRE-K.....-P--A-----S.R--R-RG-D.E-----AB..GQ.GT--LS-----V-----R-----E-----		104
9DCR056	---N-A-QQRE-K.....-P--A-----S.R--R-RG-D.E-----AB..GQ.GT--LS-----V-----R-----E-----		104
CONSENSUS-Q	---?A-ASGGHE-QL.....CA--TS?PI--?.....G-GSE.....GT?-S?E?Q??R...???CL--P--D--I--A?VG-H-C-?-----NMIQ-E-?-?-----		84
ANT70	---Q.I-ASGGHE-QL.....CA--TS?PI--D.....G-GSE.....GTG-S-TS-GPER...-L-VCL--P--D--I--ARVG-H-C-V-----NMIQ-E-----		101
NVP5180	---V-ASGGHE-QL.....CA--TS?PI--M.....G-GSE.....GTR-SBSGGGGR...-PICL--P--D--I--A-VG-E-C-----NMIQ-E-R-T-----		101
CONSENSUS-CP2	---???	CA-???	48
CP2GAB	---R-A-ORE-QL.....CA--N--G--D--N-P-GRE.....-P-ES-GREQ...SI-T-L-----IPV-V--C-----I-RIQ-Q-L-----		103
CP2ANT	---T-PHVGVQT--L.....CA-GG5GA--S-H-D-SGGAQEDSGG.....--GGTT-LV--E-P----MHEVL-Q--RCQ-----V-GIH-Q-N--T--		105

HIV1 POL

	protease \ / p51, p66 Reverse Transcriptase	
WEAU160	IGGFIKVRQYDQVFEICGSKAIGTVLWGPTFVNIIGRNLLTQIGCTLMFFDISPIETVPVKLRPGHOGPKVKQWPLTEEKIRALVEICTENEBGKISXIGPENPYNTPVFAIRKKDSTKWRKLVDFREL	235
CONSENSUS-A	-----P-----IL-----K-?-?-----H-?-?-----S-----RI-----T-----AD-R-R-----I-----	206
92UG037	-----K-----IL-----K-----H-----L-----S-----RI-----T-----AD-R-R-----I-----	236
U455	-----IL-----K-T-----K-----E-----T-----N-----R-----I-----	234
Q23	-----K-----IL-----K-----N-----E-----T-----I-----	235
92UG059	-----IL-----K-----N-----D-----T-----K-----T-----I-----	175
92UG031	-----K-E-IL-----K-----K-----D-----T-----D-----I-----	175
92RW008	-----K#-IL-----K-----K-----S-----R-----I-----	173
92RW016	-----IL-----K-----N-----D-----T-----D-----R-----	175
92UG035	-----K-----IL-----K-----N-----T-----RI-----T-----R-----	175
92TH003	-----IL-----K-----K-----H-----D-T-----T-----K-E-----	175
92TH021	-----IL-----K-----K-----D-----T-----T-----K-E-----	174
92TH007	-----IL-----K-----K-----D-----T-----T-----K-E-----	175
92TH009	-----I-----K-V-----H-----D-T-----T-----K-E-----I-----	175
92TH011	-----IL-----K-----N-----D-----T-----T-----K-E-----	175
92TH015	-----IL-----K-----K-----D-----T-----T-----K-E-----	175
92TH019	-----IL-----K-----K-----D-T-----T-----K-E-----	175
92TH005	-----IL-----K-----N-----D-----T-----T-----K-E-----	175
92TH020	-----IL-----K-----N-----D-----T-----T-----K-E-----G-----	175
92TH022	-----IL-----K-----K-L-----D-T-----T-----K-E-----	175
92TH034	-----IL-----K-----K-----D-T-----T-----T-----K-E-----	175
CONSENSUS-B	-----I-----	235
SP2	-----I-V-----	235
L47	-----IL-----	247
HXB2R	-----IL-----	235
NL43	-----G-----IL-----	235
NH	-----IT-G-----L-----I-----	230
JRC5F	-----I-D-----V-----	239
JRFL	-----IL-----	235
P895	-----E-ID-----K-----	235
VE1RT	-----T-----K-----E-----	80
VE2RT	-----L-----	80
VE3RT	-----Y-----L-----	80
VE4RT	-----	80
VE5RT	-----	80
VE6RT	-----	80
VE7RT	-----	80
VE8RT	-----L-----	80
RTPK	-----L-----D-----H-D-----	80
POI37B	-----L-----	89
92BR017	-----ILV-----L-----I-----I-----	174
92BR018	-----IL-----L-----I-----	175
92BR020	-----ILV-----I-----E-----	175
L.L0	-----I-----	179
92BR004	-----IL-----	174
92BR028	-----I-----Q-----	175
92BR003	-----I-----T-----	175
92BR021	-----L-----R-----R-----I-----	174
92BR024	-----K-NIL-----	175
92BR030	-----IL-----I-----	175
92BR026	-----E-IRV-----I-----	175
BP_93BR019	-----K-I-----V-----NG-----N-D-----	175
AP6.4	-----L-----D-----	80

[illegible]

II-A-24
DEC 97

HIV1 POL

	polymerase motif										
	- -										
WEAU160	NKRTQDFWEVQLGIPHSQKK	KKSVTVLVGDAYTSVPLDEDFKTPAFFIPSINNETPGIAYQYNVLPQGWKGFPAIFQSSNTHILEFFAKQHPDIVIYQYMDLYVGSCLLEIGQHRTRIEELRML									364
CONSENSUS-C	-A-	-Y-	-G-	-S-	-T-	-A-	-E-	-I-	-A-	-E-	339
92BR025	-S-	-A-	-G-	-S-	-T-	-A-	-E-	-I-	-A-	-E-	362
IND9	-A-	-Y-	-R-	-C-	-DR-	-EL-	-A-	-E-	-A-	-E-	216
IND7	-A-	-Y-	-C-	-DR-	-EL-	-A-	-E-	-A-	-E-	-A-	216
IND6	-A-	-Y-	-DR-	-EL-	-A-	-E-	-A-	-E-	-A-	-E-	216
IND5	-N-	-Y-	-R-	-R-	-S-	-A-	-DR-	-EL-	-A-	-E-	216
IND4	-A-	-Y-	-R-	-A-	-E-	-A-	-E-	-A-	-E-	-A-	216
ETH2220	-A-	-G-	-P-	-P-	-PQ-	-AP-	-E-	-AP-	-E-	-E-	371
92RW026	-A-	-I-	-S-	-S-	-A-	-EM-	-A-	-E-	-A-	-E-	299
92BR023	-A-	-T-	-A-	-E-	-A-	-E-	-A-	-E-	-A-	-E-	304
CONSENSUS-D	-A-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	357
ELI	-A-	-G-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	363
2226	-A-	-K-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	363
NDK	-A-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	-E-	363
94UG114.1	-A-	-H-	-T-	-E-	-EMI-	-A-	-E-	-I-	-G-	-G-	365
92UG001	-A-	-V-	-T-	-C-	-E-	-D-	-E-	-D-	-E-	-E-	304
92UG018	-A-	-I-	-E-	-EM-	-E-	-E-	-E-	-E-	-E-	-E-	304
92UG021	-A-	-E-	-E-	-EM-	-E-	-E-	-E-	-E-	-E-	-E-	304
92UG024	-A-	-Y-	-T-	-E-	-E-	-I-	-E-	-E-	-E-	-E-	303
92UG046	-A-	-Y-	-T-	-EM-	-E-	-I-	-E-	-E-	-E-	-E-	304
92UG067	-A-	-Y-	-N-	-EM-	-E-	-I-	-G-	-E-	-E-	-E-	304
92UG065	-R-	-Y-	-C-	-I-	-EM-	-E-	-I-	-G-	-E-	-E-	304
92UG070	-A-	-Y-	-K-	-L-	-C-	-EM-	-E-	-K-	-E-	-E-	304
92UG053	-A-	-K-	-Q-	-EM-	-E-	-L-	-G-	-E-	-E-	-E-	304
CONSENSUS-F	-A-	-K-	-G-	-T-	-V-	-Y-	-D-	-AK-	-E-	-E-	364
93BR020.1	-A-	-K-	-S-	-T-	-V-	-Y-	-D-	-AK-	-E-	-E-	364
CONSENSUS-G	-A-	-Y-	-V-	-X-	-R-	-AN-	-EM-	-A-	-E-	-E-	344
AG_SE6165	-A-	-R-	-K-	-S-	-TR-	-EM-	-A-	-E-	-E-	-E-	210
AG_92NG083	-A-	-S-	-KE-	-E-	-EM-	-A-	-E-	-A-	-E-	-A-	363
CONSENSUS-H	-A-	-S-	-KE-	-E-	-EM-	-A-	-E-	-A-	-E-	-A-	363
90CR096	-A-	-S-	-KE-	-E-	-EM-	-A-	-E-	-A-	-E-	-A-	363
CONSENSUS-O	-?	-Q-	-P-	-V-	-?	-D-	-E?	-E?	-?	-?	326
ANT70	-E-	-Q-	-P-	-V-	-?	-D-	-E?	-E?	-?	-?	360
MVP5180	-G-	-Q-	-P-	-V-	-?	-D-	-E?	-E?	-?	-?	360
CONSENSUS-CP2	-?	-?	-?	-?	-?	-?	-?	-?	-?	-?	251
CP2GRB	-A-	-K-	-Y-	-S-	-EK-	-T-	-D-	-K-	-V-	-E-	362
CPZANT	-I-	-Q-	-V-	-C-	-A-	-A-	-DKY-	-AVE-	-N-	-TA-	364

HIV1 POL

MSAU160	LRWGFTTPDKKHQKDEPPFLNMGVELHPDKMTVQPIKLPEKSENTFRDQKLVGKLNMAHQIYAGIKVKQLCKLLRGTKALTEIIPITEAELELAENREI...LRSPVHGYYDPSKDLIAELQKQGGQ	491
CONSENSUS-C	-k-----E-----Q-----D-----P-----R-----A-----D-V-L-----I-----d-----	466
92BR025	-K-----E-----Q-----D-----P-----R-----A-----D-V-L-----I-----N-----	489
IND8	-K-----E-----Q-----D-----P-----R-----A-----D-L-----I-----D-----	343
IND7	-----E-----R-Q-----D-----P-----R-----A-----D-V-L-K-----I-----D-----	343
IND6	-----E-----Q-----D-----P-----R-----R-----D-V-L-K-----I-----D-----	343
IND5	-----E-----Q-----D-----P-----R-----A-----D-V-L-----I-----R-----D-----	343
IND4	-K-----E-----Q-----D-----P-----R-----V-----D-V-L-----I-----D-----	343
ETH2270	-K-----E-----Q-----D-----P-----R-----A-----D-VTL-----I-----ND-----	498
92BR026	-K-----E-----E-----D-----P-----R-----A-----VSL-----	386
92BR023	-K-----E-----Q-----D-----P-----R-----A-----G-V-L-----	392
CONSENSUS-D	-----E-----?-----D-----P-----R-----V-----L-----I-----h-----	483
ELI	-----R-----E-----S-----V-----L-----N-----ER-----V-----L-----I-----H-----	490
2226	-----E-----S-----P-----R-----V-----L-----I-----H-----	490
NDK	-----E-----N-----P-----R-----V-----L-----I-----D-----	490
94UG114.1	-K-----Y-----E-----H-----P-----R-----C-----A-----V-----L-----A-----I-----D-----	492
92UG001	-----E-----P-----R-----R-----A-----V-----L-----K-----	391
92UG038	-K-----E-----Q-----P-----S-----A-----V-----L-----K-----	392
92UG021	-K-----E-----N-----P-----R-----C-----A-----V-----L-----	392
92UG024	-K-----E-----E-----P-----C-----A-----V-----L-----	391
92UG046	-K-----E-----P-----IR-----CI-----A-----V-----L-----	392
92UG067	-K-----E-----P-----R-----C-----A-----V-----L-----	392
92UG065	-K-----E-----P-----V-----R-----C-----A-----V-----L-----	392
92UG070	-K-----E-----I-----P-----R-----CI-----V-----L-----	391
92UG053	-K-----E-----E-----P-----R-----CI-----V-----L-----	392
CONSENSUS-E	-K-----L-----E-----Q-----D-----D-----P-----R-----A-----D-V-L-T-----I-----	491
93BR020.1	-K-----L-----E-----Q-----D-----D-----P-----R-----A-----D-V-L-T-----I-----	491
CONSENSUS-G	-k-----l-----E-----Q-----?-----?-----P-----?-----?-----A-----D-V-L-?-----?-----E-----?-----?-----D-----	462
AG_SB6165	-K-----L-----K-----EX-----	226
AG_92UG083	-K-----L-----E-----Q-----D-----P-----H-----R-----A-----D-V-L-A-----M-----H-----E-----V-----PD-----	490
CONSENSUS-H	-K-----E-----TV-----D-----P-----N-----A-----D-----L-----K-----R-----I-----IR-----	490
90CR056	-K-----E-----TV-----D-----P-----N-----A-----D-----L-----K-----R-----I-----IR-----	490
CONSENSUS-I	YQ-----E-----?-----Q-----?-----?-----V-----?-----Q-----R-----?-----E-----I-----S-----V-----L-----S-----E-----?-----?-----Q-----D-----W-----I-----?-----?-----	441
AWI70	YQ-----E-----S-----Q-----N-----DV-----I-----Q-----R-----RE-----I-----S-----V-----L-----S-----E-----R-----Q-----D-----W-----NI-----GE-----	487
MVP5180	YQ-----E-----Q-----D-----V-----Q-----R-----E-----I-----S-----V-----L-----S-----E-----R-----Q-----D-----W-----SI-----H-----E-----	487
CONSENSUS-CF2	??-----?-----E-----?-----?-----?-----?-----?-----P-----?-----?-----I-----?-----?-----D-----V-----?-----?-----?-----?-----?-----?-----?-----?-----?-----?-----?-----?-----	341
CP2QAB	-K-----E-----Q-----V-----I-----P-----I-----I-----K-----DVV-----L-----P-----VST-----D-----E-----I-----NC-----	489
CPZANT	QV-----LE-----E-----Q-----K-----PDD-----P-----T-----R-----I-----V-----S-----DRVQH-----R-----E-----KQ-----KKK-----QKIE-----Y-----Q-----GLP-----K-----TI-----S-----	494

HIV1 POL

	p56	
	Reverse N / Integrase	
WEA1160	VSLTPTNQKTELQAIHLALQDSGLEVMIVTDSQYALGIIQAQPDQSESELVSGIIEQLIKKKEKYLAWVPANRIGIGNEQVDKLVSIGIRRVLPLOGIDKAQSEKKEKYHSHNRAMASDFNLPPVVAKEI	751
CONSENSUS-A	--S-----H-----S-----R-----?--M--X--E--?--S-----?--?--?-----?-----	691
92UG037	--E-----H-----S-----R-----N--X--E--D--S-----R-----I-----	751
D455	--E-----H-----S-----R-----I--N--K--E--S-----D-----C-----	750
Q23	--E-----H-----S-----K-----I--N--K--E--D--S-----R-----T-----I-----	749
CONSENSUS-B	--S-----H-----S-----R-----N-----A-----	750
SF2	--IA-----K-----A-----N-----	751
LAI	--T-----K-----N-----A-----D-----	763
HXB2R	--T-----Y-----N-----A-----D-----	751
NL43	--P-----K-----G-----A-----	751
HN	--K-----K-----A-----D-----I-----	753
JRCSE	--K-----K-----A-----D-----I-----	755
JRFL	--K-----K-----A-----D-----K-----	751
P096	--A-----K-----K-----A-----T-----	749
VE1RT	--V-----K-----S-----K-----E-----	560
VE2RT	--Y-----K-----N-----A-----R-----	560
VE3RT	--I-----I-----N-----T-----	560
VE4RT	--IA-----P-----K-----I--N-----A-----I-----	560
VE5RT	--P-----K-----N-----A-----	560
VE6RT	--I-----K-----A-----	560
VE7RT	--T-----R-----S-----K-----I-----A-----I-----	560
VE8RT	--Q-----K-----I-----T-----R-----	560
RTPX	--P-----S-----K-----A-----Q-----D-----	573
AP6.4	--T-----E-----A-----K-----I-----A-----	560

	p56	
	Reverse N / Integrase	
WEA1160	VSLTPTNQKTELQAIHLALQDSGLEVMIVTDSQYALGIIQAQPDQSESELVSGIIEQLIKKKEKYLAWVPANRIGIGNEQVDKLVSIGIRRVLPLOGIDKAQSEKKEKYHSHNRAMASDFNLPPVVAKEI	751
CONSENSUS-C	--?--E-----Q-----S-----?--K--?--N-----?--R--S-----?-----?E--?--?--?---	708
92BR025	--I--E-----Q-----S-----K-----M-----R--S-----E-----N-----I-----	748
ETH2220	--E-----Q-----S-----I-----K-----I--N-----S--R--S-----N-----I-----P-----	750
CONSENSUS-D	--P?-----N-----K-----N-----Q-----N-----	739
ELI	--P-----N-----K-----N-----Q-----N-----	750
Z2Z6	--PF-----N-----K-----Q-----N-----	750
MDK	--PF-----N-----K-----Q-----N-----	750
94UG114.1	--N-----N-----K-----V-----S-----N-----I-----N-----E-----	752
CONSENSUS-F	--E-----A-----Q-----S-----K-----N-----S-----A-----A-----N-----I--A-----	751
93BR020.1	--E-----A-----Q-----S-----K-----N-----S-----A-----N-----I--A-----	751
CONSENSUS-G	IT??E-----?--R--Q-----??-----R--?--N-----?-----?--R-----?-----	704
AQ_92NG083	IT--E-----A--H--Q-----RP-----R--Q-----N-----S-----R-----	750
CONSENSUS-H	--E-----Y-----K-----N-----E-----S-----V-----R--N-----V-----I-----	750
90CR056	--E-----Y-----K-----N-----E-----S-----V-----R--N-----V-----I-----	750
CONSENSUS-I	IK-?E-----A--M--L?-----KE?-----?--?--SS--T--?--PI--Q-----E--P--?--T-----KI-----KD--R-----E--Q--D-----?--L--?--G--?-----	676
ANT70	IK-DE-----A--M--L-----KET-----V--SS--T--?--PI--Q-----E--P--?--Q--T-----KI-----KD--R-----E--Q--D-----K--L--E--G-----	747
MVP5180	IK-ES-----A--M--VLI-----KEQ-----V--SS--T--D--PI--Q-----E--T--R--T-----KI-----KD--R-----E--Q--D-----L--G-----I-----	747
CONSENSUS-CPZ	?--??-----QA--?--L?-----?-----?--?--?-----?--?--?-----?--?--?-----?--?--?-----?--?--?-----?--?--?-----	518
CPZGAB	I--EN-----QA--K--LL-----DQQ-----V-----S--H-----N-----E-----I--S-----A-----R-----R-----K-----I-----	749
CPZANT	KH--CK-----QA-----LM--E--TGP-----VL--GT-----P--EE--CK-----R--QI--S-----Q-----Q-----E-----D--D-----SL--DEY-----I-----	754

HIV1 POL

WBAU160	VASCDKCOLKGEANHQVDCSPGIWQLDCTHLEGGKILVAVHVASGYIEAEVI PAETGQETAYFILKLAGRNPFVKTITDNGENFTSTTVKAAACWAGIKQEPGIPVNPQSQGVIESHMKELKKIIGQVR	881
CONSENSUS-A	---?-----?-----?-----?-----V?-----AA---?---N??-----V-----	814
92UG037	-----V-----V-----L-----VV-----AA-----NV-----V-----	881
U455	---N-----V-----V-----AA---V---N-Q-----V-----	880
Q23	-----V-V-----L-----IV-----AA-----N-Q-----V-----	879
CONSENSUS-B	-----L-----L-----L-----V-----V-----N-----	880
SP2	-----L-----L-----L-----V-----V-----N-----	881
LAI	-----V-----L-----L-----V-----V-----	893
EXB2R	-----V-----L-----L-----QA---R-----V-----	881
NL43	-----V-----L-----V-----V-----	881
HN	-----V-----L-----F-----T-----	883
JRCSP	-----L-----T-----V-----	885
JRFL	-----L-----V-----	881
P896	---N-----V-----L-----V-----	879

WBAU160	VASCDKCOLKGEANHQVDCSPGIWQLDCTHLEGGKILVAVHVASGYIEAEVI PAETGQETAYFILKLAGRNPFVKTITDNGENFTSTTVKAAACWAGIKQEPGIPVNPQSQGVIESHMKELKKIIGQVR	881
CONSENSUS-C	---?-----?-----?-----?-----?V-----?---M?-----Q-----V-----	831
92BR025	-----T-----V-----I-N-----Q-----V-----	878
BTH2220	---C-----I---N-----L-----RV-----NA-----Q-----V-----	888
CONSENSUS-D	-----V-----L-----VV-----AA-----V-----	869
ELI	-----V-----L-----VV-----AA-----V-----	880
2226	-----V-----L-----IV-----AA-----V-----	880
NDR	-----V-----L-----VV-----A-----V-----	880
94UG114.1	---V---L-----G-----L-----VV-----AA-----V-----	882
CONSENSUS-F	-----L-----L-----T-----A-----Q-----V-----I-----	881
93BR020.1	-----L-----L-----T-----A-----Q-----V-----I-----	881
CONSENSUS-G	-----?---I-----?-----V-----?---AA?-----N-?-----V-----?	877
AQ_92NG083	-----I-----V-----P-I-AA-----N-T-----V-----G	880
CONSENSUS-H	-----QV-----R-----L-----S-----V-----AA-----D-Q-----V-----	880
90CR056	-----QV-----K-----L-----S-----V-----AA-----D-Q-----V-----	880
CONSENSUS-O	I---P---HI---?---E?---?---?---I-----?P-----L---A---V---P---??M-----??-QH-----V-A---S---Q---	796
ANT70	I---P---HI---I---EY---I---N---I---F-----L---A---V---P---M-----N-QH-----V-A---S---Q---	877
MVPS180	I---P---HI---T---Y---E---M---I---DF-----L---A---V---P---AAH-----T---QH-----V-A---S---Q---	877
CONSENSUS-CPZ	?-?-?-?V---?-?-?-V-----V-I---?-?-?-??-?-?-?-L---?-?-?-P---AA-----?-?-?-V---?-?-?-?-?	629
CPZQAB	---H---V-----V-----V-I---L-----L-----P---AA-----D-----V---L-----	879
CPZANT	I-Q---HV---R-----V-----V-I---S---F---MAD---RS---L---S-----A---AA-----N-Q-----V---Q---Q---I---	884

vif cds ->

HEAD160	DQAEHLKFAVQMAVFNNFKKGGIGGYSAGERIIDIIATDIQTQLOKQITKIQNFRVYVYRDSRDPLNKGPAKLLNKGEGAVVIQDNSDIKVVPRKAKIIRDYKQKMGDDCVASRQDED	1003
CONSENSUS-A	E-----?--?--KE-----?--?--?--?--I-----?--?--?--G-----	928
920Q037	E-----S-L-KE-----K--C-----I-----V--K-----G-----	1003
U455	E-----KE-----S-----I-----H--Q-----	1002
Q23	E-----KE-----H-----NYSQ-VK-P-----R-----G-----	1001
CONSENSUS-B	-----V-----KE-----N-----	1002
SP2	-----V-----KE-----NK-----	1003
LAI	-----V-----KE-----	1015
HXB2R	-----V-----KE-----N-----	1003
NL43	-----V-----KE-----V-----	1003
NN	-----R-----VG-----KE-----N-----V-----T-----	1005
JRCSE	-----KE-----N--I-----V-----	1007
JRFL	-----V-----KE-----E-----V-----	1003
P896	-----V--S--KE-----	1001

vif cds ->

HEAD160	DQAEHLKFAVQMAVFNNFKKGGIGGYSAGERIIDIIATDIQTQLOKQITKIQNFRVYVYRDSRDPLNKGPAKLLNKGEGAVVIQDNSDIKVVPRKAKIIRDYKQKMGDDCVASRQDED	1003
CONSENSUS-C	?-----?--?--?--?--KE-----?--?--?--?--I-----?--?--?--A--?--?--	941
92BR025	-----L--I-----KE-----N-----I-----L-----V--K-----A--N-----	959
ETH2220	E-----R-----S-----KE--N--L-----I-----A--G-----	1010
CONSENSUS-D	-----KE-----I-----I-----V-----	991
ELI	-----RR-----KE-----I-----I-----K-----V-----	1002
2226	-----KE-----I-----I-----V-----	1002
NOK	-----KE-----I-----I-----V-----	1002
94UG114.1	E-----KE-----I-----V-----E-----	1004
CONSENSUS-F	-----T-----RE-----I-----V-----E-----G-----	1003
93BR020.1	-----T-----RE-----I-----V-----E-----G-----	1003
CONSENSUS-G	-----S-----KE-----I-----I-----?E-----?--?K-----G--G-----	946
AG_92NG083	-----S-----KE-----I-----I-----NS-----LK-----G--G-----	1002
CONSENSUS-H	-----KE-----SN--K-----I-----E-----E-----	1002
90CR056	-----KE-----SN--K-----I-----E-----E-----	1002
CONSENSUS-O	-----?--V-----T--?--L--SQ--TE--L--??--I--Q-----KG-----?--T--SN--?--T--SES?EQP?EIP	919
ANT70	-----R-----V-----T-----L--SQ--TE--L--NK-----I--Q-----KG-----E-----T--SN--G--T--SESVEQPSEIP	1009
MVPS180	-----V-----T-----L--SQ--TE--L--N-----I--Q-----KG-----T--SN--N--T--SESEMQPSEIP	1009
CONSENSUS-CPZ	-----?--?--?--T--?--?--?--?--?--L--?--?--?--?--?--?--QGEI-----N--	735
CPZGAB	-----T-----SE-----L--V--K-----I--T-----QGEI-----N--	1001
CPZANT	-----Q--V--H-----TP--Q--L--L-----T--N--L--Q--N-----V--Q-----X-----	970

HIV1 VIF

Protein Alignment Summary Table (VIF)

Name	Accession No.	Region	First Author	Reference
CONSENSUS-A				
U455	M62320	complete genome	Oram, J.D.	ARHR 6 (9), 1073-1078 (1990)
IBNG	L39106	complete genome	Howard, T.M.	ARHR 12(15), 1413-1425 (1996)
IFA86	Z30637	vif	Wieland, U.	Virology 203 (1), 43-51 (1994)
92UG037	U51190	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
Q23	AF004885	complete genome	Poss, M.,R.A.	Unpublished
CONSENSUS-B				
SF2	K02007	complete genome	Sanchez-Pescador, R.	(in) Weiss, R.L., Teich, N., Varmus, H. and Coffin, J. (Eds.), <i>RNA Tumor Viruses</i> , Second Edition, 2, Vol. 2:1124-1141, Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
LAI	K02013	complete genome	Wain-Hobson, S.	Cell 40, 9-17 (1985)
HXB2R				
NL43	M19921	complete genome	Adachi, A.	J. Virol. 59, 284-291 (1986)
MN	M17449	complete genome	Gurgo, C.	Virology 164, 531-536 (1988)
JRC5F	M38429	complete genome	Koyanagi, S.	Science 236, 819 (1987)
JRFL	U63632	complete genome	O'Brien, W.A.	Nature 348, 69 (1990)
IFA19	Z30686	vif	Wieland, U.	Virology 203 (1), 43-51 (1994)
IFA65	Z30630	vif	Wieland, U.	Virology 203 (1), 43-51 (1994)
P896	U39362	complete genome	Collman, R.	J. Virol. 66, 7517 (1992)
WEAU160	U21135	complete genome	Ghosh, S.K.	Unpublished (1995)
PA-3799	U24443	vif, vpu, tat, rev, env	Michael, N.L.	J. Virol. 69 (7), 4228-4236 (1995)
U39-32	X91365	vif	Wieland, U.	Unpublished (1995)
3.7	U42234	vif complete	Sova, P.	J. Virol. 69 (4), 2557-2564 (1995)
8.18	U42258	vif complete	Sova, P.	J. Virol. 69 (4), 2557-2564 (1995)
GMK1	U42270	source vif complete	Sova, P.	J. Virol. 69 (4), 2557-2564 (1995)
I2	U42278	source vif complete	Sova, P.	J. Virol. 69 (4), 2557-2564 (1995)
U13-2	X91354	vif	Wieland, U.	Unpublished (1995)
CONSENSUS-C				
92BR025	U52953	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1667 (1996)

ETH2220	U46016	complete genome	Salminen, M.O.	ARHR 12(14), 1329-1339 (1996)
CONSENSUS-D				
ELI	K03454	complete genome	Alizon, M.	Cell 46, 63-74 (1986)
Z2Z6	M22639	complete genome	Theodore, T.	Unpublished (1988)
Z6	K03458	pol, vif, vpr, tat, rev, vpu, env, nef, 3' LTR	Srinivasan, A.	Gene 52, 71-82 (1987)
U12-0	X91351	vif	Wieland, U.	Unpublished (1995)
NDK	M27323	complete genome	Spire, B.	Gene 81, 275-284 (1989)
94UG114.1	U88824	complete genome	Gao, F.	Unpublished
84ZR085	U88822	complete genome	Gao, F.	Unpublished
CONSENSUS-E				
90CF402	U51188	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CM240	U54771	complete genome	Carr, J.K.	J. Virol. 70 (9), 5935-5943 (1996)
93TH253	U51189	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CONSENSUS-F				
93BR020	AF005494	complete genome	Gao, F.	Unpublished
CONSENSUS-G				
92NG003.1	U88825	complete genome	Gao, F.	Unpublished (1997)
92NG083.2	U88826	complete genome	Gao, F.	Unpublished (1997)
CONSENSUS-H				
90CR056	AF005496	complete genome	Gao, F.	Unpublished
CONSENSUS-O				
MVP5180	L20571	complete genome	Gurtler, L.	J. Virol. 68, 1581-1585 (1994)
ANT70C	L20587	complete genome	Vanden Haesevelde, M.	J. Virol. 68 (3), 1586-1596 (1994)
CONSENSUS-CPZ				
STVCPZGAB	X52154	complete genome	Huet, T.	Nature 345 (6273), 356-359 (1990)
STVCPZANT	U42720	gag, pol, vif, vpr, tat, rev, vpu, partial env	Vanden Haesevelde, C.	Virology 221, 346 (1996)

HIV1 VIF

	<- pol cds										hydrophilic region										
	MENRNQVNIWVQVDRMRIRTFWESLYKHHMYSIKKAKGNSVMMHYESTHERISEVNIPLGE..QKLVITTYWGLHTGERDWEQQGVSTEWKQRYSTOVDPDLADQLIKLYHFDCESESIRMAILGNL										- -										
HEA0160	MENRNQVNIWVQVDRMRIRTFWESLYKHHMYSIKKAKGNSVMMHYESTHERISEVNIPLGE..QKLVITTYWGLHTGERDWEQQGVSTEWKQRYSTOVDPDLADQLIKLYHFDCESESIRMAILGNL																				128
CONSENSUS-A	-----M-----V-----P-----E-----K-----d-----AR-VR-----h-----qk-----hY-----d-----k-----al																				128
U485	-----K-----N-----V-----Q-----F-----R-----S-----V-----d-----AR-VR-----K-----H-----LK-----H-----HY-----R-----QI																				128
1FA86	-----N-----V-----P-----F-----CR-----KA-----d-----AR-VR-----D-----K-----H-----LK-----H-----HY-----D-----K-----EV																				128
92UG037	-----N-----RR-----P-----R-----KV-----I-----D-----ARI-VR-----Q-----K-----H-----LK-----I-----HY-----N-----D-----K-----QV																				128
Q23	-----A-----N-----HV-----R-----F-----R-----KV-----d-----AT-VR-----K-----H-----LK-----IT-----MHY-----D-----K-----V-----QV																				128
CONSENSUS-B	-----?-----E-----d-----AY-----k-----k-----y-----i																				127
SP2	-----P-----V-----D-----A-----E-----B-----A-----EK-----G-----HY-----K-----YR																				128
LA1	-----V-----G-----R-----F-----P-----d-----AR-----K-----E-----Y-----D-----K-----L-----I																				128
KXB2R	-----V-----G-----R-----F-----P-----d-----AR-----K-----E-----Y-----D-----K-----L-----I																				128
HL43	-----N-----R-----R-----D-----F-----M-----K-----d-----A-----K-----E-----Y-----D-----K-----L-----I																				128
HN	-----R-----A-----R-----F-----d-----AR-----K-----E-----Y-----D-----K-----L-----I																				128
JRC37	-----N-----G-----I-----K-----N-----V-----Q-----d-----AR-----K-----E-----Y-----D-----K-----L-----I																				128
JRFL	-----T-----G-----I-----d-----AR-----K-----E-----Y-----D-----K-----L-----I																				128
1FA19	-----G-----F-----N-----T-----d-----A-----K-----E-----Y-----D-----K-----L-----I																				128
1FA65	-----N-----R-----V-----I-----K-----d-----A-----Q-----K-----T-----I-----Y-----D-----K-----L-----I																				128
P896	-----G-----N-----d-----A-----V-----K-----Q-----R-----Y-----D-----K-----L-----I																				128
PA-3799	-----G-----F-----Y-----d-----A-----K-----E-----Y-----D-----K-----L-----I																				128
U39-32	-----N-----K-----F-----R-----V-----d-----AR-VR-----K-----Q-----H-----LK-----I-----HY-----D-----K-----L-----QV																				128
3.7	-----I-----G-----R-----F-----P-----d-----AR-VR-----K-----Q-----H-----LK-----I-----HY-----D-----K-----L-----QV																				127
6.18	-----I-----R-----T-----C-----N-----KT-----d-----AR-IV-----N-----H-----HY-----D-----K-----L-----LR																				128
GNK1	-----F-----R-----V-----d-----AR-----K-----E-----Y-----D-----K-----L-----I																				128
I2	-----A-----P-----D-----K-----d-----AR-----K-----E-----Y-----D-----K-----L-----I																				128
U13-2	-----N-----F-----R-----V-----d-----ARI-----K-----K-----Q-----H-----LR-----I-----HY-----N-----D-----K-----L-----QV																				128
CONSENSUS-C	-----L-----K-----N-----??-RR-?-?-?-R-K?-d-----AR-?-?-Q-----H-----LR-?-?-G-----?-MHY-----A-----K-----L-----QV																				114
92BR025	-----L-----K-----N-----V-----RR-----S-----Y-----R-----K-----d-----AR-----I-----Q-----H-----LR-----R-----G-----MHY-----AD-----K-----L-----QV																				128
ETH2220	-----L-----K-----N-----H-----RR-----N-----V-----D-----R-----KV-----d-----AR-I-K-----Q-----H-----LR-----N-----G-----MHY-----A-----K-----L-----QV																				128
CONSENSUS-D	-----V-----P-----F-----D-----P-----K-----d-----AR-VK-----E-----R-----G-----MHY-----A-----K-----L-----QV																				125
BLICQ	-----K-----V-----NR-----F-----P-----K-----d-----AR-----K-----E-----R-----G-----MHY-----A-----K-----L-----QV																				128
Z226	-----V-----SR-----F-----D-----P-----N-----d-----AM-VK-----R-----G-----MHY-----A-----K-----L-----QV																				128
Z6	-----V-----SR-----F-----D-----P-----K-----d-----AR-VK-----R-----G-----MHY-----A-----K-----L-----QV																				128
U120	-----KA-----I-----V-----RR-----Q-----L-----D-----L-----K-----d-----A-----VK-----N-----E-----MHY-----A-----K-----L-----QV																				128
NEK	-----N-----Y-----V-----NR-----F-----D-----H-----K-----d-----AR-VK-----E-----R-----G-----MHY-----A-----K-----L-----QV																				128
94UG114	-----L-----DCPM-----K-----d-----AR-VK-----E-----R-----G-----MHY-----A-----K-----L-----QV																				128
842R065	-----N-----Y-----N-----F-----D-----P-----K-----d-----AR-VK-----E-----R-----G-----MHY-----A-----K-----L-----QV																				128
CONSENSUS-E	-----?-N-----?-??-F-----Q-----KV-----d-----AR-?-R-----?-K-----?-H-----QK-----I-----QY-----D-----P-----?-?-QV																				115
CM240	-----N-----K-----F-----Q-----KV-----d-----AR-?-R-----?-K-----?-H-----QK-----I-----QY-----D-----P-----?-?-QV																				128
90CP402	-----A-----N-----S-----AK-----F-----Q-----KV-----d-----AR-I-R-----K-----H-----QK-----I-----QY-----D-----P-----?-?-QV																				127
93TH253	-----N-----Q-----F-----Q-----KV-----d-----AR-?-R-----?-K-----?-H-----QK-----I-----QY-----D-----P-----?-?-QV																				128
CONSENSUS-F	-----N-----Y-----N-----F-----F-----R-----K-----ET-----AE-----LP-----E-----QY-----R-----I-----G-----I-----Y-----K-----K																				128
93BR020	-----N-----Y-----N-----F-----F-----R-----K-----ET-----AE-----LP-----E-----QY-----R-----I-----G-----I-----Y-----K-----K																				128
CONSENSUS-H	-----N-----Y-----H-----R-----R-----F-----F-----d-----AR-----N-----E-----LK-----E-----G-----MHY-----K-----RV																				128
90CR056	-----N-----Y-----H-----R-----R-----F-----F-----d-----AR-----N-----E-----LK-----E-----G-----MHY-----K-----RV																				128
CONSENSUS-O	-----L-----?-QKVK-----N-----Y-----K-----?-?-?-N-----RN-----TV-----?-Y-----V-----d-----A-----V-----N-----E-----MHY-----K-----RV																				111
HVP5180	-----L-----I-----QKVK-----N-----Y-----K-----N-----AN-----R-----RN-----KV-----A-----Y-----VA-----ADI-----V-----N-----E-----MHY-----K-----RV																				128
ANT70	-----L-----QKVK-----N-----Y-----K-----R-----TEN-----W-----RN-----V-----S-----Y-----V-----d-----A-----V-----N-----E-----MHY-----K-----RV																				128
CONSENSUS-CPZ	-----??-?-?-?-I-----V-----N-----F-----F-----R-----K-----ET-----AE-----LP-----E-----QY-----R-----I-----G-----I-----Y-----K-----K																				73
CPZGAB	-----K-----N-----Y-----I-----R-----R-----F-----DHPN-----KVA-----I-----PRDY-----S-----IV-----A-----SP-----A-----H-----Q-----LGS-----V-----FT-----R-----SQY-----A-----T-----R-----Q																				129
CPZANT	-TASVG-IA-----NI-----IWET-VL-P-A-----ND-----KKGE-----PTLDK-----V-VF-----QC-----P-----H-----CGK-----I-----ET-----N-----Q-----Y-----P-----DQ-----V-----Q-----ER																				130

	vpr cds ->	
WEAU160	VIPRCEYQAGHMKVGSLOYLAL.TAL...ITPKIKPPLPSVYKLTEDRMKPKQTKGHRGSHYNGH	192
CONSENSUS-A	-R-----K-----V-a-t-a-----e-----R-----R?-----	191
U455	-R-----K-----V-TRA-----R-----R-L-R-----	192
IFAR6	-R-----K-----VA--A-----L-----E-----R-----NRS-----	192
92UG037	-S--D--T-----K-----V-SRM-----A-----R-R-E-----C-----	192
Q23	-S-K-----K-----V--T-----RI-----R-L-E-----C-----	192
CONSENSUS-B	-S-----S-----?-----	190
SP2	-S-----A-----P-----	192
LAI	-S-----A-----T-----	192
NXB2R	-S-----A-----T-----	192
NI43	-S-----A-----K-Q-----R-----	192
MM	-S-I--F-----B-----Y-----	192
JRC6F	-S-----S-----K-----	192
JRF1	-S-----K-----	192
IFA19	-S-----R-----V-----	192
IPA6B	--SS--K-----A-----K--KR-----S-----	192
P896	-S-S-----A-----T--RR--P--T-----T-----	192
PA-3799	-S-----A-----R-----	192
Q39-32	-S-S--RT-----K-----V-T-----R-----L--R-P-----C-----	192
3.7	-S-----V-----A-A-----	191
B.10	-S-----S-----S-----A-----A-----B-----H-----	181
GMR1	-S-----A-----V--T-----E-----	192
I2	-S-----R-----A-----TKS--T-----T-----	192
U13-2	-S-----T-----K-----V--TAT-----R-----S-----R--F-----S-C-----	192
CONSENSUS-C	-ST--D-----K-----?-----?--V--?-----R?R--N-----	173
92BR025	-SS--D-----K-----V-----RDR--N-----	192
ETH2220	-S--D-----K--A-----S-V--K-----R--R--N-----	192
CONSENSUS-D	-ST-----?-----A-----R-----?-----	186
ELICG	-S-----A-Q-----R-----Q-R-----	192
2226	-SH-----S-----VA-----R-----K-----	192
26	-SH-----S-----A-----R-----K-----	188
U120	-S-S-D-----K-----R-----QW--KQ--LT-----	190
NDK	-S-S-----A-----A-----R-----R-----	192
94UG114	-V--N-----V--R-----G-----R-----	192
84XR085	-S-----A--R-----RQ--R-----	192
CONSENSUS-E	-R?----F?-----K-----T??-??-?-----?-----ETP-----	169
CM240	-RR----PS-----K-----Y--R-R-----XI-----R--ENP-----	190
90CF402	-R-----P-----K-----TKT--T-----R-----E--P-----	191
93TH251	-RR----PS-----K-----T--R-R-----I-----ENP-----	192
CONSENSUS-F	IS--N-----A--F-----Q--V-----R--E-----	192
93BR020	IS--N-----A--T-----Q--V-----R--E-----	192
CONSENSUS-H	-R--N-P--KQ--T-----VA-----R--V-----R-----	192
90CR056	-R--N-P--KQ--T-----VA-----R--V-----R-----	192
CONSENSUS-O	-LTR--??--SQ--T--?--?--V--VK???--Q?-----??PIRDQL?--?S-----	161
MVP5180	-LTK--L--SQ--T--F--R-V--VKV--RN--QR-----W--IRDQL--E-----	192
ANT70	-LTK--PT--SQ--T--L--R-V--VKARS--Q-----HLRIKDQLK--PS-----	192
CONSENSUS-CPZ	????-?K?--?Q--?--?--??K?--??????R??--??--??R??--?EN?--R--	107
CFZGAB	-A-----KE--RQ--F--K-----SERPER--A-----E--R--V--QENL--R--	193
CPZANT	ILAT-H-KK--SQ--T-----PCKI-EFRGY--GPRRQF--LSI-----RRMR--ENQ-----	193

HIV1 VPR

Protein Alignment Summary Table (Vpr)

Name	Accession No.	Region	First Author	Reference
CONSENSUS-A				
U455	M62320	complete genome	Oram, J.D.	ARHR 6 (9), 1073-1078 (1990)
Q23	AF004885	complete genome	Poss, M.,R.A.	Unpublished
IBNG	L39106	complete genome	Howard, T.M.	ARHR 12(15), 1413-1425 (1996)
92UG037	U51190	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CONSENSUS-B				
SF2	K02007	complete genome	Sanchez-Pescador, R.	(in) Weiss, R.L., Teich, N., Varmus, H. and Coffin, J. (Eds.), <i>RNA Tumor Viruses</i> , Second Edition, 2, Vol. 2:1124-1141, Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
LAI	K02013	complete genome	Wain-Hobson, S.	Cell 40, 9-17 (1985)
HXB2	K03455	complete genome	Rosen, C.A.	Cell 41, 813-823 (1985)
NL43	M19921	complete genome	Adachi, A.	J. Virol. 59, 284-291 (1986)
MN	M17449	complete genome	Gurgo, C.	Virology 164, 531-536 (1988)
JRC5F	M38429	complete genome	Koyanagi, S.	Science 236, 819 (1987)
JRFL	U63632	complete genome	O'Brien, W.A.	Nature 348, 69 (1990)
P896	U39362	complete genome	Collman, R.	J. Virol. 66, 7517 (1992)
3202A12	U34603	complete genome	Guillon, C.	ARHR 11, 1537 (1995)
3202A21	U34604	complete genome	Guillon, C.	ARHR 11, 1537 (1995)
WEAU160	U21135	complete genome	Ghosh, S.K.	Unpublished (1995)
1883	U41704	vpr	Ge, Y.C.	ARHR 12 (4), 351-354 (1996)
1188-4	U41705	vpr	Ge, Y.C.	ARHR 12 (4), 351-354 (1996)
1062-1-vpr	U73371	vpr	Wang, B.	ARHR 13(3): 275-279 (1996)
VR-1	U41722	vpr	Ge, Y.C.	ARHR 12 (4), 351-354 (1996)
F12CG	Z11530	complete genome	Carlini, F.	J. Viral Diseases 1, 40-55 (1992)
WR27	U26546	complete genome	Salminen, M.O.	Virology 213 (1), 80-86 (1995)
CONSENSUS-C				
92BR025	U52953	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1667 (1996)
ETH2220	U46016	complete genome	Salminen, M.O.	ARHR 12(14), 1329-1339 (1996)
CONSENSUS-D				

ELI	K03454	complete genome	Alizon, M.	Cell 46, 63-74 (1986)
Z2Z6	M22639	complete genome	Theodore, T.	Unpublished (1988)
Z6	K03458	pol, vif, vpr, tat, rev, vpu, env, nef, 3' LTR	Srinivasan, A.	Gene 52, 71-82 (1987)
NDK	M27323	complete genome	Spire, B.	Gene 81, 275-284 (1989)
	U88822	complete genome	Gao, F.	Unpublished
94UG1141	U88824	complete genome	Gao, F.	Unpublished
CONSENSUS-E				
CM240	U54771	complete genome	Carr, J.K.	J. Virol. 70 (9), 5935-5943 (1996)
90CF402	U51188	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
93TH253	U51189	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CONSENSUS-F				
93BR020.1	AF005494	complete genome	Gao, F.	Unpublished
CONSENSUS-G				
92NG003	U88825	complete genome	Gao, F.	Unpublished (1997)
H92NG083	U88826	complete genome	Gao, F.	Unpublished (1997)
CONSENSUS-H				
U40561	AF005496	complete genome	Gao, F.	Unpublished
CONSENSUS-O				
ANI70C	L20587	complete genome	Vanden Haesevelde, M.	J. Virol. 68 (3), 1586-1596 (1994)
MVP5180	L20571	complete genome	Gurtler, L.	J. Virol. 68, 1581-1585 (1994)
CONSENSUS-CPZ				
CPZGAB	X52154	complete genome	Huet, T.	Nature 345 (6273), 356-359 (1990)
CPZANT	U42720	gag, pol, vif, vpr, tat, rev, vpu, partial env	Vanden Haesevelde.C.	Virology 221, 346 (1996)

Protein Alignment Summary Table (Tat)

Name	Accession No.	Region	First Author	Reference
CONSENSUS-A				
U455	M62320	complete genome	Oram, J.D.	ARHR 6 (9), 1073-1078 (1990)
UG273A	L22957	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
UG275A	L22951	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
DJ263A	L22941	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
DJ264A	L23064	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
SP1702	M66535	tat, rev	Evans, L.	PNAS 85, 2815 (1988)
92UG037	U51190	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
Q23	AF004885	complete genome	Poss, M.	Unpublished
CONSENSUS-B				
SF2	K02007	complete genome	Sanchez-Pescador, R.	(in) Weiss, R.L., Teich, N., Varmus, H. and Coffin, J. (Eds.), <i>RNA Tumor Viruses</i> , Second Edition, 2, Vol. 2:1124-1141, Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
HXB2	K03455 M38432	complete genome	Rosen, C.A.	Cell 41, 813-823 (1985)
MN	M17449	complete genome	Gurgo, C.	Virology 164, 531-536 (1988)
JRC5F	M38429	complete genome	Koyanagi, S.	Science 236, 819 (1987)
JRFL	U63632	complete genome	O'Brien, W.A.	Nature 348, 69 (1990)
P896	U39362	complete genome	Collman, R.	J. Virol. 66, 7517 (1992)
tat	X64650	tat	Siderovski, D.P.	Nucleic Acids Res. 20 (20), 5311-5320 (1992)
R1-18	U11179	tat (partial)	Diaz, R.S.	J. Virol. 69, 3273-3281 (1995)
SWB882	M31462	tat	Meyerhans, A.	Cell 58, 901-910 (1989)
R250	U11204	tat (partial)	Diaz, R.S.	J. Virol. 69, 3273-3281 (1995)
WEAU160	U21135	complete genome	Ghosh, S.K.	Unpublished (1995)
CONSENSUS-C				
DJ259A	L22940	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
DJ373A	L23065	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
ETH2220	U46016	complete genome	Salminen, M.O.	ARHR 12(14), 1329-1339 (1996)
92BR025	U52953	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1667 (1996)

HIV1 TAT

92RW009	U88823	complete genome	Gao, F.	Unpublished;
CONSENSUS-D				
ELI	K03454	complete genome	Alizon, M.	Cell 46, 63-74 (1986)
Z2Z6	M22639	complete genome	Theodore, T.	Unpublished (1988)
NDK	M27323	complete genome	Spire, B.	Gene 81, 275-284 (1989)
SE365A2	L22945	env tat nef vpu rev	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
UG274A2	L22950	env tat nef vpu rev	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
84ZR085	U88822	complete genome	Gao, F.	Unpublished
94UG1141	U88824	complete genome	Gao, F.	Unpublished
CONSENSUS-E				
90CF402	U51188	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CM240	U54771	complete genome	Carr, J.K.	J. Virol. 70 (9), 5935-5943 (1996)
93TH253	U51189	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CONSENSUS-F				
BZ126A	L22082	env tat rev nef vpu	Louwagie, J.	ARHR 10 (5), 561-567 (1994)
BZ163A	L22085	env tat rev nef vpu	Louwagie, J.	ARHR 10, 561 (1994)
93BR0201	AF005494	complete genome	Gao, F.	Unpublished
U40561	AF005496	complete genome	Gao, F.	Unpublished
CONSENSUS-G				
92NG083.2	U88826	complete genome	Gao, F.	Unpublished (1997)
92NG003	U88825	complete genome	Gao, F.	Unpublished (1997)
CONSENSUS-O				
ANT70C	L20587	complete genome	Vanden Haesevelde, M.	J. Virol. 68 (3), 1586-1596 (1994)
MVP5180	L20571	complete genome	Gurtler, L.	J. Virol. 68, 1581-1585 (1994)
CONSENSUS-CPZ				
CPZGAB	X52154	complete genome	Huet, T.	Nature 345 (6273), 356-359 (1990)
CPZANT	U42720	gag, pol, vif, vpr, tat, rev, vpu, partial env	Vanden Haesevelde, C.	Virology 221, 346 (1996)

II-A-41
DEC 97

HIV1 REV

Protein Alignment Summary Table (Rev)

Name	Accession No.	Region	First Author	Reference
CONSENSUS-A				
U455	M62320	complete genome	Oram, J.D.	ARHR 6 (9), 1073-1078 (1990)
UG273A	L22957	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
SF1702	M66535	tat, rev	Evans, L.	PNAS 85, 2815 (1988)
Q23	AF004885	complete genome	Poss, M.	Unpublished
92UG037	U51190	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CONSENSUS-B				
SP2	K02007	complete genome	Sanchez-Pescador, R.	(in) Weiss, R.L., Teich, N., Varmus, H. and Coffin, J. (Eds.), <i>RNA Tumor Viruses</i> , Second Edition, 2, Vol. 2:1124-1141, Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
LAI	K02013	complete genome	Wain-Hobson, S.	Cell 40, 9-17 (1985)
MN	M17449	complete genome	Gurgo, C.	Virology 164, 531-536 (1988)
JRCSF	M38429	complete genome	Koyanagi, S.	Science 236, 819 (1987)
JRFL	U63632	complete genome	O'Brien, W.A.	Nature 348, 69 (1990)
WEAU160	U21135	complete genome	Ghosh, S.K.	Unpublished (1995)
92BR020.4	U08797	env, partial nef	Gao, F.	ARHR 10 (11), 1327-1343 (1994)
92TH014.12	U08801	env, partial nef	Gao, F.	ARHR 10 (11), 1327-1343 (1994)
92UG005	U08803	env	WHO-GPA	ARHR 10 (11), 1327-1343 (1994)
92US657.1	U04908	tat, rev, env, and partial vpr, nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
91HT652.11	U08443	env, partial nef cds	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
9001.A2	U44069	rev	Wade, N.	AIDS Res. Hum. Retroviruses 12 (9), 845-847 (1996)
CONSENSUS-C				
93MW959.18	U08453	env, partial nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
93MW960.3	U08454	env, partial nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
93MW965.26	U08455	env, partial nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
92BR025	U52953	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
ETH2220	U46016	complete genome	Salminen, M.O.	ARHR 12(14), 1329-1339 (1996)
CONSENSUS-D				

ELI	K03454	complete genome	Alizon, M.	Cell 46, 63-74 (1986)
Z2Z6	M22639	complete genome	Theodore, T.	Unpublished (1988)
Z6	K03458	pol, vif, vpr, tat, rev, vpu, env, nef, 3' LTR	Srinivasan, A.	Gene 52, 71-82 (1987)
NDK	M27323	complete genome	Spire, B.	Gene 81, 275-284 (1989)
92UG024D	U08805	env	WHO-GPA	ARHR 10 (11), 1327-1343 (1994)
84ZR085	U88822	complete genome	Gao, F.	Unpublished
94UG1141	U88824	complete genome	Gao, F.	Unpublished
CONSENSUS-E				
90CF402	U51188	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CM240	U54771	complete genome	Carr, J.K.	J. Virol. 70 (9), 5935-5943 (1996)
CONSENSUS-F				
BZ163A	L22085	env, tat, rev, nef, vpu	Louwagie, J.	ARHR 10, 561 (1994)
BZ126A	L22082	env, tat, rev, nef, vpu	Louwagie, J.	ARHR 10 (3), 561-567 (1994)
93BR0201	AF005494	complete genome	Gao, F.	Unpublished
CONSENSUS-G				
92NG003.1	U88825	complete genome	Gao, F.	Unpublished (1997)
H92NG083.2	U88826	complete genome	Gao, F.	Unpublished (1997)
LBV217	U09664	env	Janssens, W.	ARHR 10, 877-879 (1994)
CONSENSUS-H				
U40561	AF005496	complete genome	Gao, F.	Unpublished
CONSENSUS-O				
ANT70C	L20587	complete genome	Vanden Haesevelde, M.	J. Virol. 68 (3), 1586-1596 (1994)
MVP5180	L20571	complete genome	Gurtler, L.	J. Virol. 68, 1581-1585 (1994)
CONSENSUS-CPZ				
CPZGAB	X52154	complete genome	Huet, T.	Nature 345 (6273), 356-359 (1990)
CPZANT	U42720	gag, pol, vif, vpr, tat, rev, vpu, partial env	Vanden Haesevelde, C.	Virology 221, 346 (1996)

[illegible]

Protein Alignment Summary Table (Vpu)

Name	Accession No.	Region	First Author	Reference
CONSENSUS-A				
U455	M62320	complete genome	Oram, J.D.	ARHR 6 (9), 1073-1078 (1990)
UG273A	L22957	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
UG275A	L22951	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
92RW009	U88823	complete genome	Gao, F.	Unpublished;
92UG037	U51190	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CONSENSUS-B				
SF2	K02007	complete genome	Sanchez-Pescador, R.	(in) Weiss, R.L., Teich, N., Varmus, H. and Coffin, J. (Eds.), <i>RNA Tumor Viruses</i> , Second Edition, 2, Vol. 2:1124-1141, Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
HXB2	K03455 M38432	complete genome	Rosen, C.A.	Cell 41, 813-823 (1985)
LAI	K02013	complete genome	Wain-Hobson, S.	Cell 40, 9-17 (1985)
MN	M17449	complete genome	Gurgo, C.	Virology 164, 531-536 (1988)
NL43	M19921	complete genome	Adachi, A.	J. Virol. 59, 284-291 (1986)
JRC5F	M38429	complete genome	Koyanagi, S.	Science 236, 819 (1987)
JRFL	U63632	complete genome	O'Brien, W.A.	Nature 348, 69 (1990)
P896	U39362	complete genome	Collman, R.	J. Virol. 66, 7517 (1992)
WEAU160	U21135	complete genome	Ghosh, S.K.	Unpublished (1995)
92US657.1	U04908	tat, rev, env and partial vpr, nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
HN0008	Z68505	vpu	Kuiken, C.L.	J. Gen. Virol. 77: 783-792 (1996)
JGV86	Z68602	vpu	Kuiken, C.L.	J. Gen. Virol. 77: 783-792 (1996)
JGV90	Z68607	vpu	Kuiken, C.L.	J. Gen. Virol. 77: 783-792 (1996)
JGV97	Z68614	vpu	Kuiken, C.L.	J. Gen. Virol. 77: 783-792 (1996)
CONSENSUS-C				
ETH2220	U46016	complete genome	Salminen, M.O.	ARHR 12(14), 1329-1339 (1996)
92BR025	U52953	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1667 (1996)
CONSENSUS-D				
Z226	M22639	complete genome	Theodore, T.	Unpublished (1988)

HIV1 VPU

Z6	K03458	pol, vif, vpr, tat, rev, vpu, env, nef, 3' LTR	Srinivasan, A.	Gene 52, 71-82 (1987)
NDK	M27323	complete genome	Spire, B.	Gene 81, 275-284 (1989)
TZ012	U12407	envelope V1-V3 vpu partial	Siwka, W.	ARHR 10 (12), 1753-1754 (1994)
TZ064	U12413	envelope V1-V3 vpu partial	Siwka, W.	ARHR 10 (12), 1753-1754 (1994)
TZ030	U12411	envelope V1-V3 vpu partial	Siwka, W.	ARHR 10 (12), 1753-1754 (1994)
TZ017	U12409	envelope V1-V3 vpu partial	Siwka, W.	ARHR 10 (12), 1753-1754 (1994)
MAL	K03456	complete genome	Alizon, M.	Cell 46, 63-74 (1986)
84ZR085	U88822	complete genome	Gao, F.	Unpublished
94UG1141	U88824	complete genome	Gao, F.	Unpublished
CONSENSUS-E				
CM240X	L14572	V3 region, envelope tat nef rev vpu	Mascola, J.J.	J. Infect. Dis. 169, 48-54 (1993)
93TH253	U51189	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CONSENSUS-F				
BZ126A	L22082	env tat rev nef vpu	Louwagie, J.	ARHR 10 (5), 561-567 (1994)
BZ163A	L22085	env tat rev nef vpu	Louwagie, J.	ARHR 10, 561 (1994)
93BR029.4	AF005495	complete genome	Gao, F.	Unpublished
93BR020.1	AF005494	complete genome	Gao, F.	Unpublished
U40561	AF005496	complete genome	Gao, F.	Unpublished
CONSENSUS-G				
92NG083.2	U88826	complete genome	Gao, F.	Unpublished (1997)
92NG003.1	U88825	complete genome	Gao, F.	Unpublished (1997)
CONSENSUS-O				
ANT70C	L20587	complete genome	Vanden Haesevelde, M.	J. Virol. 68 (3), 1586-1596 (1994)
MVP5180	L20571	complete genome	Gurder, L.	J. Virol. 68, 1581-1585 (1994)
CONSENSUS-CPZ				
CPZGAB	X52154	complete genome	Huet, T.	Nature 345 (6273), 356-359 (1990)
CPZANT	U42720	gag, pol, vif, vpr, tat, rev, vpu, partial env	Vanden Haesevelde.C.	Virology 221, 346 (1996)

HIV1 ENV

Protein Alignment Summary Table (Env)

Name	Accession No.	Region	First Author	Reference
CONSENSUS-A				
U455	M62320	complete genome	Oram, J.D.	ARHR 6 (9), 1073-1078 (1990)
SF1703	M66533	env	Evans, L.	PNAS 85, 2815 (1988)
92RW020.5	U08794	env, nef	Gao, F.	ARHR 10 (11), 1327-1343 (1994)
92UG031.7	L34667	env, partial nef	Gao, F.	ARHR 10 (11), 1359-1368 (1994)
92UG037.8	U09127	env nef	Gao, F.	ARHR 10 (11), 1359-1368 (1994)
UG275A	L22951	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
UG273A	L22957	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
DJ264A	L23064	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
DJ263A	L22941	env, tat, vpU, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
DJ258A	L22939	env, vpU, and nef sequences, tat rev	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
KENYA	L22943	envelope	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
VI191A2	L22952	envelope	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
Q23-17	AF004885	complete genome	Poss, M.	Unpublished
CONSENSUS-B				
SF2	K02007	complete genome	Sanchez-Pescador, R.	(in) Weiss, R.L., Teich, N., Varmus, H. and Coffin, J. (Eds.), <i>RNA Tumor Viruses</i> , Second Edition, 2, Vol. 2:1124-1141, Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
LAI	K02013	complete genome	Wain-Hobson, S.	Cell 40, 9-17 (1985)
HXB2	K03455 M38432	complete genome	Rosen, C.A.	Cell 41, 813-823 (1985)
NL43	M19921	complete genome	Adachi, A.	J. Virol. 59, 284-291 (1986)
JRC5F	M38429	complete genome	Koyanagi, S.	Science 236, 819 (1987)
JRFL	U63632	complete genome	O'Brien, W.A.	Nature 348, 69 (1990)
BAL1	M68893	tat, rev, vpu, env, nef	Reitz, M.	Science 233, 215 (1986)
MN	M17449	complete genome	Gurgo, C.	Virology 164, 531-536 (1988)
SF162	M38428	tat, rev, vpu, env, env, 3' LTR, nef	Cheng-Mayer, C.	J. Virol. 64, 4390-4398 (1990)
91US006.10	U27443	env, nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)

HIV1 ENV

91US005.11	U27434	env, nef	Gao, F.	J. Virol. 70 (3),1651-1657 (1996)
92TH014.12	U08801	env, partial nef	Gao, F.	ARHR 10 (11), 1327-1343 (1994)
92BR020.4	U08797	env, partial nef cds	Gao, F.	ARHR 10 (11), 1327-1343 (1994)
93TH067A	U39259	source envelope complete	Penny, M.A.	ARHR 12 (8), 741-747 (1996)
91HT651.11	U08441	env, partial nef	Gao, F.	J. Virol. 70 (3),1651-1657 (1996)
92HT593.1	U08444	env, nef partial	Gao, F.	J. Virol. 70 (3),1651-1657 (1996)
WEAU160	U21135	complete genome	Ghosh, S.K.	Unpublished (1995)
CONSENSUS-C				
92BR025.9	U09126	env nef	Gao, F.	ARHR 10 (11), 1359-1368 (1994)
ETH2220	U46016	complete genome	Salminen, M.O.	ARHR 12(14), 1329-1339 (1996)
BU910112	U39233	env	Penny, M.A.	ARHR 12 (8), 741-747 (1996)
BR911513	U39234	env	Penny, M.A.	ARHR 12 (8), 741-747 (1996)
93MW959.18	U08453	complete env, partial nef	Gao, F.	J. Virol. 70 (3),1651-1657 (1996)
93MW960.3	U08454	complete env, partial nef	Gao, F.	J. Virol. 70 (3),1651-1657 (1996)
93MW965.26	U08455	complete env, partial nef	Gao, F.	J. Virol. 70 (3),1651-1657 (1996)
UG268A2	L22948	env nef vpu rev, tat	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
SM145A	L22946	env tat nef vpu rev	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
ZAM18A	L22954	envelope	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
ZAM20A	L22956	envelope	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
DJ259A	L22940	env, tat, vpu, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
DJ373A	L23065	env, tat, vpu, rev, and nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
SE364A	L22944	env tat vpu rev nef	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
BU91007A	U39249	source envelope complete	Penny, M.A.	ARHR 12 (8), 741-747 (1996)
BU91002A	U39248	source envelope complete	Penny, M.A.	ARHR 12 (8), 741-747 (1996)
CONSENSUS-D				
84ZR085	U88822	complete genome	Gao, F.	Unpublished
WHO15-726	U36884	envelope complete	Douglas, N.W.	Aids 10, 39-46 (1996)
87TZ4622	U65075	env	Robbins, K.E.	ARHR 12(14), 1389-1391 (1996)
94UG1141	U88824	complete genome	Gao, F.	Unpublished
ELI	K03454	complete genome	Alizon, M.	Cell 46, 63-74 (1986)
Z226	M22639	complete genome	Theodore, T.	Unpublished (1988)
Z6	K03458	pol, vif, vpr, tat, rev, vpu, env, nef, 3' LTR	Srinivasan, A.	Gene 52, 71-82 (1987)

HIV1 ENV

NDK	M27323	complete genome	Spire, B.	Gene 81, 275-284 (1989)
92UG024.2	U43386	env	WHO-GPA, C.	ARHR 10, 1327 (1994)
JY1	J03653	vpu, env, tat (exon 3), rev (exon 3), nef	Youno, J.	ARHR 4, 165-173 (1988)
UG269A	L22949	env, vpu, and nef, tat rev	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
UG274A2	L22950	env tat nef vpu rev	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
SE365A2	L22945	env tat nef vpu rev	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
CONSENSUS-E				
92TH011A	U08811	gp120 of env	WHO-GPA	ARHR 10 (11), 1327-1343 (1994)
KH03	U48264	env	McCutchan, F.E.	J. Virol. 70 (6), 3331-3338 (1996)
KH08	U48266	env	McCutchan, F.E.	J. Virol. 70 (6), 3331-3338 (1996)
93TH966.8	U08456	env, partial nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
93TH975.15	U08457	env, partial nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
TN235	L03698	env	McCutchan, F.E.	ARHR 8 (11), 1887-1895 (1992)
TN242	L03701	env	McCutchan, F.E.	ARHR 8 (11), 1887-1895 (1992)
92TH022.4	U09131 U15120	env	Gao, F.	J. Virol. 70 (3), 1651-1667 (1996)
93TH976.17	U08458	env, partial nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CM240	U54771	complete genome	Carr, J.K.	J. Virol. 70 (9), 5935-5943 (1996)
CONSENSUS-F				
93BR029.4	AF005495	complete genome	Gao, F.	Unpublished
93BR020.1	AF005494	complete genome	Gao, F.	Unpublished;
BZ163A	L22085	env, tat, rev, nef, vpu	Louwagie, J.	ARHR 10, 561 (1994)
BZ126A	L22082	env, tat, rev, nef, vpu	Louwagie, J.	ARHR 10 (5), 561-567 (1994)
CONSENSUS-G				
92RU131.9	U30312	env, nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
VI525A2	L22953	envelope	Louwagie, J.J.	J. Virol. 69 (1), 263-271 (1995)
92NG083.2	U88826	complete genome	Gao, F.	Unpublished (1997)
92NG003.1	U88825	complete genome	Gao, F.	Unpublished (1997)
LBV217	U09664	env	Janssens, W.	ARHR 10, 877-879 (1994)
92UG975.10 JW3	U27426	env gp160 nef	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CONSENSUS-H				
90CR056	AF005496	complete genome	Gao, F.	Unpublished

CONSENSUS-O

ANT70	M31171	LTR	De Leys, R.	J. Virol. 64, 1207-16 (1990)
MVP5180	L20571	complete genome	Gurtler, L.	J. Virol. 68, 1581-1585 (1994)
VAU	X80020	env	Charneau, P.	Virology 205 (1), 247-253 (1994)
VI686	X96526	env	Delaporte, E.	AIDS 10, 903-910 (1996)
CONSENSUS-CPZ				
CPZGAB	X52154	complete genome	Huet, T.	Nature 345 (6273), 356-359 (1990)
CPZANT	U42720	gag, pol, vif, vpr, tat, rev, vpu, partial env	Vanden Haesevelde.C.	Virology 221, 346 (1996)

HIV1 ENV

[illegible]

[illegible]

HIV1 ENV

	V2	V3 Loop	
NEAU160	9.....YTLINCK19TITQACPKVSFEPIPIHYCAPAGFALLKCKDKRFNGKGPCKNVSTVQCTHGIRPVVSTQLLNGSLAEE.DIVIRSENFDMANNIIVQLNVSIBINCTAP.NRQVTRKX..IT		312
CONSENSUS-A	??..???.R--NT-A-----S--L-----K-----2.?v?--i-m-?--??pvt?-----s--vr		272
U455	R--NT-----R-PE--R-----R-E-R-----VNPVK--S--Y-----M--R		305
SP1703	M...IRTQ-R--NT-A-----AE--T-----I-----Q.RVK--I-N-T-----KTV-----S--VR		314
92RU020.5	T...DK-R--NT-A-----K-TE--T-----I-----G-Q--I-N-T-----DKAVK-----G--VR		296
92UG031.7	-D...NFSE-R--NT-A-----T-----R-EDD--T-L-----K-----K.GVR--ISM--T-----HTPMN-S-I-----S--VR		307
92UG037.8	SNL-R--NT-AL-----T-----R--Y--E--T-L-----Q.KVM--I-N-V-----ETVT-----S--VR		309
UG275A	Q.....R--NT-A-----OE--T-I-N-----K-----G.KVK--I-N-----L-S-TTPVT-----TS--VR		305
UG273A	R--NT-A-----E--T-L--S-----T--EVK-R--IS--TL-----TTPVK-----G--TS--VR		312
DJ264A	Q-R--NT-A-K-----T-----T-----K-----EV--I-N-----VTPVG-----S--VR		304
DJ263A	Q-R--NT-A-K-----T-----T-L-----K-----EV--I-N-Q-----VAPVR-----S--VR		308
DJ258A	Q-R--NT-A-----T-----T-----K-A-----G.EV--I-N-----VTPVR--S--G--S--VR		305
KENYA	E.....R--NT-A-----T-----R--T--T--S-----K-----G.EVM--I-N-----FAEPVK-----MS--R		306
VI191A	AQ-R--NT-A-----T-----B-E--T--R--S-----K-----K-M-----G-T-----VEPVK-----S--G--H		309
Q23	E-R--NT-A-----T-----K-EQ--T-L-----K-----K.N-T--I-N-----VQPVK-X-I-----S--R		300
CONSENSUS-B	???.r--s--nc-v-----ev--?--e-v-----s--?		290
SF2	T.....N-R--R-MR-V-----T-----N-T-----T-----I-----EV--D--N-T-----E-VA-----S--Y		311
LA1	T...TS--TS-NT-V-----EV--N-T--T-----T-----EV--A-----T-----Q-V-----S--R		313
HXB2R	T...TS-K-TS-NT-V-----N-T--T-----T-----EV--V-----T-----T-V-----R--R		309
HL43	R--S-NT-V-----N-T--T-----T-----V--A-----T-----T-V-----S--R		306
JRC5F	T...K-R--S-NT-V-----N-T--Q-----T-----KV--D-----T-----E-VK-----S--H		306
JRFL	T...S-R--S-DT-V-----I-----T-----EV--D--N-T-----KE-V-----S--H		305
BAL1	MNR-R--S-NT-V-----K-----T-----EV--A--A--V-----E-V-----S--H		312
NI	TS-R--S-NT-V-----I-----S--S-----EV-----T--H--E-VQ-----Y-K--R--H		313
SF162	K--NT-V-----S--T-----GV-----T-----KE-V-----S--		306
91US005.10	T.....R-RS-NT-V-----K-----T-----EV--K--S-----E-VV-----S--		313
91US005.11	R-RS-NT-V-----T-----K-----T-----E--A--SM--T-----A-VV-----G--S--P		315
92TH014.12	R--NT-V-----T-----Q--M-----T-----T-----EV--S-----RV-----E-V-----S--H		299
92BR020.4	R-TS-NT-V-----T-----T--Y-----G-----T-----T-----K-----L--N-----T-----KDPVD-----S--H		306
93TH067A	R-L--NT-V-----T-----T-----N-----T-----T-----EV--S-----RV-----E-V-----S--P		301
91HT651.11	R--NT-V-----R-P--K--T-----T-----K-----EVI-----T-----T--Q-----S--M		304
92HT593.1	T.....L-R--NT-V-----T-----RM--T-----K-----R.EV--N--T--I-----ETV-----S--R--S		330
CONSENSUS-C	???.r--e--nt-a-----a--y--n--c--l--n-----K-----a--i--i--n--c--h--e--v--?		285
92BR025.9	QD-R--NT-A-----D-----Y-----N-T--T--N-----I-----K--H-----E-I--K-L--V-T--H--E-VG-----S--R		305
ETH2220	TD-R--NT-----LD-----Y-----R--T--T--H-----K-----I--G.ETI--F--L--N-----I-----E-V--T-----S--ES--R		305
BU910112	R--S-NT-A-----D-----Y-----N-T--T--Q--N-----K--I-----E-I--L--N--T--H--KDPV--E-----S--R		303
BR911513	T...GQD-R--NT-A-----D-----Y-----N-T--T--N-----K--I-----E-I--K-L--V-T--H--KEPV-----S--R		313
93NN959.18	-S...GE-R--ST--V-----D-----T--Y-----N-T--T--H-----K-----G.E-I--L--N-V-T--H--E-V--V-----RS--R		308
93NN950.3	R--NT-----D-----Y-----N-T--D--T--H-----K-----G.E-I--L--N--QT--HF--E-V--V-----S--R		298
93NN955.26	SE-R--NT-V-----T--D-----Y-----N-T--T--I-----K-----E-IV--L--V-T--H--E-V--V-----S--VR		307
UG268A2	R--NT-A-----TLD-----Y-----N-T--T--N-----K--I-----I--E-I--L--N--I-----K-V--A-----ES--R		313
SM145A	R--NT-A-----D-----V-----NET--T--RY-----K-----G.E-M--L--N--T--H--Q-V--R--Y--A-----S--VR		309
2AN18A	E-R--NT-A-----TLD-----Y-----N-T--T--H-----K-----E-I--L--N--T--H--E-V--V-----S--S--R		306
2AN20A	T...KPTR-I--NT-----LG-----G-----TN--T--T--H-----K-----S--G-I--L--N--QT--K--V--A-----G-----S--R		321
DJ259A	-E...R--NT-A-----LD-----Y-----R--M--T--N-----K-----G--I--L--N--I-----E-VA-----ES--R		299
DJ373A	R--NT-S-----Y-----R--T--T-----K-----T-----I--L--N--I-----Q-V-----QS--R		305
SE364A	E-R--NT-A-----LD-----Y-----N-T--T--N-----K-----G.E-I--L--M--T--H--E-V--V-----S--MR		308
BU91007A	Y-R--NT-----D-----Y-----T--T--N-----K-----NVT-----K-N-I-I-----KEPV-----G--G		312
BU91002A	-R...D-R--NT-AT-----D-----Y-----N-T--T--N-----K-----K-I--L--N-V-T-----KDPV--E-I-----ES--R		315

HTV1 KENV

[illegible]

II-A-59
DEC 97

	V4	CD4 binding	CD4	V5				
WEAU160	...WONTGAD...	...NHITLBCRI...	...KQIINR...	...WQEVSKAMYPPIE...	...GQIRCSNITGLLLTRDGGSE...	...NQT...	...EIFRPGGGMKDNWSELYKYKVKIEFLOVAPTR...	AKRRV
CONSENSUS-D	??Tn?-??T??T??T??T??	-----	m?-g	-----	?-?-?-s	-----	??,??-s	-----
842R085	...ND-	...I-I-	...M-	...N-S-	...A-	...N-	...QEDT-	...D-R-
WHO15-726	...NG-RPD-	...I-I-	...M-	...L-K-T-	...I-AN-	...GS-	...NN-T-G-	...D-R-
87724622	...GVNES-	...SI-R-	...N-G-	...A-	...N-S-	...A-	...SS-	...D-T-
94UG1141	...SDWSDNTP-	...DET-Q-	...M-	...F-N-S-	...AI-	...SS-	...QN-T-	...D-RN-
ELI	...NI-	...ESNNSTNP-	...Q-	...KMF-	...A-I-I-	...RN-L-S-	...IN-	...S-
2226	...WA-	...ESDDK...	...P-Q-	...M-G-	...N-S-	...TN-	...SS-	...M-T-
26	...EQANS-	...ESDNK...	...L-Q-	...M-G-	...N-S-	...TN-	...SS-	...M-T-
NDK	...NS-	...GFNNQTV-	...V-L-	...R-	...L-K-S-	...AN-	...SS-	...E-TI-
92UG024.2	...IS-	...NGV-	...NT-	...K-	...V-M-	...Q-	...L-S-	...TN-
JY1	...TWNND-	...LNSEGT-K-	...M-G-	...L-K-T-	...VN-	...S-	...N-T-	...D-R-
UG269A	...STOGNES-	...E-	...M-	...I-S-	...IT-	...NS-	...N-T-	...R-L-T-
UG274A2	...NS-	...GGNET-R-Q-	...M-	...I-	...K-S-	...V-	...M-S-	...NETSSON-
SE365A2	...	...	...	...	...	...	...	...
CONSENSUS-E	mTgen?	...??T-i-	...K-?	...M-	...GA-Q-	...S-	...R-N-V-	...I-
92TH011A	...NGOCND-	...I-K-	...M-G-Q-	...S-	...R-N-V-	...I-	...DI-	...DN-
KH03	...KEGCHG-	...K-	...KN-	...G-Q-	...S-	...I-N-	...I-	...FN-
KH08	...NGOCNDA-	...I-K-	...M-G-Q-	...S-	...R-N-V-	...I-	...A-P-	...S-N-
93TH966.0	...IGOCND-	...I-K-	...M-GA-L-	...S-	...R-N-V-	...I-	...A-	...N-S-
93TH975.15	...MEGRNNS-	...DTP-V-	...K-	...M-GA-Q-	...S-	...K-N-V-	...I-	...D-
TH235	...NGOCHO-	...I-K-	...M-GA-Q-	...S-	...R-N-V-	...I-	...M-A-	...I-T-
TH242	...NGOCNS-	...I-K-	...M-GA-QP-	...R-	...K-N-V-	...I-	...A-	...N-N-
92TH022.4	...NAGCHG-	...I-K-	...M-GA-Q-	...S-	...R-N-V-	...I-	...A-	...NA-
93TH976.17	...REGA-	...K-	...V-M-	...GA-Q-	...S-	...I-K-V-	...I-	...A-
CM240	...NAGEND-	...K-	...M-G-Q-	...S-	...R-N-V-	...I-	...V-	...N-
CONSENSUS-F	...??T??	...G?NGT-	...V-M-	...R-	...S-S-	...N-T-N-	...GN?	...T-
BP_93BR029.4	...VONGT-	...V-M-	...R-	...A-A-	...N-T-S-	...GN-	...E-T-	...E-
93BR020.1	...XFP-ND-	...GFNGT-	...V-M-	...R-	...N-A-	...N-T-N-	...L-	...S-
B2163A	...GNGT-	...L-	...MV-M-	...R-	...A-S-	...EN-T-N-	...GN-	...T-
B2125A	...GNGT-	...L-	...R-V-M-	...R-	...A-	...N-T-N-	...GN-	...T-
CONSENSUS-G	...??T??	...??T-	...K-	...VRM-	...R-Q-	...A?	...R-T-K-	...?
AG_92RU131.9	...EM-	...NST-	...RM-	...R-Q-	...A-	...N-T-I-	...NM-	...NS-
AG_V1525A2	...ND-R-	...KD-Q-K-	...VRM-	...R-Q-	...AA-N-E-R-S-	...I-	...NM-	...N-
AG_92NG083	...	...K-	...VRM-	...K-Q-	...L-A-	...MLV-K-	...I-	...NM-
AG_93NG003	...	...K-	...VRM-	...R-Q-	...A-	...D-T-R-	...V-	...N-
AG_LBV217	...NSS-	...K-	...VRM-	...R-Q-	...A-	...N-T-R-	...I-	...S-
AG_92UG975.10	...	...K-	...VRM-	...R-Q-	...RA-	...N-T-R-	...Y-	...N-
CONSENSUS-H	...??Tng-	...??T??	...V-M-	...R-Q-	...?	...K-T-V-	...I-	...S-HV??
CA13	...TE-	...NG-	...Q-	...VSM-	...RA-Q-	...N-K-T-V-	...I-	...S-HV-
V1557	...IS-	...NG-	...TI-E-	...M-	...A-Q-	...K-K-T-B-	...I-	...S-HV-
90CR056	...NYTSND-	...KONE-	...V-M-	...R-R-	...Q-	...N-M-V-	...I-	...I-E-NA-
CONSENSUS-O	...??Tn?-n??	...ngT??I-	...L-R-VVRs-	...meG-Sgl-	...?	...NLT-	...QI-?	...PM-
AMT70	...VS-	...MYSQ-	...NGT-	...KL-R-VVRs-	...TROQSGL-	...K-	...NLT-M-	...MI-QM-M-TM-
MYP5180	...IKGSNE-NK-	...NGT-	...I-KL-R-LVRs-	...MKGBSRI-	...P-	...NLT-H-	...MI-QL-Q-PW-	...S-
VAU	...NISNNS-	...NGTQAI-	...L-R-VVRD-	...MRG-SGL-	...P-	...NLT-R-	...MI-QL-T-PW-	...K-
V1686	...QNMNA-H-	...NDTQ-I-	...L-R-VVRs-	...MRG-SGL-	...R-	...NLT-I-H-	...MI-QM-M-PW-	...KS-
CONSENSUS-CPZ	...	...NGT??I?	...?-V??	...??-??QI-??-??R-	...N??-??-??	...??-??-??-??	...??	...??-??-??-??-??-??-??-??
CPZGAB	...	...NGI-I-	...R-VSS-	...NR-RQI-	...R-N-T-N-	...S-TMVT-	...NS-	...GNLT-
CPZANT	...	...NGALIAK-	...V-M-	...GI-S-GI-LA-R-	...NVS-T-S-	...IN-E-	...QIY-	...E-

- fusion peptide -1		gp120\gp41		* *		***		***				
WEAU160	V...	QREKRA	VGM	LGMFLGFLGAAGSTMGASMTLTQVQRL	LLSGIVQQQNLRLRAI	EQQHLFELTVWGI	INQLQARVLAVERY	LRDQQLLG	GWGCSQKLI	CTTVPNNA	SN...	620
CONSENSUS-A	-	-	-	-	-	-	-	-	-	-	-	560
U455	-	-	-	-	-	-	-	-	-	-	-	610
SP1703	-	-	-	-	-	-	-	-	-	-	-	622
92RW030.5	-	-	-	-	-	-	-	-	-	-	-	599
92UG031.7	-	-	-	-	-	-	-	-	-	-	-	615
92UG037.8	-	-	-	-	-	-	-	-	-	-	-	611
UG275A	-	-	-	-	-	-	-	-	-	-	-	616
UG273A	-	-	-	-	-	-	-	-	-	-	-	607
DJ264A	-	-	-	-	-	-	-	-	-	-	-	615
DJ263A	-	-	-	-	-	-	-	-	-	-	-	609
DJ256A	-	-	-	-	-	-	-	-	-	-	-	609
KENYA	-	-	-	-	-	-	-	-	-	-	-	615
VI191A	-	-	-	-	-	-	-	-	-	-	-	607
QJ3	-	-	-	-	-	-	-	-	-	-	-	579
CONSENSUS-B	-	-	-	-	-	-	-	-	-	-	-	616
SP2	-	-	-	-	-	-	-	-	-	-	-	622
LAI	-	-	-	-	-	-	-	-	-	-	-	617
HXB2R	-	-	-	-	-	-	-	-	-	-	-	615
HL43	-	-	-	-	-	-	-	-	-	-	-	609
JRCSP	-	-	-	-	-	-	-	-	-	-	-	608
JRPL	-	-	-	-	-	-	-	-	-	-	-	615
BAL1	-	-	-	-	-	-	-	-	-	-	-	618
MD	-	-	-	-	-	-	-	-	-	-	-	608
SP162	-	-	-	-	-	-	-	-	-	-	-	620
91U6006.10	-	-	-	-	-	-	-	-	-	-	-	624
91US005.11	-	-	-	-	-	-	-	-	-	-	-	604
92TH014.13	-	-	-	-	-	-	-	-	-	-	-	612
92BR020.4	-	-	-	-	-	-	-	-	-	-	-	600
93TH067A	-	-	-	-	-	-	-	-	-	-	-	609
91HT651.11	-	-	-	-	-	-	-	-	-	-	-	634
92HT593.1	-	-	-	-	-	-	-	-	-	-	-	562
CONSENSUS-C	-	-	-	-	-	-	-	-	-	-	-	609
92BR025.9	-	-	-	-	-	-	-	-	-	-	-	605
ETH2320	-	-	-	-	-	-	-	-	-	-	-	602
BUS10112	-	-	-	-	-	-	-	-	-	-	-	613
BR911513	-	-	-	-	-	-	-	-	-	-	-	607
93MM959.18	-	-	-	-	-	-	-	-	-	-	-	603
93MM960.3	-	-	-	-	-	-	-	-	-	-	-	605
93MM965.26	-	-	-	-	-	-	-	-	-	-	-	610
UG268A2	-	-	-	-	-	-	-	-	-	-	-	615
SH145A	-	-	-	-	-	-	-	-	-	-	-	611
ZAM18A	-	-	-									

	- fusion peptide -		immunodominant - region -			
	gp120\gp41					
WEAU160	V.....	QREKRA..VGNLGAFLGLPLGAAGSTNGAASMTLTVDARLLLSGIVQQNNLLRAIEAQHFLFELTVWGIHQQLQARVLAVERYLADQQLLGINGCSGLICTTTTPVWNASW.....	SNR	620		
CONSENSUS-D		I.....?.....I.....Q.....LQ.....I.....h.....?.....s.....?	548			
842R085		E.....I.....QV.....LQ.....I.....H.....S.....	617			
WHO15-726		E.....I.....L.....Q.....LQ.....I.....Q.....B.....RH.....S.....K	621			
87T24622		E.....I.....V.....L.....T.....K.....D.....LQ.....IP.....S.....R.....RH.....K	612			
94UG1141		E.....I.....L.....T.....V.....L.....QV.....LQ.....I.....S.....H.....N.....S.....	605			
ELI		E.....I.....E.....R.....V.....Q.....N.....LQ.....I.....Q.....H.....N.....S.....	613			
3226		E.....I.....E.....R.....L.....Q.....LQ.....I.....I.....N.....S.....	614			
26		E.....I.....E.....V.....Q.....M.....LQ.....I.....I.....S.....	616			
MDK		E.....I.....I.....V.....V.....Q.....M.....H.....LQ.....I.....R.....RH.....N.....S.....	607			
92UG024.2		A.....I.....V.....L.....Q.....M.....LQ.....I.....S.....V.....RH.....P.....R.....S.....K	611			
JY1		E.....I.....V.....V.....VA.....G.....Q.....MLQ.....S.....H.....S.....K	624			
UG269A		E.....I.....L.....L.....QV.....LQ.....I.....S.....Q.....RH.....N.....S.....	613			
UG274A2		E.....I.....L.....T.....QV.....K.....LQ.....I.....I.....K.....S.....	623			
SE365A2		K.....I.....V.....A.....Q.....LQ.....I.....R.....RH.....N.....S.....K	612			
CONSENSUS-E		E.....I.....I.....IF.....I.....Q.....S.....LQ.....I.....K.....I.....L.....I.....A.....ST.....	603			
92TH011A		E.....I.....I.....IF.....I.....H.....S.....LQ.....I.....KF.....L.....I.....A.....ST.....	615			
KRO3		E.....I.....I.....IF.....I.....Q.....S.....LQ.....I.....KF.....L.....I.....A.....ST.....	612			
KH00		E.....I.....I.....IF.....I.....Q.....S.....LQ.....I.....KF.....L.....I.....A.....ST.....K	621			
93TH966.8		E.....I.....I.....IF.....I.....Q.....V.....S.....LQ.....I.....KF.....L.....I.....A.....ST.....	609			
93TH975.15		E.....I.....I.....IF.....I.....Q.....S.....LQ.....I.....KF.....L.....I.....A.....ST.....	607			
TN235		E.....I.....E.....I.....IF.....I.....Q.....S.....LQ.....I.....KF.....L.....I.....A.....ST.....	612			
TN242		E.....I.....I.....IF.....I.....Q.....S.....K.....LQ.....I.....KF.....L.....I.....A.....S.....	609			
92TH022.4		E.....I.....I.....IF.....I.....Q.....S.....MLQ.....I.....KF.....L.....I.....A.....ST.....K	607			
93TH975.17		E.....I.....I.....IF.....I.....Q.....S.....LQ.....I.....KF.....L.....I.....A.....ST.....K	608			
CN240		E.....I.....I.....IF.....I.....Q.....S.....LQ.....I.....KF.....L.....I.....A.....ST.....	611			
CONSENSUS-F		K.....?.....M.....L.....I.....Q.....LQ.....I.....L.....N.....S.....K	590			
BF_93BR029.4		K.....?.....M.....L.....I.....A.....Q.....LQ.....I.....L.....D.....S.....K	603			
93BR020.1		K.....R.....L.....L.....I.....Q.....S.....LQ.....I.....L.....N.....S.....K	607			
B2163A		R.....?.....M.....L.....S.....I.....Q.....S.....LQ.....I.....Q.....L.....M.....S.....K	598			
B2126A		K.....V.....I.....L.....S.....I.....Q.....E.....LQ.....I.....Q.....L.....N.....S.....K	602			
CONSENSUS-G		E.....I.....V.....I.....V.....Q.....S.....LQ.....I.....K.....F.....N.....T.....K	608			
AG_92RU131.9		G.....I.....V.....V.....T.....I.....V.....Q.....S.....LQ.....I.....K.....F.....N.....T.....K	628			
AG_V1525A2		E.....I.....I.....V.....V.....Q.....S.....LQ.....I.....F.....P.....N.....T.....K	603			
AG_92NG083		E.....G.....I.....V.....V.....Q.....I.....V.....Q.....S.....LQ.....I.....F.....P.....N.....T.....K	599			
AG_93NG003		G.....I.....T.....V.....V.....Q.....I.....V.....Q.....S.....LQ.....I.....F.....P.....N.....T.....K	618			
AG_LBV217		G.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	604			
AG_92UG975.10		G.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	584			
CONSENSUS-H		G.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	301			
CA13		G.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	296			
V1557		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	609			
90CR056		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	549			
CONSENSUS-O		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	618			
ANT70		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	630			
MVP3180		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	631			
VAU		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	632			
VI686		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	297			
CONSENSUS-CPZ		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	608			
CP2GAB		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K	619			
CP2ANT		E.....I.....V.....V.....V.....L.....I.....V.....Q.....S.....LQ.....I.....L.....R.....N.....T.....K				

[illegible]

II-A-65
DEC 97

<- rev cds end

```

WEAU160      GRIVDGFILIMVDRSLCLFLYHRLIDLLI AKRIVELLORR.....GNEALKYC.WNLLOYWQELKNSAVSLAHATAIABEGTDRVIEIVQRTCRALHIP.RAIRQGLERALL 859
CONSENSUS-A i---S---A-A-D-----S---R-FI---A-T---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 799
U455         I---S---AIA-D---N---S---R-PA---VA-A---SSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 855
SF1703       I---S---A-A-D-----S---R-FI---A-T---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 868
92RM020.5    I---S---A-A-D-----S---R-FIS---A-T---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 845
92UG031.7    I---S---A-A-D-----S---R-FI---A-T---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 861
92UG037.8    I---S---A-A-D-----S---Q-R-FI---VA-T---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 857
UG275A       I---S---A-A-D-----S---R-FI---A-T---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 846
UG273A       I---S---A-A-D-----S---R-FI---VA-T-K-H---G.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 855
DJ264A       V---S---A-A-D-----S---H-R-FV---V-A---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 853
DJ263A       I---S---A-A-D-----S---R-FV---V-A---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 861
DN258A       I---S---A-A-D-----S---Q-R-FV---V-A---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 856
KENYA        I---T---A-A-D-----S---R-FI---FET---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 855
VI191A       I---S---A-A-D-----S---R-FI---VA-T---G.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 854
Q23          I---S---A-A-D-----S---R-FI---A-T---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 853
CONSENSUS-B ?---?---A---D-----S---R---VA---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 810
SF2          V---A---A---D-----S---R---VA---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 853
LAI          I---N---S-A---D-----S---R---VT---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 861
HXB2R        I---N---S-A---D-----S---R---VT---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 856
NL43         I---N---S-A---D-----S---R---VT---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 854
JRC5F        -Q---N---A---D-----F---S---R---T-VT---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 848
JRFL         -N---A---D-----S---R---T-VT---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 847
BAL1         -P---N---A---D-----F---S---R---VN---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 854
MH           -H---AI---D-----F---S---H-R---A---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 856
SF162        SP---H-L-A---D-----S---R-I---A---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 847
91US006.10   -P---A---D-----S---R---VT---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 859
91US005.11   ---T---E---D-----S---R---VA---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 851
92TH014.12   E---S-L---E---D-----S---R---VA---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 843
92BR020.4    ---T-L-EI---D-----F---S-R-R---VA---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 851
93TH067A     E---SN---E-Q---S---R---VA---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 839
91HT651.11   SP-AH---A---D-----S---R---IA---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 848
92HT593.1    V---M---A---D-----S---R---IA---HSLKGLRLG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 873
CONSENSUS-C i---S---A-A-D-----S---R-i---A-A---SLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 802
92BR025.9    I---S---A-A-D-----S---R-I---A-A---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 855
ETH0220      I---N---AIF-D-----S---R-I---A-T---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 851
BU910112     I---S---A-A-D-----S---R-I---T-V---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 848
BR911513     V---T---A-A-D-----S---R-I---A-T---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 859
93MM959.18   I---S---A-A-D-----S---R-FI-V---A-A---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 853
93MM960.3    I---N---A-P-D-----S---Q-R---V---A-V---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 849
93MM965.26   I---S---P-A-D-----S---Q-R-FI---VA-A---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 851
UG268A2      V---IS---A-A-D-----S---R-R---A-A---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 856
SM145A       I---N---A-A-D-----S---R-FI-V---A-A---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 861
ZAM18A       I---N---A-P-D-----S---R-FI---A-AA---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 848
ZAM20A       I---N---A-V-D-----R---S-Q-R-FT---A-A---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 875
DJ259A       I---S---A-A-D-----S---R-R-I---A-A---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 831
DJ373A       T---S---A-A-D-----F---S---R-I---VA-V---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 843
SE364A       I---S---SIA-D-----S---R-FI---VA-A---I---QNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 841
BU91007A     I---S---B-A-D-----S---R-R-V---A-AA---HSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 852
BU91002A     I---N---A-A-D-----S---R---T-A---SSLRGLQNG.....G---L---L---GR---I---IT---D-V-V---GNI---G-TIG---N-----P---X 865

```

II-A-67
DEC 97

HIV1 NEF

Protein Alignment Summary Table (Nef)

Name	Accession No.	Region	First Author	Reference
CONSENSUS-A				
U455	M62320	complete genome	Oram, J.D.	ARHR 6 (9), 1073-1078 (1990)
IBNG	L39106	complete genome	Howard, T.M.	ARHR 12(15), 1413-1425 (1996)
SP1704	M81729	nef	Evans, L.	PNAS 85, 2815 (1988)
92UG037	U51190	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
Q23	AF004885	complete genome	Poss, M.	Unpublished
CONSENSUS-B				
SF2	K02007	complete genome	Sanchez-Pescador, R.	(in) Weiss, R.L., Teich, N., Varmus, H. and Coffin, J. (Eds.), <i>RNA Tumor Viruses</i> , Second Edition, 2, Vol. 2:1124-1141, Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)
LAI	K02013	complete genome	Wain-Hobson, S.	Cell 40, 9-17 (1985)
NL43	M19921	complete genome	Adachi, A.	J. Virol. 59, 284-291 (1986)
MN	M17449	complete genome	Gurgo, C.	Virology 164, 531-536 (1988)
JRCSF	M38429	complete genome	Koyanagi, S.	Science 236, 819 (1987)
JRFL	U63632	complete genome	O'Brien, W.A.	Nature 348, 69 (1990)
NH53	L07425	nef, tat, rev, env	Weidt, G.	Unpublished (1992)
P896	U39362	complete genome	Collman, R.	J. Virol. 66, 7517 (1992)
BO1	U16863	nef	Huang, Y.	J. Virol. 69 (1), 93-100 (1995)
CD1	U16869	nef	Huang, Y.	J. Virol. 69 (1), 93-100 (1995)
WEAU160	U21135	complete genome	Ghosh, S.K.	Unpublished (1995)
HEI39CS	X64745	gp41 nef	Steuler, H.	ARHR 8 (1), 53-59 (1992)
HXB2	K03455 M38432	complete genome	Rosen, C.A.	Cell 41, 813-823 (1985)
NE023	M58195	nef	Delassus, S.	J. Virol. 65, 225-231 (1991)
93JW-3	U73339	nef	Wang, B.	ARHR 13(8), 729-732 (1997)
1062-1-nef	U73370	nef	Wang, B.	ARHR 13(3): 275-279 (1996)
CONSENSUS-C				
ETH2220	U46016	complete genome	Salminen, M.O.	ARHR 12(14), 1329-1339 (1996)
92BR025	U52953	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1667 (1996)
92RW009	U88823	complete genome	Gao, F.	Unpublished;

CONSENSUS-D

ELI	K03454	complete genome	Alizon, M.	Cell 46, 63-74 (1986)
Z6	K03458	pol, vif, vpr, tat, rev, vpu, env, nef, 3' LTR	Srinivasan, A.	Gene 52, 71-82 (1987)
NDK	M27323	complete genome	Spire, B.	Gene 81, 275-284 (1989)
84ZR085	U88822	complete genome	Gao, F.	Unpublished
94UG114.1	U88824	complete genome	Gao, F.	Unpublished
CONSENSUS-E				
17-38	U48904	nef	Artenstein, A.W.	ARHR 12, 557-560 (1996)
235-32	U48907	nef	Artenstein, A.W.	ARHR 12, 557-560 (1996)
74-24	U48927	nef	Artenstein, A.W.	ARHR 12, 557-560 (1996)
89-4	U48930	nef	Artenstein, A.W.	ARHR 12, 557-560 (1996)
90CF402	U51188	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CM240	U54771	complete genome	Carr, J.K.	J. Virol. 70 (9), 5935-5943 (1996)
9-58	U48932	nef	Artenstein, A.W.	ARHR 12, 557-560 (1996)
93TH253	U51189	complete genome	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
CONSENSUS-F				
93BR0201	AF005494	complete genome	Gao, F.	Unpublished
93BR029.2	U27413	env, nef partial	Gao, F.	J. Virol. 70 (3), 1651-1657 (1996)
U40561	AF005496	complete genome	Gao, F.	Unpublished
CONSENSUS-G				
92NG083.2	U88826	complete genome	Gao, F.	Unpublished (1997)
92NG003.1	U88825	complete genome	Gao, F.	Unpublished (1997)
CONSENSUS-O				
ANT70C	L20587	complete genome	Vanden Haesevelde, M.	J. Virol. 68 (3), 1586-1596 (1994)
MVP5180	L20571	complete genome	Gurtler, L.	J. Virol. 68, 1581-1585 (1994)
CONSENSUS-CPZ				
CPZGAB	X52154	complete genome	Huet, T.	Nature 345 (6273), 356-359 (1990)

	Beta turn	SH3-binding	

MEAD160-open	DLMVYHTQGFPPDMQNTPTGPGTRY	FLCTGACFKLVPEPEKVE	EANESEENSLHHPNSLHGMDCHKEVIMDKFDSKLAFHSIVARELHPEYF
CONSENSUS-A	---n---y---T---D-?E---t?---ICQ---e?E---?---?---?---?---		
U455	---I---Y---I---Y---D-AB---TG---ICQ---V---E---T---LK-R-Y---FY---		
SP1704	---N---Y---F---T---D-E---KMTV---T---ICQ---EKRT---R---LK-R-Q-N---FY---		
92UG037	---Y---I---T---DEDE---TG---ICQ---E---T-R---S---SV-K---FY---		
Q23	---Y---F---T---D-DE---K-T---ICQ---E-R---K---R---LK-R---MY---		
CONSENSUS-B	---i---y---?---T---?---?---?---?---?---?---?---?---		
SP2	---I---Y---I---T---E-A---V-R---K---Y---Y---		
LA1	---I---Y---V---T---Y---D---K---T---V---P-R---E-R---R---N---		
HL43	---I---Y---V---T---Y---D---K---T---V---P-R---E-R---R---N---		
MM	---Y---I---T---I---K---C---Q---I---P-R---X---		
JRC87	---I---Y---A---V-P---T---D---C---Q---F---V---L---Y---		
JRFL	---Y---I---F---T---C---Q---IE-E---E-R---Y---		
NW53	---I---Y---I---T---G-DE---S---A---TE-T-R---K-R---H---M---Y---		
PE96	---I---Y---T---Y---DEG---NR-D---AMQ---VE-S-RQ---V-S---R---N---		
BO1	---I---Y---V---T---I---K---Q---T-R---E-R---R---M-Q---Y---Y---N---		
CD1	---N---Y---I---T---F---I---M-P-R---E---R---M---Y---N---		
NEPCA	---x---x---Y---x---x---D-D---x---x---C---x---x---Q---x---x---Y---		
NYB2GG	---I---Y---F---V---T---Y---D-I---K---T---V---P-R---E-R---R---N---		
NB023	---Y---F---TY---Q---E-F---V---S---K---Y---Y---		
93JM-3	---Y---P-F---T---D---K---L-Q---P---V-RL---H---R---Y---N---		
1062-1-NEF	---I---Y---T---G-E-S---NQ---TE-T-R---K---FY---		
CONSENSUS-C	---n---y---T---d-re---c---?---E-edr---k---h---rr-m---y?---		
ETH2220	---N---I---V---T---D-SE---I---A---E-EDR---K---H---RR-M---Y---		
92BR028	---N---Y---V-F---T---D-RE---T---E-SNR---Q---L---RR-M---YI-TA---		
AC_92RM009	---N---Y---V---T---Y---D-RE---D-C---L-Q---E-EDR---K---H---RR-M---Y---		
CONSENSUS-D	---n---i---I---T---e---d---T---D-C---ICQ---E-P-r---T---R-N-R---E-R---FY---		
ELI	---N---I---I---T---YE---D-QE---DT---T---ICQ---E-P-RQ---K-R-N-R---E-K---M---FY---N---		
Z6	---N---I---I---T---E---D-RE---T---T-C---VQ---E-T-R---K-R-N-R---E-K---M---FY---		
NDK	---N---I---I---T---Q---D-QE---T-R-D-C---CQ---E-P-RQ---R-N-R---LE-K---FY---		
84ZR085	---N---I---I---T---E---D-V---K-T---D-C---ICQ---E-E---V-R-N-R---E-K-K-KY---N---		
94UG1141	---I---T---E---N-KE---MT---D-C---INQ---E-P-R---V-R-N-R---E-K-KMK---Y---		
CONSENSUS-E	---n---i---I---T---D-RE---D-K---C---Q---IE-E-R---A-RQ---Y---		
17-36	---N---I---I---T---D-RE---D-KE---C---D---I---E-R---L---A-RK-I---FY---		
235-32	---N---I---I---T---D-RE---D-KE---C---D---I---E-R---L---A-RK-I---FY---		
74-24	---N---I---I---T---D-RE---D-KE---C---D---I---E-R---L---A-RK-I---FY---		
89-4	---N---I---I---T---D-RE---D-KE---C---D---I---E-R---L---A-RK-I---FY---		
90CF402	---N---Y---V---T---D-RE---D-K---C---Q---I---D-R---S---RR-I---Y---		
CM240	---N---I---I---T---D-RE---D-K---C---Q---IE-E-R---A-RK-I---FY---		
9-59	---N---D---I---T---D-RE---D-K---C---Q---IE-E-R---A-RK-I---FY---		
93TH253	---N---H---I---T---D-RE---D-K---C---Q---IE-E-R---I---A-RK-I---R---FY---		
CONSENSUS-F	---Y---?---?---?---?---?---?---?---?---?---?---?---?---		
93BR0201	---Y---I---TM---D-E---K---C---Q---E-ED---K-E---R---LR-I---R---YI-TA---		
90CR056	---N---Y---E-F---T---N-QE---Q---E-DGR---R---LT-L---VR---		
CONSENSUS-G	---N---?---?---?---?---?---?---?---?---?---?---?---?---		
AG_92NG083	---N---Y---L---T---MD-AEI---K---I---ICQ---E-EDR---V-R-N-S---RR-L---Y---		
AG_93MG003	---N---Y---F---T---R---MD-AEI---K---I---ICQ---LE-ADR---V-R---S---RR-I---Q---Y---		
CONSENSUS-H	---?---?---?---?---?---?---?---?---?---?---?---?---		
AMT70	---N---F---T---L---SE-EA-RLG-TC-RAN---ACA---FE-TW---I---RS-GNT---MT---L-Q---		
NVP5180	---I---C---P-F---Y---L---SA-EA-RLG-TW-DA---ACN---AE-ANG---I-K-Q---RS-GLT---I-LOK---L-P---		
CP2GAB	---T---F---L---L---L---L---L---L---L---L---L---L---L---L---		

204
192
205
206
215
204
201
210
206
206
182
216
216
200
204
206
214
210
205
202
201
207
206
212
207
215
206
202
204
204
204
211
204
206
204
209
206
183
207
206
166
213
211
205

This page intentionally left blank.

Protein Alignment Summary Table (Gag)

Name	Accession No.	Region	First Author	Reference
ROD	M15390	complete genome	Guyader,M.	Nature 326 (6114), 662-669 (1987)
NIHZ	J03654	complete genome	Zagury,J.F.	PNAS 85, 5941-5945 (1988)
ISY	J04498	complete genome	Franchini,G.	PNAS 86, 2433-2437 (1989)
ST	M31113	complete genome	Kumar,P.	J. Virol. 64, 890-901 (1990)
UC2	U38293	complete genome	Barnett,S.W.	Unpublished (1997)
BEN	M30502	complete genome	Kirchhoff,P.	Virology 177, 305-311 (1990)
CAM2	D00835	complete genome	Tristem,M.F.	J. Gen. Virol. 72, 721 (1991)
D194	J04542	complete genome	Kuehnelt,H.	PNAS 86, 2383-2387 (1989)
GH1	M30895	complete genome	Hasegawa,A.	ARHR 5, 593-604 (1989)
MDS	Z48731	gag pol vif tat rev nef env	Becker,M.	Unpublished
U22047 KR	U22047	complete genome	Kraus,G.K.	PNAS 90 (9), 4226-4230 (1993)
7924A1	L33081	partial p16/p28 gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
60415K1	L33076	p16/p28 partial gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
FA1	L33082	partial p16/p28 gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
HIV2ALT HIV2D205	X61240	complete genome	Dietrich Et Al,U.	ARHR 8, 1619 (1992)
UC1	L07625	complete genome	Barnett,S.W.	J. Virol. 67, 1006-1014, (1993)
EHOA	U27200	complete genome	Rey-Cuille,M.A.	Virology 202 (1), 471-476 (1994)
60667K1	L33078	p16/p28 partial gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
JA1	L33085	partial p16/p28 gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
FT	L33084	partial p16/p28 gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
7810A1	L33080	partial p16/p28 gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
ON	L33086	partial p16/p28 gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
22381	L33077	p16/p28 partial gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
FO784	L33083	gag, p16/p28 region	Gao,F.	J. Virol. 68, 7433-7477 (1994)
MM251	M19499	complete genome	Franchini,G.	Nature 328, 539-543 (1987)
MM32H	D01065	complete genome	Rud,E.W.	J. Gen. Virol. 75, 529 (1994)
MM1A11	M76764	complete genome	Luciw,P.A.	ARHR 8: 395-402 (1992)
MM239	M33262	complete genome	Regier,D.A.	ARHR 6, 1221-1231 (1990)
MM142	M16403	complete genome	Chakrabarti,L.	Nature 328, 543-547 (1987)
MNE	M32741	complete genome	Benveniste,R.B.	Unpublished (1990)

HIV2/SIV GAG

SMMH4	X14307	complete genome	Hirsch,V.M.	Nature 339, 389-391 (1989)
SIVSMMH9	M80194	complete genome	Courgnaud,V.	J. Virol. 66, 414-419 (1992)
SMMPBJ	M31325	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
SIVsmSL92a	U17646	source gag	Chen,Z.	J. Med. Primatol. 24, 108-115 (1995)
SIVsmSL92f	U48815	gag	Chen,Z.	J. Virol. 70 (6), 3617-3627 (1996)
PA1	L33087	gag ltr	Gao,F.	J. Virol. 68, 7433-7447 (1994)
STM	M83293	complete genome	Novembre,F.J.	Virology 186, 783-787 (1992)
STM22579	X60667	gag, pol	Khan,A.S.	J. Med. Primatol. 20, 167-171 (1991)

HIV2/SIV GAG

	-> Gag P17 Matrix		
CONSENSUS-A	MGAmSVLAGKKaDELEKRLRFgGKKKpRLKHIVWaaNEdrFgLaEsLLSEKESGCKIltVLGPIVPTGSENLkSLdNTVCVivWC7HAEKKVHDEaAk.kivqRLVAETgTaeKMPntSRPLAPps		126
ROD	-----R-----K-----N-----I-----G-----Q-----S-----S-----		129
NINZ	-----K-----T-----IS-E-----Y-T-----L-----R-G-----E-----I-----		129
ISY	-----I-----A-----QK-----K-T-----S-----		129
ST	-----R-----N-----L-----L-----LA-----D-----		129
UC2	-----V-----K-----R-----K-----E-----L-----LA-----I-----		129
BSN	-----V-----K-----R-----K-----E-----L-----LA-----I-----		129
CAN2	-----V-----K-----R-----K-----E-----L-----LA-----I-----		129
D194	-----V-----N-----R-----V-----K-----E-----L-----LA-----I-----		129
GH1	-----S-----V-----S-----K-----L-----L-----G-----S-----		129
MDS	-----S-----V-----S-----K-----L-----L-----G-----S-----		129
KR	-----S-----V-----S-----K-----L-----L-----G-----S-----		129
7924A	X-----R-IG-----V-I-----G-----Q-----D-----S-----A-----		93
60415K	X-----R-IG-----V-I-----G-----Q-----D-----S-----A-----		93
FA	X-----I-----N-----Q-----L-----R-----L-----		93
CONSENSUS-B	---g---s---t---v---k-m---?v---h---a---y-l---a-?-a-d---h-?-		123
D205	---g---s---t---v---k-m---?v---h---a---y-l---a-?-a-d---h-?-		126
UC1	---g---s---t---v---k-m---?v---h---a---y-l---a-?-a-d---h-?-		126
EXOA	---g---s---t---v---k-m---?v---h---a---y-l---a-?-a-d---h-?-		126
60667K	xv---e---n---g---h---a---e---y-l---g---a---gn---a-k-sk-t-		90
JA	xv---e---n---g---h---a---e---y-l---g---a---gn---a-k-sk-t-		90
FT	xv---t---h---a---v-l---q---a-s---vda---v-a---ks---		90
7610A	x---t---g---h---a---a---v-l---q---a-s---vda---v-a---ks---		90
ON	xv---r---h---a---a---v-l---q---a-s---vda---v-a---ks---		90
CONSENSUS-C	x---a---t---e---y-t---l---q---h---rn, ev-e---a---kn---a---s---		93
2236	x---a---t---e---y-t---l---q---h---rn, ev-e---a---kn---a---s---		93
CONSENSUS-D	x-----n---s---a---y---l---g---q---v---d---s---		93
FO764	x-----n---s---a---y---l---g---q---v---d---s---		93
CONSENSUS-SD	---s---v---n---v---n---s---a---y---l---h---q---v---dc---a---		129
NI251	---s---v---n---v---n---s---a---y---l---h---q---v---dc---a---		129
NI32H	---s---v---n---v---n---s---a---y---l---h---q---v---dc---a---		129
NI1A11	---s---v---n---v---n---s---a---y---l---h---q---v---dc---a---		129
NI239	---v---s---n---v---n---s---a---y---l---h---q---v---dc---a---		129
NI142	---s---v---n---v---n---s---a---y---l---h---q---v---dc---a---		129
NIE	---s---v---n---v---n---s---a---y---l---h---q---v---dc---a---		129
SDNH4	---s---e---v---n---v---n---s---a---y---l---h---q---v---dc---a---		129
SDNH9	---v---s---n---v---n---s---a---y---l---h---q---v---dc---a---		129
SDNFB14.441	---s---v---n---v---n---s---a---y---l---h---q---v---dc---a---		129
SID17646-SMEL92A	X---K---R-R-S-V---I---N---N---Q---AV-ME-DQL-A---		93
SID48815-SMEL92F	X---R-S-V---L---M---H---Q---AV-ME-DQL-A---		90
CONSENSUS-E	x-R-----N---HR---E-M---A---VY-L-AV---KHVQ-H---GGK-T-L-PQ---		94
FA	x-R-----N---HR---E-M---A---VY-L-AV---KHVQ-H---GGK-T-L-PQ---		94
CONSENSUS-SYM	---s---s---v---n---v---?---E---I---H---QV-X---V---?---A---		127
SYM	---s---s---v---n---v---?---E---I---H---QV-X---V---?---A---		129
SYM22519	---s---s---v---n---v---?---E---I---H---QV-X---V---?---A---		129

HIV2/SIV GAG

Gag P17 Matrix \ / P24 Capsid		
CONSENSUS-A	gkqgNEPVQq??gQNYcHvPLSPtTLAAWVKLvEeKKFGASVV?PGQALSEGCTFYDINQmLNCVGDhQAAMQIIRETIIneAAAdMaqHPiP.GPLPAGQLR?PRGSDIAGTTSTVeeEQIQmIR?qNP	253
ROD	E---Y---H-V-----I-----E-V-----E-----P---	257
NIH2	---VA---I---G-----VA-----E-----P---	255
ISY	---I---V-L-----V-----D---Y-QE---	257
ST	-N---VA-----E-----P---	257
UC2	---A---V-----Q-----D-----D---Y-Q---	257
BEN	--R-Y---A---V-----S-----D---Y-P---	257
CAM2	---Y---S-I-----N-----D---A---	257
D194	---A---I-----D-----D---Y-QP---	257
GW1	-R-R---TG---I-----D-----D---Y-P---	258
MDS	-R-Y---N-I-----E-----D---A---	257
KR	-R-Y---IA---S-----V-----E-----A---	257
7924A	---Y---S-I-----V---VA-----E---A---	221
60415K	-E---H-V---I-----E-V-----E---D---P---	221
PA	-S---TG---V-----I-D-----D-----D---Y-A---	222
CONSENSUS-B	-..I-Y---VA---V-L-----E-----Q---S---d---Y-a---	249
D205	---Y---LA---V-L-----E-----Q---S---M---D---Y-A---	252
UC1	---Y---IA---V-M-----Q-----Q---S---D---Y-A---	252
BHOA	R..LAV---IA---S-L-----E-----Q---S---M---E---Y-P---	252
60667K	---Y---IA---S-L-----EN-----IE---E-Q---S---M---E---C-A---	216
JA	---R-Y---VA---V-L-----E-----Q---V---D---E---Y-A---	216
FT	---R-Y---VA---V-----I---Q-----D---Y-QP---	216
7810A	---R-Y---VA---V-L-----D---E---Q---S---D---Y-A---	216
OW	---R-Y---VA---V-----I---Q-----E---Q---S---D---Y-QP---	216
CONSENSUS-C	-GR-Y---VA---V-H-----T-----L-----EH-V---D---Y-A---	221
2236	-GR-Y---VA---V-H-----T-----L-----EH-V---D---Y-A---	221
CONSENSUS-D	DR-R-Y---V---V-L-----D-----E-----D---Q---Q---D---Y-Q---	221
FO784	DR-R-Y---V---V-L-----D-----E-----D---Q---Q---D---Y-Q---	221
CONSENSUS-SD	-R---V---i---V-L-----i-----E-----L---Q---q-?-E-?-d---Y-q---	255
NM351	-R---Y---I---V-L-----I-----D-----L---Q---A-QQ---E-S---S-D---Y-Q---	256
NM12H	-R---Y---I---V-L-----I-----D-----L---Q---A-QQ---E-S---S-D---Y-Q---	256
NM1A11	-R---Y---I---V-L-----I-----D-----L---Q---A-QQ---E-S---S-D---Y-Q---	256
NM239	-R---Y---I---V-L-----I-----D-----L---Q---A-QQ---E-S---S-D---Y-Q---	256
NM142	-R---Y---I---V-L-----I-----S-----L---Q---A-QQ---E-S---S-D---Y-Q---	256
NM2	-R---Y---V---L-----I-----E-----L---Q---A-QQ---E-S---S-D---Y-Q---	256
SHM4	-R---Y---V---V-L-----E-----L---Q---E---D---Y-Q---	257
SHM9	-R---Y---V---L-----X-----E-----L---XQ---I---E-X---XQ---X-Q---	257
SHM13J14.441	-R---Y---I---L-----I-----E-----L---Q---I-P---E---D---Y-Q---	257
SIU17646-SN9L92A	-R---Y---V---V-L-----E-----L---Q---E---GR---Y-AA---	221
SIU48915-SN5L92F	-RE-Y---V---V-L-----A---E-----L---Q---E---D---Y-AA---	218
CONSENSUS-E	-G---Y---I-N---V-S-----E-----V---R---QP-QG---D-S---PA---E---Y-MF---	221
PA	-G---Y---I-N---V-S-----E-----V---R---QP-QG---D-S---PA---E---Y-MF---	221
CONSENSUS-STM	-R---Y---V---V-L-----?---?---E---?---?---Q-?-E-S---??---?---Q---	248
STM	-R---Y---V---V-L-----L---E---K---H---Q-P---E-S---SP---Y-Q---	258
STM22579	-R---Y---V---V-L-----S---E---V---Q---E-S---H-Q---	257

HIV2/SIV GAG

	P24 Capsid \ / X peptide	/ P9 NC
CONSENSUS-A	vPVGnLYRRWlQIGLQKCVRmTNPTMILD?kQGPKSpFQEYVDRFYKSLRAEQcDpAVIQdMMTQTLLvGmaNPDCILVLKGLOmNPTLEEMLTaCQGvGGFGqXARLIAGBALKEamPaPIPFPAARQQ..	379
RGD	I-----I-----I-----VIG-----.	385
NTHZ	IN-----I-----T-----V-A-----.	383
ISY	I-S-----A-----S-----I-----R-----.	384
ST	-S-I-----I-----A-----J-----A-----.	385
UC2	V-V-S-----A-----I-----LT-P-----.	385
BEN	K-I-----I-----G-S-----.	385
CAM2	I-I-S-----C-F-----.	385
D194	V-S-----I-----S-----LT-P-----.	385
GHI	V-----I-----LT-P-----.	386
MDS	V-----A-----I-----T-----.	385
KR	I-----V-----I-----LA-----.	385
7924A	V-S---	261
60415K	I----	261
FA	I---V----V--V---S---	262
CONSENSUS-B	--?-I-----I-----I-----LT-----ka	378
D205	L-I-----I-----I-----LT-----V-ka	382
UC1	I-I-----I-----I-----LT-----ka	382
BMOR	L-I-----I-----I-----LT-STH-----PRA	382
60667K	I-----N	256
JA	L-I-E-----	256
FT	V-----	256
7810A	I-L-----I-----	256
ON	S-V-----	256
CONSENSUS-C	--L-----V-----	262
223B	L-----V-----	261
CONSENSUS-D	I-----L-----I-T---	261
FOT84	I-----L-----I-T---	261
CONSENSUS-SD	I-I-v-----I-lag-l-ksg	384
NN251	I-L-V-----A-J-V-KRG	386
NN12H	I-L-V-----A-I-V-RG	386
NN1A11	I-L-V-----A-I-V-RG	386
NN239	I-L-V-----A-I-V-RG	386
NN142	I-L-V-----A-I-T-KG	386
NN5	I-L-S-V-----I-L-G-L-KG	386
NNKH4	I-L-V-----I-LR-DCL-V-KG	387
NNKO9	I-x-L-V-S-x-x-I-xI-x-k-x-D-LTGQ-L-x-y-KG	387
NNMPB714_441	I-L-V-V-----I-I-LTGQ-L-Y-KG	387
SIU17646-SMSL92A	I-L-V-I-----x	276
SIU46815-SMSL92F	I-S-R-V-I-G-x	273
CONSENSUS-E	I-D-L-V-g-	261
FA	I-D-L-V-s-	261
CONSENSUS-STH	I-L-V-?-A-?-?-I-?FQ-f-L-QG	371
STH	I-L-V-V-T-A-S-X-I-VFQ-d-L-QG	388
STH22579	I-L-V-I-A-P-I-FQ-G-L-QG	387

HIV2/SIV GAG

	P9 NC \ / P6	
CONSENSUS-A	..rttikoWNCCKeGHSArQcrAPRRQOCWKCK?GH1HancPdrBQAGFLG?GpsGKKpRNPpvaq?pgGL.....tPTAPP7DFavdLLSKYmQQGrkQrE...QreRPYKEVT2	481
ROD	..-KAF-----P-----T-----L-----V-----V-----V-----KR-----	491
NIH2	..-K-P-----MS-----S-V-----I-H-----A-----A-----V-----L-----KR-----	489
ISY	..-E-A-----S-----L-----VPSS-----M-----Q-----	490
ST	..-----A-----K-----E-----I-----I-----VE-----KR-----	491
UC2	..-R-----K-----P-----N-SR-----I-----A-----A-----L-----	491
BEW	..-KA-RY-----P-----E-----L-R-----T-A-----I-----A-----AE-----R-----	491
CAG2	..-----P-----T-----I-----V-----L-----KR-----Q-----	491
D194	..-A-R-----K-----S-----E-----M-R-Q-----A-----I-----I-----	491
GH1	..-KV-R-----T-----V-K-----E-----M-----A-----P-----I-----A-----R-----E-----	492
MD9	..-----T-----K-----I-----V-----A-----	491
KR	..-----D-----S-----V-----E-----I-----TRV-----A-----A-----L-----K-----KR-----	491
CONSENSUS-B	GK-g-VT-----?-T-----q-----SK-----E-----l-----mt-v-----vtPSAPPmPAeQNTPrQA-S-A-EM-Xs-m-q-...S-----	501
D205	GK-G-VT-----Q-T-----T-----SK-----E-----L-----MT-V-----VTPSAPPmPAeQNTPrQA-S-A-EM-XS-M-Q-...S-----	505
UC1	GK-G-VT-----V-T-R-----Q-----SK-----E-----L-----MT-V-----VTPSAPPmPAeQNTPrQA-S-A-EM-XS-KM-Q-...S-----	506
EHQA	GK-, -VT-----A-?-K-----QQ-----SK-----E-----F-----A-----IVPSAPPmPAeQNTPrQA-S-A-EM-XN-L-K-K-...N-----	504
CONSENSUS-SD	P.-Kp-----m-V-X-?-l-----M-mb-----B-----KN-?l-kq-??77s-k-----	487
MM251	P.-Kp-----ND-V-K-----L-----M-VH-----E-----KN-L-KQ-...S-K-----	491
MM32H	P.-Kp-----ND-V-X-----L-----M-VH-----E-----KN-L-KQ-KQRES-K-----	495
MM1A11	P.-Kp-----ND-V-X-----L-----M-VH-----E-----KN-L-KQ-KQRES-K-----M-----	495
MM239	P.-Kp-----ND-V-X-----L-----M-VH-----M-----E-----KN-L-KQ-KQRES-K-----	495
MM142	P.-Kp-----ND-V-K-N-----L-----M-VH-----E-----KN-HL-KQ-...S-GK-----	491
MNE	P.-Kp-----T-----QM-V-K-----F-----M-NH-----E-----KN-L-KQ-...SKRK-----	491
SMH4	Q.-K-----K-----T-V-X-E-----L-----M-N-----I-----E-----KN-KM-...N-----	492
SMH9	Q.XKI-x-x-x-----F-----X-----A-V-K-E-----L-----M-N-----x-E-----KN-KV-R-----N-----	492
SMHFBT14.441	Q.-KI-----A-V-X-E-----L-----M-N-----B-----KN-KM-R-----N-----	492
CONSENSUS-STH	..-V-----?-T-K-K7-----?-Q-K-E-V-F-----?-M-I-777777...??-A-BS-L-K-...S-KT-----	468
STM	..-V-----A-T-K-KQ-----QQ-Q-K-E-V-F-----H-M-I-----E-A-BS-L-K-...S-KT-----	492
STM22579	..-V-----T-K-X-----P-Q-K-E-V-F-----M-I-TPTAPP...EM-V-A-BS-L-K-...S-KT-----	498

HIV2/SIV GAG

		P6 \	
CONSENSUS-A	DLh18qgeLP?re??TEDLLHLNGLPGKDO9		509
R0D	-----Y--PP-----		522
NH2	---RF--A--C--T-----		519
ISY	-----H--T-----N--		520
ST	-F-Q--RQ--C--T-----		521
UC1	-----R--A-----		521
BEN	-----R--H--E-----		521
CAN2	-----A-C--T-----		521
D154	-----H-G-A-----		521
GH1	-----KA-H--A-----		522
MDS	-----K-CK--T-----T--		521
KR	-----HK--A-----		521
CONSENSUS-B	-----E--		516
D205	-----E--		521
UC1	-----E--		521
BR0A	-----E--		519
CONSENSUS-SD	-----E--		502
MN251	-----G--		506
MN32H	-----G--		510
MN1A11	-----G--		510
MN239	-----G--		510
MN142	-----G--		506
MNE	-----E--		506
SHMH4	-----E--		507
SHMH9	-----E--		507
SHMFB014.441	-----E--		507
CONSENSUS-STK	-?------E--		482
STK	-----E--		507
STM22579	-----V-----		510

HIV2/SIV POL

Protein Alignment Summary Table (Pol)

Name	Accession No.	Region	First Author	Reference
ROD	M15390	complete genome	Guyader,M.	Nature 326 (6114), 662-669 (1987)
NIHZ	J03654	complete genome	Zagury,J.F.	PNAS 85, 5941-5945 (1988)
ISY	J04498	complete genome	Franchini,G.	PNAS 86, 2433-2437 (1989)
ST	M31113	complete genome	Kumar,P.	J. Virol. 64, 890-901 (1990)
BEN	M30502	complete genome	Kirchhoff,F.	Virology 177, 305-311 (1990)
CAM2	D00835	complete genome	Tristem,M.F.	J. Gen. Virol. 72, 721 (1991)
D194	J04542	complete genome	Koehnel,H.	PNAS 86, 2383-2387 (1989)
GH1	M30895	complete genome	Hasegawa,A.	ARHR 5, 593-604 (1989)
UC2	U38293	complete genome	Barnett,S.W.	Unpublished (1997)
MDS	Z48731	gag pol vif tat rev nef env	Becker,M.	Unpublished
U22047 KR	U22047	complete genome	Kraus,G.K.	PNAS 90 (9), 4226-4230 (1993)
FA2	L33089	partial pol integrase	Gao,F.	J. Virol. 68, 7433-7477 (1994)
5132P1	L16970	pol partial	Gao,F.	ARHR 9, 703-704 (1993)
S52907	S52907	pol RT region	Warren,T.C.	Protein Expr. Purif. 3, 479-487 (1992)
HIV2ALT HIV2D205	X61240	complete genome	Dietrich Et Al,U.	ARHR 8, 1619 (1992)
UC1	L07625	complete genome	Barnett,S.W.	J. Virol. 67, 1006-1014, (1993)
EHQA	U27200	complete genome	Rey-Cuille,M.A.	Virology 202 (1), 471-476 (1994)
JA2	L33090	partial pol integrase	Gao,F.	J. Virol. 68, 7433-7477 (1994)
ONFT2	L33091	partial pol integrase	Gao,F.	J. Virol. 68, 7433-7477 (1994)
22POLB7	M87138	partial pol	Gao,F.	Nature 358, 495-499 (1992)
FORTC2	M87111	pol, RT region, partial	Gao,F.	Nature 358, 495-499 (1992)
POLC12	M87110	pol, partial integrase	Gao,F.	Nature 358, 495-499 (1992)
MM1A11	M76764	complete genome	Luciw,P.A.	ARHR 8: 395-402 (1992)
MM251	M19499	complete genome	Franchini,G.	Nature 328, 539-543 (1987)
MM32H	D01065	complete genome	Rud,E.W.	J. Gen. Virol. 75, 529 (1994)
MM239	M33262	complete genome	Regier,D.A.	ARHR 6, 1221-1231 (1990)
MM142	M16403	complete genome	Chakrabarti,L.	Nature 328, 543-547 (1987)
MNE	M32741	complete genome	Benveniste,R.E.	Unpublished (1990)
SMMH4	X14307	complete genome	Hirsch,V.M.	Nature 339, 389-391 (1989)
SIVSMMH9	M80194	complete genome	Courgnaud,V.	J. Virol. 66, 414-419 (1992)

HIV2/STV POL

SIVSMMPBJA	M31325	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
STM	M83293	complete genome	Novembre,F.J.	Virology 186, 783-787 (1992)
SMMLIB1	M62651	pol	Marx,P.A.	J. Virol. 65, 4480-4485 (1991)
SMMM7	M27256	pol	Li,Y. et al.	Unpublished (1989)
PA2	L33093	partial pol RT	Gao,F.	J. Virol. 68, 7433-7477 (1994)

HIV2/SIV POL

		-> Pol reading frame (-1 from Gag) P15		P15 peptide w/ Protease	
CONSENSUS-A	VLELWKGSTLGETVPSTQKTLLEWMO?RTHNGKLEckTG?FFRdgp?GKeapQlPrqPasaGhPNSp?rassapptgeiyAarekaegAE.?etiQrgDrgLcApRagrdL?Qr??dRGIAaPOPSLW				121
ROD	--R--T--L-----S-----S-----T-R--R--GS-----G-I-QATM-----				91
N1H2	--W--W-L--G-----P--N-----IG-----R--X--R--V-GS-----M-G-D-----				90
ISY	--W--AWTH-----KF--N-----MG-----VM-----T-R--TK--S--A-S-R--T--D-----				90
ST	--R-----G-----N--G--V--M-----F-C--NP-----D-P-R--R--VH-----R--R-A--S--P-A-ET--M--D-----				110
BEN	--R--V--T-----G--D--P--S-----SG-----TV-----G-----G--A--E--S--G-----				110
CAN2	--Q--W-L--F-----T--N-----IG--S-----T-----TR--GN--G--DN-----				90
D194	-----V--V--G--R--T--A-----S--N-----V-----R--G--Q--Q-----AP--G--T-----				128
GH1	-----D--R--N-----R--SM-----S-----SI--X--G--R-----G--GR-----KS--S--G-----				110
UC2	-----A-----GQ--G-----LT-----S-----S--G--VR--F--G-----G--G--GR-----S--G-----				110
MD8	-----D--Y--V--Q--G--W--L-----G--N-----S-----R-----A--K--N--G--DN--F-----				110
KR	-----I--C-----H--W--K-----S--D--F-----S--F--R--AR--VL--E--R--N--G-----TR--T--G--F-----				110
CONSENSUS-B	--K--k--V--M--g--VRL--S--hd--ss--e--IC--DeF--r--HDTSGGdcICaPCRSSGGDAaKkLHAd--E--E--Ep--E--L--G--G--F-----				110
D205	--M--KN--V--N--G--VRL--S--HD--AS--S--IC--DEF--R--HDTSGGDTICAPCRSSGGDAEKkLHAD--E--TS--EF--E--L--G--G--F-----				111
UCL	--N--KT--V--N--W--VRL--S--HD--AS--S--IC--DGP--R--HDTSGGDTICAPCRSSGGDAEKkLHAD--E--E--EF--E--L--G--G--F-----				111
EHCAQ3	--A--M--TA--V--N--G--VR--L--S--F--GTP--DSAICR--DEP--IRHDTSGGDSICTPCRSSGGDAEKkLHAD--E--EQ--E--L--G--G--F-----				110
CONSENSUS-SD	-----e??--CKANQ--PK--m--kngPCY--QM--Q--G--PWm-----F--H--?AS--a--CS--RG7--C--SAK--LH--VQQA--RK-----E--KQ--EAL--G--G--F-----				104
NH1A112A	-----ER--CKANQ--PK--M--KNGPCY--QM--Q--G--PWSN-----F--H--S--AS--A--CS--RGP--C--SAK--LH--VQQA--RK-----E--KQ--EAL--G--G--F-----				115
NH251	-----E--CKANQ--PK--M--KNGPCY--QM--Q--G--PWSN-----F--H--S--AS--A--CS--RGP--C--SAK--LH--VQQA--RK-----E--KQ--EAL--G--G--F-----				111
NH32H	-----EER--CKANQ--PK--M--KNGPCY--QM--Q--G--PWSN-----F--H--S--AS--A--CS--RGP--C--SAK--LH--VQQA--RK-----E--KQ--EAL--G--G--F-----				115
NH239	-----ER--CKANQ--PK--M--KNGPCY--QM--Q--G--PWSN-----F--H--S--AS--A--CS--RGP--C--SAK--LH--VQQA--RK-----E--KQ--EAL--G--G--F-----				115
NH142	-----S--R--CKANQ--PK--M--KNGPCY--QM--Q--G--PWSN-----F--H--S--AS--A--CS--RGP--C--SAK--LH--VQQA--RK-----E--KQ--EAL--G--G--F-----				111
NHE	-----ER--CKANQMPK-----TNGPCY--QM--Q--R--LH--M-----F--H--NAS--A--CS--RGP--C--SAK--LH--VQQA--RK-----E--KQ--EAL--G--G--F-----				111
SHH4	-----G--AN--N-----F--H--DAS--CS--RG--C--S--E--LHEDGQ-----E--EQ--E--L--G--G--F-----				74
SHH9	-----M--G--AN--N-----F--H--DAS--CS--RG--C--S--E--LHEDGQ-----E--EQ--E--L--G--G--F-----				77
SHH9BJ14.442	-----R--CK--NQ--PS-----V--KS--PCY--SM--SG--AN--M-----F--H--DAS--CS--RG--C--S--E--LHEDGQ-----E--EQ--E--L--G--G--F-----				109
SHH9BJ14.15	-----R--CK--NQ--PS-----V--KS--PCY--SM--SG--AN--M-----F--H--DAS--CS--RG--C--S--E--LHEDGQ-----E--EQ--E--L--G--G--F-----				109
SHH9BJ6.6	-----M--SG--AN--N-----F--H--DAS--CS--RG--C--S--E--LHEDGQ-----E--EQ--E--L--G--G--F-----				77
CONSENSUS-STM	RPE--R--TTGPSD--QM--G--LW--M-----H--NTS--A--CPSR--P--C--SAEKLH--GQE-----E--EQ--E--L--G--G--F-----				96
STM	RPE--R--TTGPSD--QM--G--LW--M-----H--NTS--A--CPSR--P--C--SAEKLH--GQE-----E--EQ--E--L--G--G--F-----				96

		Protease \ Reverse Transcriptase	
CONSENSUS-A	kRPVVTAyIEgQPVEVLLEtGAGDSIVAGieLGSNySPKLVGGIGGFINTREyKvEI7VLnRrVxATINTGDTPIINIFGRMLtLcLONSLNlpvAkIePiXimLKPGKDGp?irQWPLTkeKiealkDI		249
ROD	-----H-----R-N-----E-----K-----V-----K-----		221
NIH2	-----H-----R-N-----E-----G-----K-----V-----R-----K-----		220
ISY	-----D-----D-----R-----K-----V-----VT-----EQ-----R-----R-----		220
ST	-----KV-----V-----R-----K-----R-----K-----		240
BEN	-----D-----D-----T-----R-----VT-----R-----K-----		240
CAM2	-----K-----E-----G-----R-----R-----		220
D194	-----F-----D-----D-----P-----K-----AT-----LD-----VT-----R-----K-----		258
GH1	-----V-----Q-----D-----V-----I-----I-----K-----I-----VT-----R-----R-----		240
UC2	-----D-----D-----T-----V-----K-----R-----T-----R-----K-----V-----		240
NDS	N-----K-----E-----G-----K-----K-----V-----VI-----KV-----		240
KR	-----V-----K-----K-----V-----VI-----KV-----		240
S52907		MA-----V-----K-----	38
CONSENSUS-B	R---k-c---s---T---d---E-VQ---M---T---F---V-V?---KI---S---L---		239
D205	R---K-C---S---V---T---D---E-VQ---N---T---F---V---V-VB---KI---SR---L---		241
UC1	R---R-C---S---T---D---E-VQ---N---T---F---V---V-VK---KI---S---L---		241
EHQACG	R---K-T---S---T---H---E-VQ---V---N---T---F---R---V-VQ---E---KI---S---L---		240
CONSENSUS-D		X---L---E---F---V-VR---N---K---S---	51
PORTC2		X---L---E---F---V-VR---N---K---S---	51
CONSENSUS-S0	R-----h-----ph-T-----R-e-G-IKG-----L-----I-V-v-vt-----K-----S---v-R-		234
HN1A21AA	R-----H-----PH-T-----E-G-I-G-----L-----I-V-v-VV-----V-K-K-----S---v-R-		245
HN251	R-----H-----PH-T-----K-E-G-ING-----L-----I-V-v-VT-----V-K-K-----S---v-R-		241
HN32H	R-----H-----PH-T-----E-G-ING-----L-----I-V-v-VV-----K-K-----S---v-R-		245
HN239	R-----H-----PH-T-----E-G-ING-----L-----F-I-V-v-VV-----K-K-----S---v-R-		249
HN142	R-----H-----PH-T-----E-G-ING-----L-----I-V-v-VV-----K-K-----S---v-R-		241
HN5	R-----H-----A-----PH-T-----K-E-G-ING-----L-----E-F-I-V-v-VT-----K-----S---v-R-		241
SNM14	R-----B-----P-T-----D-K-K-G-VING-----L-----N-----I-V-v-VT-----E-K-----S---I-R-		204
SNM19	R-----B-----P-T-----D-K-K-G-VING-----L-----N-----I-V-v-VT-----K-----S---I-R-		207
SNMPEJ14.441	R-----I-----E-----P-T-----D-K-K-G-VING-----L-----N-----I-V-v-VT-----K-----S---I-R-		239
CONSENSUS-STM	R-----B-----LQ-T-V-----F---N-E-G-KING-----L-----F---V---V-VT-----KIK---S-----		226
STM	R-----H-----LQ-T-V-----F---N-E-G-KING-----L-----F---V---V-VT-----KIK---S-----		226

HIV2/SIV POL

CONSENSUS-A	CEKMEKEGQLKEAPPTNPYNTPTFAIKKKDKKKKKRLIDFRLNKVTPQFTSILQIGIPHPAGLAKKRRIcvtLbuGOAYFSIPhEDPRqYTAPTLpavvNaEPgKRYIYKVLPQGMKGSPIFQycNRQv	379
ROD	-----P-----H-----	351
NIH2	-----I-----	350
ISY	-R-----V-----Y-----	350
ST	-R-----K-----I-----S-----	370
BEN	-----K-SI-----A-M-----	370
CAM3	-----R-----	350
D194	-R-----R-----K-----E-V-----FM-I	398
GH1	-----K-----K-----H-----H-----	370
UC2	-----R-----R-----K-----Y-P-----V-----I	370
MD9	-R-----R-----	370
XR	-R-----E-----I-----T-----H-----	370
S52907	-----P-----H-----	168
CONSENSUS-B	-----V-----7-?-DPD-----S-K-	367
D205	-----I-----GR-E-I-----DPH-----QS-C-S-K-	169
UC1	-----R-----V-----E-----I-----S-K-	371
ENOMOG	-----S-----E-V-----S-K-----V-DP-----A-----L-K-----AK-	370
CONSENSUS-D	-----R-----S-----V-V-----R-----DKE-----T-----R-----NI	181
FORTC2	-----R-----S-----V-V-----R-----DKE-----T-----R-----NI	181
CONSENSUS-SD	-----D-----V-----RK-----D-E-----n-	364
HN1A11AA	-----D-----R-----V-----RK-----I-----D-E-----H-	375
HN251	-----D-----R-----V-----RK-----I-----D-E-----H-	371
HN32H	-----D-----R-----V-----RK-----I-----D-E-----H-	375
HN339	-----D-----R-----V-----RK-----I-----D-E-----H-	375
HN142	-----D-----R-----V-----RK-----I-----D-E-----H-	371
HNVE	-----D-----I-----V-----RK-----D-E-----H-----H-	371
SHMH4	-----D-----V-----R-----D-E-----M-	334
SHMH9	-----D-X-----XX-G-----V-X-----R-----D-E-----XH-X-N-	337
SHMPBJ14.441	-----D-----V-----R-----D-E-----M-----N-N-	369
CONSENSUS-STM	-----D-----R-----D-G-----NI	356
STM	-----D-----R-----D-G-----NI	356
CONSENSUS-E	x-----R-----R-S-----A-V-----KENEQ-----I-----V-DPE-----I-Q-----V-E-----	126
PAR	x-----R-----R-S-----A-V-----KENEQ-----I-----V-DPE-----I-Q-----V-E-----	126

HIV2/SIV POL

CONSENSUS-A	LEPPRKANpDvIPIQYNDILIASDRtdLEHDkvvLQLKELLNgLGFSTpDENFQKDEPygmHgyELNPTAMQLQKIQLPQKevNTVNDIQKLVQvLNMAAQIYpGZKThLCrLIrGKMTlTEEvQWTE	508
ROD	-----K--I-----R-----H-----T-----L-----	481
NIN2	-----E--I-----R-----	480
ISV	-----IV-----	480
ST	-----I-L-----R-----R-----RN-----	500
BEN	-----L-----G-----P--C-----L--DI-----S-----	500
CAM2	-----S--I-----H-----R-----L-----	480
D194	-----L-----G-----F-----I-----K-----	518
GR1	-----L-----G-----L-----L--I-----K-----	500
UC2	-----L-----G-----E--F-----T-----I-----	500
NDS	-----Q--I-----R-----S-----R-----I-----	500
KR	-----LV-----K-----K-----	500
S52907	-----K--I-----R-I-----H--Y-----I-----L-----	298
CONSENSUS-B	-D-----S--I-----S--R--S-----SM--E-----FK--K-----E-----LF-----R-I-K-----	497
D205	-D-----S--I-----S--R--S-----DM--E-----FK--K-----E-----A-----LF-----R-I-K-----	499
UC1	-D-----S--I-----S--R--S-----DM--E-----FK--KR-----E-----LF-----R-I-K-----	501
ENDACG	-D-----N--TI-----V--S--R--S-----NM--E-----FK--K-----E-----LF-----R-I-K-----	500
CONSENSUS-D	-----S--TL-----D--L--E-----H-----E--R-D-----K-----	311
PORTC2	-----S--TL-----D--L--E-----H-----E--R-D-----K-----	311
CONSENSUS-SD	-----TLV-----R-----SI--E-----P-----E--R-T-----	494
NC1A11AA	-----TLV-----N--R-----SI--E-----F-----E--R-T-----	505
NC251	-----TLV-----R-----SI--E-----F-----E--R-T-----	501
NC12H	-----TLV-----R-----SI--E-----F-----E--R-T-----	505
NC239	-----TLV-----R--S--SI--E-----P-----E--R-T-----	505
NC143	-----TLV-----R-----SI--S--E-----P-----E--R-T-----	501
NNE	-----TLV-----V-----R-----SI--E-----F-----E--R-T-----	501
SMNH4	-----TL-----R-----I--E-----F-----E--R-T-----	464
SMNH9	-----TL-----R-----SI--E-----F-----E--R-T-----H-----	467
SMNPBJ14.441	-----TL-----R-----SI--E-----F-----E--R-T-----	499
CONSENSUS-STM	-----R-----TL-----R-----N-----E--F-----E--R-D-----K-----A-----	486
STM	-----R-----TL-----R-----N-----E--P-----E--R-D-----K-----A-----	486
CONSENSUS-E	-----S--LL-----L--G--N--SMT--RM--V--DN--N-----ED--RE--LK--L--Y--K-----E--E--R-----I-----V-----N--K-----	256
PAR	-----S--LL-----L--G--N--SMT--RM--V--DN--N-----ED--RE--LK--L--Y--K-----E--E--R-----I-----V-----N--K-----	256

HIV2/SIV POL

CONSENSUS-A	LAELAELEN?iLLaQaQEGHYQEEkeLSATvqKQqdnQHCYKiHQ?ekiLKvQKyaKikNhtngvrLLhQ/VQKIGKEALVWGRIPKFHLFVBRetNEQWNDnYKQVTWlPQWDPvSTPPLVRLLePM	636
ROD	-----R-----K-----K-----V-G-----G-IC-D-KYPYQR-----D-----	611
NIH2	-----K-----K-----V-Q-----K-----I-----	610
ISY	-----K-----C-----GG-----V-----O-----I-----V-----	610
ST	-----K-----Y-----I-S-QH-----E-----E-----E-----T-----	630
BEN	-----R-----S-----E-----I-S-----A-----G-RV-----V-----D-----V-----E-----T-----	609
CAM2	-----K-----Y-----I-N-----B-----E-----E-----E-----T-----	648
D194	-----R-----D-----I-S-----E-----E-----E-----E-----T-----	630
GK1	-----RV-----Y-----G-T-----V-----I-----D-----V-----	630
UC2	-----K-----C-----G-----H-----I-----	630
MD5	-----R-----ED-----V-----I-----I-----	628
KR	-----Q-X-----E-----S-K-GVP-----NLA-----GN-----V-----H-----E-V-----D-----TD-----E-----i-----Y-----	627
S52907	-----Q-K-----E-----S-K-RVP-----NLA-----GN-V-----V-----H-----E-V-----D-----TD-----E-----I-----Y-----	629
CONSENSUS-B	-----Q-X-----E-----S-K-GVP-----NLA-----GN-----V-----H-----E-V-----D-----TD-----E-----i-----Y-----	631
D205	-----Q-K-----E-----S-K-RVP-----NLA-----GN-V-----V-----H-----E-V-----D-----TD-----E-----I-----Y-----	630
UC1	-----Q-K-----E-----S-K-GVP-----NLA-----GN-----V-----H-----E-V-----D-----TD-----E-----i-----Y-----	630
EHQACG	-----Q-K-----E-----S-K-GVP-----NLA-----GN-----V-----H-----E-V-----D-----TD-----E-----i-----Y-----	630
CONSENSUS-D	M-----Y-----K-----C-----S-P-----V-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	623
PORTC2	M-----Y-----K-----C-----S-P-----V-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	635
CONSENSUS-SD	M-----Y-----K-----C-----G-P-----I-S-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	631
MM1A11AA	M-----Y-----K-----C-----G-P-----I-N-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	635
MM251	M-----Y-----K-----C-----G-P-----I-S-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	631
MM132H	M-----Y-----K-----C-----G-P-----I-S-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	635
MM239	M-----Y-----K-----C-----G-P-----I-S-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	635
MM142	M-----Y-----K-----C-----G-P-----I-S-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	631
MM5	M-----Y-----K-----C-----G-P-----I-N-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	631
SMH14	M-----Y-----K-----C-----G-P-----I-S-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	594
SMH19	M-----Y-----K-----C-----G-P-----I-S-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	597
SMHFB314.441	M-----Y-----K-----C-----G-P-----I-S-----S-----ED-----F-----H-----I-----I-----QV-----?QV-----TD-----E-----i-----V-----	629
CONSENSUS-STM	M-----YA-----K-----Q-----R-----D-----P-----V-----S-----ED-----F-----H-----I-----I-----Q-----V-----AD-----E-----V-----	616
STM	M-----YA-----K-----Q-----R-----D-----P-----V-----S-----ED-----F-----H-----I-----I-----Q-----V-----AD-----E-----V-----	616
CONSENSUS-SO	AR?-???-???I-----V-----	16
SMH181	x-----E-----I-----V-----	21
SMH17	ARP-IRP-#CSRI-----V-----	23
CONSENSUS-E	-S-----A-----R-----G-----R-----R-----EN-----L-----N-----S-----GDR-----FPQ-----G-----	332
PAR	-S-----A-----X-----G-----R-----R-----EN-----L-----N-----S-----GDR-----FPQ-----S-----	332

Reverse Transcriptase // RNase B		
CONSENSUS-A	LVGdPipGmETFTDGCNROSKEGKagYvTDGrDkV7vLEqTtNQDAELEaPehaltDGGPKVNIiVD5QYvMGivagQPCESSEarivNqIIdEMIKKBeiYVAVvPAHKGIGGNQGVdHLVSQGIHQ	765
R0D	-----K--KK-----S-S-----K-----	741
MIH2	--E-V-----I--R-K-----A-----N-----	740
ISV	--K-----I--X--RI-----V-----V-----T--A--K-----I-----	740
ST	--K--L-----K-R--I--RL-----V--A-----K-----	760
BEN	-----K--R-----V-R-A-----N-----V-----	760
CAM2	-----T-----KI-----A-----N-----	739
D194	-----T-----R-R--S-----A-----N-----D-----V-----	778
GH1	-----R-----R-R--T-----V-----D-----V-----	760
UC2	-----T-----R-----A-----S-----D-----V--I-----	760
NDS	--K-----K-----A-----V--N-----	760
KR	--K--E-----I--RI-----A-----KL-----PL-----	760
FA	-----X--V-----N-----D-----	57
S132P1	-----X-----K-----D-----V-----I-----	57
S52907	-----X--KK-----S--K-----	338
CONSENSUS-B	--K--LE-V-Y-----I--K--K-----L-Q-Q-----I-----T-P-K-----G-----L-----	757
D205	--K--LE-R-Y-----T-----R--K-----L--E-Q-----I-A--T-P-AK-----V--G-----L-----	759
UC1	--K--LEK-V-Y-----A-----R--K-----L-Q-Q-----T-PL-----G-----R-L-----	761
EHOACG	--K--LE-V-Y-----KA-----K--KP-----L-Q-Q-----A--T-P-RE-----K--G-----L-----	760
JA	-----X--I-----T-P-K-----Q-----L-----	57
ON	-----X--I-----T-P-K-----G-----L-----	57
CONSENSUS-C	-----XL--I-----T-PL-R-----L-R--V-IS-----L-----	57
POLB7	-----XL--I-----T-PL-R-----L-R--V-IS-----L-----	57
CONSENSUS-D	--KE-LEKT--V--H-----D-E-TKA-Q-----LL-Q--A-----??--KL-----I-----?--	568
PORTC2	--KE-LEKT--V--H-----D-E-TKA-Q-----LL-Q--A-----KL-----I-----R-----	571
POLC12	-----XI-----KL-----I-----	57
CONSENSUS-6D	--K--?--a--y--v--k-----i--k--K--L--a--a-----ik-C--L-----?a-----	751
HH1A11AA	--K--E-E-Y-----K-----I--K--K-----L-----A-----IT-C--L--K--SE-----	765
HH231	--K--E-E-Y-----K-----I--K--K-----L-----T-----IT-C--L--SE-----I-----	761
HH32H	--K--K-E-Y-----K-----I--K--K-----L-----A-----IT-C--L--SE-----I-----	765
HH239	--K--E-E-Y-----K-----I--K--K-----L-----A-----IT-C--L--SE-----I-----	765
HH142	--K--E-E-Y-V--SK-----I--K--K-----L-----A-----IT-C--L--TE-----I-----	761
HNE	--K--K-E-Y-V--K-----I--K--K-----L--A--A-----IT-C--L--TE-----I-----	761
SMMH4	--KE-Q--V--R-----AKL-----YL-A--A-----I-----L-----	724
SMMH9	--KE-Q--V--R-----G-TKL-----YL-A--A-----L--K-----X--E-----E-----	727
SMMH914.441	--KE-Q--V--R-----AKL-----YL-A--A-----L-----	759
CONSENSUS-STM	--KE-LE-T--V-----I--KN-KA-----A--A--V-----IT--KL-----	746
STM	--KE-LE-T--V-----I--KN-KA-----A--A--V-----IT--KL-----	746
CONSENSUS-80	--K?-???-?--V-----??-?-I--?-?-?K7-----?--L--?--?EA-----??--??-T-?-?-?-?-?-?-?	117
SMML1B1	--KE-L#-V-----E-V-----I-----KA-----K-----L--NQ--EA-----NNL-----T-----	150
SMMH7	--K--EEIG--V--K-----I--R-S-IkP-----R--L--V--AEA-----TR--K-----T-V--T5-----L-R--I-----R-----	152

HIV2/SIV POL

	RMase H \ / Integrase	
CONSENSUS-A	VLFLIEPAQEEHRYHn7KEL7HKKgIPqUVARQIVNtCaqcQKQGEAIGHQVnaelGtWQMDcTHLeGRiITVAVHVASGFIEaVI PQeagRQTALPLLKLaSRWPTIHLHTDNGaNTSQQEVGM	893
ROD	-----V--S-----N-----S-----	871
NIH2	-----I--S-----K-----HV-----V-----	870
ISY	-----R-----G-----V--A--L--H-----L--H-----	870
ST	-----V--S-----L--K-----T-----I--V-----P-----	890
BEN	-----II--T-----L-----S-----I--V-----Y-----P-----	890
CAN2	-----T--V--C--D-----Y-----V-----	869
D194	-----I--T-----I--V-----P-----	900
GM1	-----R-----M--T-----I--V-----S-----	890
UC2	-----I--I-----I--V-----P-----	890
MDS	-----V--S-----L-----D-----V-----T-----	890
RR	-----V--S-----L--X-----D-----T-----	890
PA	-----R-----I--T-----I--V-----K--K--E-----S-----	187
5132P1	-----N-----I--T--A-----TK-----V-----G-----S-----	187
S52907	--SS	560
CONSENSUS-B	-----G--V--V-----K--S-DK-----V-----T-----D-----	887
D205	-----G--V--V-----K--S-DK-----D-----T-----PS-----	889
UC1	-----G--V--V-----L--K--S-DK-----V-----V-----T-----D-----	891
ENCAOG	I-----N--V--V-----S-DK-----S-----V-----T-----D-----	890
JA	-----G--V--V-----H--S--K--S-DK-----V-----D-----T-----D-----	187
ON	-----G--V--V-----K--S-DK-----V-----T-----R-----D-----	187
CONSENSUS-C	-----RF--A--T-----M--K-----HV-----D-----V-----T-----A-----	187
POLB7	-----RF--A--T-----M--K-----KV-----D-----V-----T-----A-----	187
CONSENSUS-D	-----D-----I--VF--L--K-----ID--DK-HR-----V?-----?-----T-----	696
PORTC3	-----D-----I--VF--L--K-----ID--DK-HR-----VT-----X-----	658
POLC12	-----D-----I--VF--L--K-----ID--DK-HR-----V--X-----T-----	187
CONSENSUS-SD	-----?-----V--VF--L--R--K--D--DK-H-----SD-----T-----A-----	880
HM1A11AA	-----D-----V--VF--L--RI--K--D--DK-H-----SD-----T-----G-----A-----	895
HM251	-----D-----V--VF--L--RI--K--D--DK-H-----SD-----T-----G-----A-----	891
HM32H	-----D-----V--VF--L--RI--K--D--DK-H-----SD-----T-----G--V-----A-----	895
HM239	-----D-----V--VF--L--RI--K--D--DK-H-----A-SD-----T-----G-----A-----	895
HM142	-----S-----I--VF--L--R--K--D--DK-H-----SD-----V-----T-----A-----	891
HNE	-----D-----V--VF--L--R--K--D--DK-H-----SD-----T-----A-----	891
SHME4	-----K-----V--VF--L--R--K--D--DK-H-----SD-----T-----G-----A-----	854
SHMH9	-----V--K--VF--XL--R--K--D--DK-HL-----X-----G-----X--T-----	857
SHNPEJ14.441	-----V--VF--L--R--K--D--DK-H-----T-----	889
CONSENSUS-STM	-----V--VF--R--K--D--DK-H-----T-----V-----	876
STM	-----V--VF--R--K--D--DK-H-----T-----V-----	876
CONSENSUS-SO	-----VF--??--R--K--D--DK--?-----?-----?-----?-----T-----?-----	234
SHML1B1	-----I--VF--Y--R--K--D--DK-H-----T-----V-----	280
SHOH7	-----V--VF--R--R--K--D--DK-----R--T--RI--I-----VV-----A-----	282

HIV2/SIV POL

CONSENSUS-A	VAMWIGIEQsFOVPVYFQSQGVYEAHHHLNQIerIRaQANT?ETIVLAAVHCNPNRRGGIGDMTPaERIIHMLTtEQEIQLqKNSklmFrVYFREGDQLNKGPGELLNKGQSAVIVRVotdIX	1072
RQD	-----I-----I-----S-----D-----E-----L-----	1001
NIH2	---V---?-----D---V-----I-----E-----	1000
ISY	---V---E---M---V---V---N---Q-----	1000
ST	-----V---A-----Q-----A-----	1020
BEH	---V---I-----R---NF---Q---Y-----E-----	1020
CAN2	---V---?-----V---S-----	999
DL94	-----I---K---Q---Y-----A-----	1038
GN1	-----I---NF---Q---Y-----A-----	1020
UC2	---V---V---R---NF---K---Q---Y-----E---A-----	1020
ND5	-----D---V---S-----	1020
RR	---T---M-----H-----	1030
PA	-----T-----I---X-----	237
5132P1	-----K-----I---X-----	237
CONSENSUS-B	---V---T-----D---D---VSI---V---E-----V---f---L-FQ---Q---Y-----E---I---E---	1017
D205	---V---?-----D---L---D---VSI---V---T-----V---F---L-FQ---Q---Y-----E---I---E---	1019
UC1	A---T---T-----D---D---VSI---Y---A---V---F---L-FQ---Q---Y-----E---L---E---	1021
ENHACC	A---T---E-----D---D---VSI---V---T-----IV---T---L-FQ---Y-----D---E---I---E---	1020
JA	---V---RT-----D---D---VSI---V---X-----	237
OM	---V---T-----D---D---VSI---V---X-----	237
CONSENSUS-C	---V---V---HT-----D-----N---X-----	237
POLB7	---V---V---HT-----D-----N---X-----	237
CONSENSUS-D	---V---T-----DK---D---SV---X-----	746
POLC12	---V---T-----DK---D---SV---X-----	237
CONSENSUS-ED	---A---HT-----D---S-----?---Q---S---F---Y-----E---L-----	1008
MM1A11AA	---A---HT-----D---SV---E---T---Q---S---F---Y-----E---L-----	1025
MM251	---A---HT-----D---SV---A---E---Q---S---F---Y-----E---L-----	1021
MM32H	---A---HT-----D---SV---Q---S---F---Y-----E---L-----	1025
MM239	---A---HT-----D---SV---Q---S---F---Y-----E---L-----	1025
MM142	---A---HT-----D---SV---Q---S---F---Y-----E---L-----	1021
MM2	---A---HT-----D---SH---L---Q---S---F---Y-----E---VL-----	1021
SHNH4	---A---T-----T---D---SI---V---Q---S---F---Y-----E---L---E---	984
SHNR9	---AX---T---X-----T---D---SI---T---I---V---Q---S---F---Y-----E---L---E---	987
SHNFB714.441	---A---T-----T---D---SI---V---Q---S---F---Y-----E---L---E---	1019
CONSENSUS-9TM	---A---T-----T---D---D---V---V---K---L---V---Q---S---F---Y-----E-----	1006
9TM	---A---T-----T---D---D---V---V---K---L---V---Q---S---F---Y-----E-----	1006
CONSENSUS-60	---?---A?---?T---?-----?---?---?---S---?-----?---Q---S---F---?-----?---E---	340
SHMLIB1	---AH---T-----T---DK---D---SI---T---V---Q---S---F---Y-----X-----	395
SHS257	---S---AD---HT---K---R---E---SV-----Q---S---F-----E---	401

HIV2/SIV POL

		Integrase \	
CONSENSUS-A	vVPRRKAKIIzDVGgRqPDS7eHLE7aRE7aGEvA?PCQVPEIQNKAFPGALCSPPQGGHGMVDLQCONI PTTAKKSSANTGILEPNTKRMALLSCSKINLVYRKVLDRCYPRLCRHPNT		1141
ROD	II-----H--G-----M-		1036
NIH2	-----K-----M--G-----M-		1035
ISY	-I-----P--H--G-----M-		1035
ST	II-----K-----H--G--N-----		1055
BEN	-----L--SP-----M-C-----		1142
CAN2	II-----L--S-----N---		1034
D194	-----L--S-----		1073
GH1	-I-----L--E-----		1055
DC2	-----L--SP-----S		1055
MDS	I-----DM--GP--T--S		1055
KR	I-----R-V--S--T--S		1055
CONSENSUS-B	-i-----h--GK-L-CsADn-DT-Q.AR-n-QS7?--A		1058
D205	-----H--GKGL-CSADN-DT-Q.AR-N-QSD		1057
UC1	-I-----N--GK-L-CGTDM-DT-Q.AR-N-QSG---A		1064
SHOACQ	-I-----N--GK-L-CSADV-DTHQ.AR--QBNQ		1058
CONSENSUS-SD	-----K--GK--S--n-DTG-.AR--S		1042
MM1A11AA	-----K--GK-M--S--N-DTG-.AR--S		1060
MM251	-----K--GK-V--S--N-DTG-.AR--		1056
MM32H	-----K--GK-V--S--N-DTG-.AR--		1060
MM239	-----K--GK-V--S--N-DTG-.AR--		1060
MM142	-----K--GK-M--S--N-DTG-.AR--		1056
MNE	-----K--GK-V--S--N-DTG-.AR--		1056
SNM14	-----K--GK-L--G--DTG-.AR--		1019
SNM19	-----K--GK-L--G--DTG-.AR--		1022
SNMPEJ14.441	-----K--GK-L--G--DTG-.AR--S		1054
CONSENSUS-STH	-----K--GK-V--G--DT--.A----		1041
STH	-----K--GK-V--G--DT--.A----		1041

Protein Alignment Summary Table (Vif)

Name	Accession No.	Region	First Author	Reference
ROD	M15390	complete genome	Guyader,M.	Nature 326 (6114), 662-669 (1987)
NIHZ	J03654	complete genome	Zagury,J.F.	PNAS 85, 5941-5945 (1988)
ISY	J04498	complete genome	Franchini,G.	PNAS 86, 2433-2437 (1989)
ST	M31113	complete genome	Kumar,P.	J. Virol. 64, 890-901 (1990)
BEN	M30502	complete genome	Kirchhoff,F.	Virology 177, 305-311 (1990)
CAM2	D00835	complete genome	Tristem,M.F.	J. Gen. Virol. 72, 721 (1991)
D194	J04542	complete genome	Kuehnel,H.	PNAS 86, 2383-2387 (1989)
GH1	M30895	complete genome	Hasegawa,A.	ARHR 5, 593-604 (1989)
JAUI	L28935	vif, complete	Jacinto,A.	Unpublished (1994)
KR	U22047	complete genome	Kraus,G.K.	PNAS 90 (9), 4226-4230 (1993)
MDS	Z48731	gag pol vif tat rev nef env	Becker,M.	Unpublished
VALI	L28934	vif, complete	Jacinto,A.	Unpublished (1994)
UC2	U38293	complete genome	Barnett,S.W.	Unpublished (1997)
UC1	L07625	complete genome	Barnett,S.W.	J. Virol. 67, 1006-1014, (1993)
D205	X61240	complete genome	Dietrich Et Al,U.	ARHR 8, 1619 (1992)
EHOA	U27200	complete genome	Rey-Cuille,M.A.	Virology 202 (1), 471-476 (1994)
MM251	M19499	complete genome	Franchini,G.	Nature 328, 539-543 (1987)
MM1A11	M76764	complete genome	Luciw,P.A.	ARHR 8: 395-402 (1992)
MM32H	D07065	complete genome	Rud,E.W.	J. Gen. Virol. 75, 529 (1994)
MM316S	L22790	vif, vpx, vpr	Kodama,T.	J. Virol. 67, 6522-6534 (1993)
MM239	M33262	complete genome	Regier,D.A.	ARHR 6, 1221-1231 (1990)
MM142	M16403	complete genome	Chakrabarti,L.	Nature 328, 543-547 (1987)
MNE	M32741	complete genome	Benveniste,R.E.	Unpublished (1990)
SMMH4	X14307	complete genome	Hirsch,V.M.	Nature 339, 389-391 (1989)
SIVSMMH9	M80194	complete genome	Courgnaud,V.	J. Virol. 66, 414-419 (1992)
SIVSMMPBJA	M31325	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
STM	M83293	complete genome	Novembre,F.J.	Virology 186, 783-787 (1992)

HIV2/SIV VIF

		← pol cds		
CONSENSUS-A	MeagkrHivVpPcNRVPGQR7HEKWKsLVKYLKyzTKDLE7VgYVPHKVKGAWWTCsAVIFPLkg7ShLEIQAYWLTPEKQWLSSysVR?TWYtEKFWTDVTPDCAD7LIHsTYFsCFLAGEVRRAIRGE			125
ROD	---D-----K-----K-----N-----I-----V-----P-----			129
NIH2	---I-----R-----K-----N-----H-----I-----T-----			129
ISY	-DQ-----A-----I-----Q-R-----N-----N-----S-G-----T-----			129
ST	---A-----R-----I-----G-----K-----E-----L-----S-----			129
BEN	---DRN-----A-----B-R-----Q-X-----HA-----L-----I-----			129
CAH2	---S-A-----K-Q-----Q-----R-D-R-----A-N-----T-----			129
D194	---N-----R-----H-----B-R-----E-E-----N-L-----S-----			129
GH1	---N-----R-----R-----E-R-----E-----N-I-----R-----Y-I-----			129
JAV1	---D-----Q-----E-C-----A-I-----S-G-----			129
KR	---E-----H-----G-----Q-N-----A-I-----R-----S-----			129
NDS	---T-----K-----E-R-----N-----A-L-----S-----T-----A-----			129
VAL1	---K-----K-----Q-X-----A-I-----S-G-----			129
UC2	---S-----R-----G-R-----E-----H-----L-----I-----C-----A-----			129
CONSENSUS-B	---E-n-----I-----L-I-----i-----e-QQ-B-----egA7-V-G-----R-F-----A-L-----eRS-Y-----V-Q-L-GS-----N-----			129
UC1	---GE-N-----I-----L-----H-----E-QQ-S-----EEAY-V-G-----R-F-----A-L-----KRS-Y-----V-Q-L-GS-----N-----			129
D205	---E-D-----I-----L-R-----I-----GE-QQ-S-----I-----NKGAM-V-G-----R-F-----A-L-----ERN-Y-----V-Q-L-GS-----S-N-----			129
EROA	---E-N-A-----I-C-----L-R-----I-----QQ-S-----BOA-V-Q-----R-F-----A-L-----ERS-Y-----V-R-L-GS-----S-N-----			129
CONSENSUS-SD	---e-----?-----Y-e-----L-r-----I-----k-----QK?-----qeg-----V-G-----T-A-i-----SRN-----y-i-L-----P-----			129
NM251	---E-----A-----I-E-----L-R-----I-----R-----QK-----F-----QEG-----V-G-H-----R-----T-A-I-----SRN-----Y-I-L-----P-----			129
NM1A11	---E-----A-----I-E-----L-R-----I-----K-----QK-----F-----QEG-----V-G-H-----R-----T-A-I-----SRN-----MY-I-L-----P-----			129
NM32H	---A-----I-E-----L-R-----I-----K-----QK-----Y-----QEG-----V-G-H-----R-----T-A-I-----SRN-----Y-I-L-----P-----V-----			129
NM3165	---E-----A-----I-E-----L-R-----I-----K-----QK-----F-----QEG-----V-G-H-----R-----T-A-I-----SRN-----MY-I-L-----P-----			129
NM239	---E-----A-----I-E-----L-R-----I-----K-----QK-----F-----QEG-----V-G-H-----R-----T-A-I-----SRN-----MY-I-L-----P-----			129
NM142	---E-----I-E-----L-R-----I-----K-----QKA-----QEG-----V-G-----R-----T-A-I-----SKD-----EY-I-L-----P-----			129
NNE	---E-----A-----I-E-----L-R-----I-----K-----QK-----Q-----Q-----V-G-----R-----THA-I-----SRN-----I-L-----P-----			129
SMNH4	---E-N-----I-E-----L-R-----I-H-----N-----QKA-----RDET-----V-G-----A-----T-A-I-----SRN-----Y-T-L-----P-----SE-----K-----			129
SMNH9	---E-N-----I-----L-----I-H-----N-----QKA-----RDEA-----V-G-----x-----T-A-I-----SRN-----A-----Y-T-L-G-----P-----SE-----			129
SMMPBJ14.441	---E-N-----I-----L-----I-H-----N-----QKA-----DEA-----V-G-----T-A-I-----SRN-----Y-T-L-G-----P-----SE-----			129
CONSENSUS-STH	---E-----I-----L-R-----I-H-----N-----E-SKA-----Q-BA-----V-G-----E-A-I-----RN-S-----Q-L-G-----P-----			129
STH	---E-----I-----L-R-----I-H-----N-----E-SKA-----Q-BA-----V-G-----E-A-I-----RN-S-----Q-L-G-----P-----			129

vpX cds ->		
CONSENSUS-A	KILGCGnYPgAHkaqVpsLQfLALVVVQGNd.rpQr??ttRlQwRrdyrtGIRvArqD?rs?kQrgseppaprayPPGvAKVLeILa9	207
ROD	-----R--RA-----DS-----R-----L-K--S--H--S--S--T--T-----E-----	215
NIN2	-----K--R--R-----S--GSA-----R-G-----M-----S--GY-----S--PT--H--L--E-----	215
ISY	-----S-----K-----K--DN-----H-----L--G--H-----CQ-----	215
ST	-----Y-----G-----DN-----H-----G--H-----	215
BEN	-----A-----Y-----XG-A-----HN-----E-H--L--G-----S--H-----	215
CAM2	-----K--R--R-----DR-----L-----S--Y-----S--G-----	215
D194	-----A-----Y-----G-----HGA--H-----Y--L--G-----E--H-----G-----	215
GH1	-----V-----Y-----XG-A-----HN-----Y--L-----S--H-----	215
JAD1	-----K-----KY-----N-----SGGH-----K-----D-----	215
KR	-----R-K--L-----G-----KNS--R--SN-W--F--L--K-G-GH-----SQ-----	215
MD9	-----T-----G-----DS-----R--A--L-----S--H-----G-S-----	215
VAL1	-----G-K--NS-----S--GL-----G-S--G--H-----D-----	215
UC2	-----A-----Y-----KQPA-----HN-----Y--LE--GR-----H-----	215
CONSENSUS-B		215
UC1	-I--y--S--eS-----R--EGKQDS--GSA-----R--Ns--SI--l--KNNRAQ--GS--Q--l--T--L--E--G--	216
D205	-I--Y--S--EG-----R--EGKQDS--GSA-----R--Ns--SI--L--KNNRAQ--GS--Q--L--TH--L--E--G--	216
EMOA	-I--W--S--TG-----R--EGKQDS--GSA-----R--Ns--SI--L--KNNRAQ--GS--Q--L--TH--L--E--G--	216
CONSENSUS-SD		211
MM251	Q-----KF-R--?-----Y--t--SHV--S--GENP-W-----N-----M--K--NS--GD-----K--TKG-N--L--G--	214
MM1A11	Q-----RF-R--Y-----Y--K--SDV--S--GENP-W-----N-----M--K--NS--GD-----CK--TRG-N--L--G--	214
MM32H	Q-----KF-R--RY-----Y--R--SYV--S--GENP-W-----S--S--M--K--NS--GD-----CK--TKG-N--L--G--	214
MM116S	Q-----RF-R--Y-----Y--K--SDV--S--GENP-W-----N-----N--K--NS--GD-----CK--TKG-N--L--G--	214
MM239	Q-----RF-R--Y-----Y--K--SDV--S--GENP-W-----N-----N--K--NS--GD-----CK--TKG-N--L--G--	214
MM142	R-----RF-R--H-----Y--R--SHV--S--GENP-W-----N--S--K--NS--GD-----CK--TRG-N--L--G--	214
MM5	Q-----RF-R--T-----Y--R--SYV--S--BNP-W-----N--S--M--K--NS--GD-----K--TRGVD--L--G--	214
SHM14	-----KF-K--N-----Y--T--SHV--S--GSDP-W-----NN-K--M--K--NS--RN--GS--KS--EG--N--L--G--	214
SHM19	-----RF-R--N-----Y--T--SHV--XS--GSDP-W-----NN-K--M--K--NS--RN--GS--SF--KG--N--L--G--	214
SHMFBJ14.441	-----KF-R--N-----Y--T--SHV--S--GENP-W-----NN-K--M--K--NS--RN--GS--SF--EGTN--L--G--	214
CONSENSUS-STM		214
STM	-----RFTK--N-----Y--K--ENV--S--EN-A-----CH--GSI-----T--NG--GH--P--K--STEGTD--L--G--	214

HIV2/SIV VPR

Protein Alignment Summary Table (Vpr)

Name	Accession No.	Region	First Author	Reference
ROD	M15390	complete genome	Guyader,M.	Nature 326 (6114), 662-669 (1987)
NIHZ	J03654	complete genome	Zagury,J.F.	PNAS 85, 5941-5945 (1988)
ISY	J04498	complete genome	Franchini,G.	PNAS 86, 2433-2437 (1989)
ST	M31113	complete genome	Kumar,P.	J. Virol. 64, 890-901 (1990)
BEN	M30502	complete genome	Kirchhoff,F.	Virology 177, 305-311 (1990)
CAM2	D00835	complete genome	Tristem,M.F.	J. Gen. Virol. 72, 721 (1991)
D194	J04542	complete genome	Kuehnelt,H.	PNAS 86, 2383-2387 (1989)
GH1	M30895	complete genome	Hasegawa,A.	ARHR 5, 593-604 (1989)
U22047 KR	U22047	complete genome	Kraus,G.K.	PNAS 90 (9), 4226-4230 (1993)
MDS	Z48731	gag, pol, vif, tat, rev, nef, env	Becker,M.	Unpublished
UC2	U38293	complete genome	Barnett,S.W.	Unpublished (1997)
UC1	L07625	complete genome	Barnett,S.W.	J. Virol. 67, 1006-1014, (1993)
D205 HIV2ALT	X61240	complete genome	Dietrich Et Al,U.	ARHR 8, 1619 (1992)
EHQA	U27200	complete genome	Rey-Cuille,M.A.	Virology 202 (1), 471-476 (1994)
MM251	M19499	complete genome	Franchini,G.	Nature 328, 539-543 (1987)
MM32H	D01065	complete genome	Rud,E.W.	J. Gen. Virol. 75, 529 (1994)
MM1A11	M76764	complete genome	Luciw,P.A.	ARHR 8: 395-402 (1992)
MM316S	L22790	vif, vpx, vpr	Kodama,T.	J. Virol. 67, 6522-6534 (1993)
MM239	M33262	complete genome	Regier,D.A.	ARHR 6, 1221-1231 (1990)
MM142	M16403	complete genome	Chakrabarti,L.	Nature 328, 543-547 (1987)
MNE	M32741	complete genome	Benveniste,R.E.	Unpublished (1990)
SMMPBJ14.441 SMMPBJ	M31325	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
SIVSMMH9	M80194	complete genome	Courgnaud,V.	J. Virol. 66, 414-419 (1992)
SMMH4	X14307	complete genome	Hirsch,V.M.	Nature 339, 389-391 (1989)
SM62A	U04982	partial env, nef, tat, rev	Hirsch,V.M.	J. Virol. 68, 2649-2661 (1994)
STM	M83293	complete genome	Novembre,F.J.	Virology 186, 783-787 (1992)

	<- vpr cds	cat cds ->	
CONSENSUS-A	MC EAPtEIPPeDgTPprepGdeuvieilLreikeEAlkHPDprLLlaLg?yIy?RHQdTLgARELir?LQRALf?HfReGc?7SRiggtRgrnPlsAiPTPRGM?6		98
ROD	-A-----L--V-----L-----I-----K-----T-----KV-----T-----GH-----G-----N--Q		105
NIH2	-----L-----R-----A-----E-----R-----T-----R-----I-----A-----GH-----G-----H		105
ISY	-----A-----L-----R-----A-----E-----R-----T-----Y-----T-----V-----T-----GH-----P-----N--Q		105
ST	-----S-----R-----T-----S-----T-----N-----A-----G-----I-----LL-----GR-----P-----R		104
SEN	-----R-----DL--SD-----T-----R-----Y-----HN-----KT-----V-----NR-----		87
CAN2	-----L-----D-----T-----QH-----A-----R-----V-----T-----MH-----G-----R--S		104
D194	-----R-----L--ST-----T-----K-----C-----N-----N-----V-----V-----I-----DR-----K-----R--A--CP--A-----H		105
GH1	-----R-----L--GD-----R-----Q-----N-----HS-----P-----I-----V-----L-----NR-----S-----R--T--FP--A-----Y		105
KR	-----A-----E-----G-----G-----LR-----TT-----N-----CA-----S-----NV-----V-----KI-----ST--F-----Q--		105
MD5	-----L-----T-----RS-----T-----RK-----T-----I-----T-----GH-----G-----S-----Q--		105
UC2	-K-----X--L--DC--X-----S--R-----DN-----N-----N-----KI-----SM--L--V--GSR-----S--R--T--CP--T-----H--		85
CONSENSUS-B	-A--AP--I--den--Q--WE--?dv--E--Q-----T--NF--S-----A--G--KI-----L-----QH-----PG--G-----PS\$		100
UC1	-A--AP--T--NES--Q--WE--EDVME--Q--R-----T--NF--S-----A--G--KI-----L-----QH-----PG--G-----PS		101
D205	-A--AP--I--den--Q--WE--IG--E--Q-----T--NF--S-----A--G--KI-----L-----QH-----SG--G-----T--P--		101
ENOA	-A--VP--I--EN--Q--WEQ--VDV--E--Q-----T--NF--N-----N-----G--KL-----L-----G--QH-----PG--G-----PSS		101
CONSENSUS-SD	E--R--NEG--Q--W--V--V--E--L-----T--NH--N-----G--I-----I--g--CH-----sg--g--c--Pc--VL		101
MM251	E--R--NEG--Q--W--V--V--E--L-----T--NH--N-----G--I-----M--G--NH-----PG--G--T--PS		97
MM32H	E--R--NEG--Q--W--V--V--E--L-----T--NH--N-----G--I-----M--G--NH-----PG--G--T--PS--VL		101
MM1A11	E--R--NEG--Q--W--V--V--E--L--G-----T--NH--N-----G--I-----M--G--IH-----PG--G-----PS--S--L		100
MM3165	E--R--NEG--Q--W--V--V--E--L-----T--NH--N-----G--I-----M--G--IH-----PG--G--x		92
MM239	E--R--NEG--Q--W--V--V--E--L-----T--NH--N-----G--I-----M--G--IH-----PG--G-----PS--S--L		101
MM142	E--R--NEG--Q--W--V--V--E--L-----T--NH--N-----G--I-----I--S--SH-----PG--G--T--PS--S--L		101
MNE	E--R--EG--Q--W--V--V--E--L-----T--NH--N-----G--KI-----M--G--TH-----SG--G--T--PS--R--L		101
SMMPBJ19.441	R--EA--Q--W--V--V--E--N-----T--N--D-----G--I-----I--C--RH-----SG--G--T--PS--VL		101
SMH49	R--EA--Q--W--xV--V--E--N-----T--N--D-----G--I-----I--G--RH-----SG--x--T--Px--VL		101
SMH4	A--R--EA--Q--W--V--V--E--L-----T--N--D-----G--I-----I--S--AH-----S--G--T--PS--VL		101
SM62A	A--R--EA--Q--W--V--V--E--L-----T--N--D-----P--G--I-----I--S--AH-----S--G--T--PS--L		101
CONSENSUS-STN	HR--EG--Q--W--V--V--E--Q--R-----S--N--D-----G--KI-----N--G--RH-----PG--G--AT--PT--VL		101
STW	HR--EG--Q--W--V--V--E--Q--R-----S--N--D-----G--KI-----N--G--RH-----PG--G--AT--PT--VL		101

HIV2/SIV TAT

Protein Alignment Summary Table (Tat)

Name	Accession No.	Region	First Author	Reference
ROD	M15390	complete genome	Guyader,M.	Nature 326 (6114), 662-669 (1987)
NIHZ	J03654	complete genome	Zagury,J.F.	PNAS 85, 5941-5945 (1988)
ISY	J04498	complete genome	Franchini,G.	PNAS 86, 2433-2437 (1989)
ST	M31113	complete genome	Kumar,P.	J. Virol. 64, 890-901 (1990)
BEN	M30502	complete genome	Kirchhoff,F.	Virology 177, 305-311 (1990)
CAM2	D00835	complete genome	Tristem,M.F.	J. Gen. Virol. 72, 721 (1991)
D194	J04542	complete genome	Kuehnelt,H.	PNAS 86, 2383-2387 (1989)
GH1	M30895	complete genome	Hasegawa,A.	ARHR 5, 593-604 (1989)
MDS	Z48731	gag, pol, vif, tat, rev, nef, env	Becker,M.	Unpublished
KR	U22047	complete genome	Kraus,G.K.	PNAS 90 (9), 4226-4230 (1993)
UC2	U38293	complete genome	Barnett,S.W.	Unpublished (1997)
UC1	L07625	complete genome	Barnett,S.W.	J. Virol. 67, 1006-1014, (1993)
D205	X61240	complete genome	Dietrich Et Al,U.	ARHR 8, 1619 (1992)
EHOA	U27200	complete genome	Rey-Cuille,M.A.	Virology 202 (1), 471-476 (1994)
MM251	M19499	complete genome	Franchini,G.	Nature 328, 539-543 (1987)
MMP11	M72323	tat, rev, nef, complete	Colombini,S.	PNAS 86, 4813-4817 (1989)
MM32H	D01065	complete genome	Rud,E.W.	J. Gen. Virol. 75, 529 (1994)
MM1A11	M76764	complete genome	Luciw,P.A.	ARHR 8: 395-402 (1992)
MM239	M33262	complete genome	Regier,D.A.	ARHR 6, 1221-1231 (1990)
MM142	M16403	complete genome	Chakrabarti,L.	Nature 328, 543-547 (1987)
MNE	M32741	complete genome	Benveniste,R.E.	Unpublished (1990)
SMMPBJ14.441 SMMPBJA	M31325	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
SIVSMMH9	M80194	complete genome	Courgnaud,V.	J. Virol. 66, 414-419 (1992)
SMMH4	X14307	complete genome	Hirsch,VM.	Nature 339, 389-391 (1989)
SM62A	U04982	partial env, nef, tat, rev	Hirsch,VM.	J. Virol. 68, 2649-2661 (1994)
STM	M83293	complete genome	Novembre,F.J.	Virology 186, 783-787 (1992)

HIV2/SIV TAT

	<- vpr cds	rev cds ->	exon \ / exon	
CONSENSUS-A	MEtPlKaPEsSL7syNEPssctTSeqdva?QelakqGaeiLaQLyrPLBaC7M?CyCH?CcyhCQLCF1kQGLGIwY?RkgrRRRtPkktKthssasD.XS16trTgnSgp7KkQkkt?ettvaz??g1G			119
ROD	K-C-F-R-T-R-T-N-S-R-M-N-C-E-P-PTF-V-A-DT-P			129
NIH2	E-C-R-T-R-T-N-S-K-D-Q-D-A-P-R-E-L-A-DL			129
ISY	G-Q-R-T-N-H-T-N-K-P-Q-F-N-C-D-S-A-P-TE-LG-L-ADF			129
ST	G-Q-A-A-Q-VSP-D-Y-Q-D-K-K-M-N-E-L-AL-IG-P			129
BBN	KP-R-TA-L-A-H-F-T-K-R-SF-S-S-E-R-P-P-D-T-E-S-A-V-TC			129
CAN2	E-G-C-R-G-ATT-L-Q-D-D-S-R-C-D-A-A-R-A-L-DF			129
D194	E-E-TA-R-L-A-H-T-S-Q-S-A-Q-R-PPP-D-T-P-VSAC			129
GH1	H-N-E-G-TA-L-H-T-S-Q-SF-A-S-R-D-T-E-T-MV-TC6			129
HDS	I-N-E-H-T-R-N-N-R-I-N-E-A-P-TE-L-A-DL-P			129
KR	S-M-C-KS-RL-Q-N-P-K-C-E-A-A-TE-T-P-AL-AR			129
UC2	K-F-WE-A-L-A-H-T-P-K-SF-V-G-R-PP-TP-IQ-D-E-P-E-P-A-TS			129
CONSENSUS-B	I-QEQ-K-Ss-S-pV-MT-G-DN-k-d-t-K-c-d-S.-k-Ss-RA-Ta-p-?L-A-D-T-gV-g?-Dl-P			125
OC1	I-QEQ-K-Ss-S-pV-MT-GSDN-D-K-K-DHG.-R-SS-RA-VFA-E-L-AN-D-T-TKV-KGL-DL-P			127
D205	I-QEQ-K-Ss-S-pV-MT-G-DN-K-D-T-K-C-D-S.-SA-RA-TAF-P-L-A-D-T-EV-OT-DF-P			127
EHOA	I-EQ-N-SSGN-S-GVMT-G-DNR-K-S-T-K-S-C-E-S.-K-SS-RA-TI-PNRR-L-A-D-T-EV-RA-DL-P			128
CONSENSUS-SD	REQ-n-a-a-?-i-a-ttp-s-nl?-Y-t-k-f-c-Eqsr?-a-ANT-n?-n-RHC-k-?k-E-V-?a-A-ap--			122
MM251	REQ-N-E-S-R-IL-A-ATTP-S-NL-Y-T-K-F-C-EQSRK-A-ANT-N-L-PN-RHC-E-AK-E-V-KA-A-AP--			129
MM211	REQ-N-E-S-R-IL-A-ATTP-S-NL-Y-T-K-F-C-EQSRK-A-ANT-N-L-PN-RHC-E-AK-E-V-KA-A-AP--			129
MM22H	REQ-N-E-S-R-I-A-ATTP-S-NL-Y-T-K-F-C-EQSRK-A-ANT-NNRP-N-RHC-E-AK-E-V-KA-A-AP--			130
MM1A11	REQ-N-E-SSGN-AAATP-S-NL-Y-T-K-F-C-EQSRK-A-ANT-NNRP-N-RHC-E-AK-E-V-KA-A-AP--			130
MM239	REQ-N-E-S-R-I-A-ATTP-S-NL-Y-T-K-F-C-EQSRK-A-ANT-N-P-N-RHC-E-AK-E-V-KA-A-AP--			129
MM142	REQ-N-E-S-R-YI-AAA-IP-S-NL-Y-T-K-F-C-EQSRK-A-ANT-N-EP-PN-IRLC-K-AK-E-V-AA-A-AP--			129
MNE	REQ-N-K-S-GR-AAAPT-L-S-NLE-E-Y-T-K-F-C-EQSRK-A-ANT-N-P-D-RHC-K-E-E-V-AA-A-AP--			129
SMMPBJ14.441.A	EQ-E-SR-H-SI-V-ADTP-S-SLE-Y-R-K-H-C-EQQ-ANT-L-R-AR-C-K-EK-E-V-AE-A-EL--			127
SM249	EQ-E-SR-H-SI-V-ADTP-S-SLEX-xV-Y-R-K-K-H-x-C-EQQ-ANTF-L-AR-C-K-EK-E-V-xB-A-DL-x			127
SM244	EQ-N-E-CR-H-SI-V-PTP-S-NLE-Y-K-R-H-C-EQH-ANPPF-N-L-AR-R-K-EK-E-V-E-AADL--			127
SM2A	EQ-N-E-CR-H-SI-V-PTP-S-NLE-Y-F-Y-K-R-H-C-EQH-ANPLP-N-L-AR-R-K-EK-EV-E-AADL--			127
CONSENSUS-STM	EQ-R-SS-AVA-TDG-M-E-W-E-C-K-F-K-VT-T-E-S.-VK-A-YPI-NWR-L-AR-K-E-EV-E-STP--			128
STM	EQ-R-SS-AVA-TDG-N-E-W-E-C-K-F-K-VT-T-E-S.-VK-A-YPI-NWR-L-AR-K-E-EV-E-STP--			128

HIV2/SIV TAT

CONSENSUS-A	r9	120
ROD	-	130
NIH2	-	130
ISY	-	130
BT	-	130
BEN	Q	130
CAN2	-	130
D194	H	130
GH1	-	130
MDS	--	130
KR	Q-	130
UC2	--	130
CONSENSUS-B	-SNTSTSRFANS	136
UC1	-	128
D205	-	128
EMOA	-SNTSTSRFANS	139
CONSENSUS-SD	--	123
MM251	-	130
MM11	-	130
MM32H	-	131
MM1A11	-	131
MM239	-	130
MM142	-	130
MNE	-	130
SMMPB314.441.A	--	128
SMKH9	-	128
SMKH4	-	128
SM62A	--	128
CONSENSUS-STM	K	129
STM	K	129

Protein Alignment Summary Table (Rev)

Name	Accession No.	Region	First Author	Reference
ROD	M15390	complete genome	Guyader,M.	Nature 326 (6114), 662-669 (1987)
NIHZ	J03654	complete genome	Zagury,J.F.	PNAS 85, 5941-5945 (1988)
ISY	J04498	complete genome	Franchini,G.	PNAS 86, 2433-2437 (1989)
ST	M31113	complete genome	Kumar,P.	J. Virol. 64, 890-901 (1990)
BEN	M30502	complete genome	Kirchhoff,F.	Virology 177, 305-311 (1990)
CAM2	D00835	complete genome	Tristem,M.F.	J. Gen. Virol. 72, 721 (1991)
D194	J04542	complete genome	Kushnel,H.	PNAS 86, 2383-2387 (1989)
GH1	M30895	complete genome	Hasegawa,A.	ARHR 5, 593-604 (1989)
KR	U22047	complete genome	Kraus,G.K.	PNAS 90 (9), 4226-4230 (1993)
MDS	Z48731	gag, pol, vif, tat, rev, nef, env	Becker,M.	Unpublished
UC2	U38293	complete genome	Barnett,S.W.	Unpublished (1997)
UC1	L07625	complete genome	Barnett,S.W.	J. Virol. 67, 1006-1014, (1993)
BHOA	U27200	complete genome	Rey-Cuille,M.A.	Virology 202 (1), 471-476 (1994)
D205	X61240	complete genome	Dietrich Et Al,U.	ARHR 8, 1619 (1992)
MM251	M19499	complete genome	Franchini,G.	Nature 328, 539-543 (1987)
SIVMMP11	M72323	tat, rev, nef, complete	Colombini,S.	PNAS 86, 4813-4817 (1989)
MM32H	D01065	complete genome	Rud,E.W.	J. Gen. Virol. 75, 529 (1994)
MM1A11	M76764	complete genome	Luciw,P.A.	ARHR 8: 395-402 (1992)
MM239	M33262	complete genome	Regier,D.A.	ARHR 6, 1221-1231 (1990)
MM142	M16403	complete genome	Chakrabarti,L.	Nature 328, 543-547 (1987)
MNE	M32741	complete genome	Benveniste,R.E.	Unpublished (1990)
SMMPBJ14.441	M31325	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
SMMH9	M80194	complete genome	Courgnaud,V.	J. Virol. 66, 414-419 (1992)
SMMH4	X14307	complete genome	Hirsch,V.M.	Nature 339, 389-391 (1989)
SM62A	U04982	partial env, nef, tat, rev	Hirsch,V.M.	J. Virol. 68, 2649-2661 (1994)
STM	M83293	complete genome	Novembre,F.J.	Virology 186, 783-787 (1992)

HIV2/SIV REV

	exon 1/ exon	
CONSENSUS-A	mnerAdeSglqrklRLiRLIHQC...NPYPqgpgTAcQRRWRRRRwzqrWrgllALAdkiytIPDPPaDspLdraiQhLQgLTiqaLPDPPTd...pae7777nncQr1Aet7nSLPaiWVRVDP	116
ROD	-----K-----I-----S-----QT-----K-----	100
NIH2	-T-----I-----K-----S-----D-----N-----PEST-----A	107
ISY	-T-----VR-----L-----R-----ER-----K-----I-----R-----P-----QT-----Q-----T-----T-----Q-----V-----	110
ST	-E-----E-----R-----K-----V-----EQT-----N-----SESID-----S-----I	107
BEN	-S-----G-----L-----R-----RQ-----LR-----V-----N-----LCAV-----T-----R-----G-----	103
CAN2	-T-----I-----R-----P-----T-----E-----D-----HP-----A	100
D194	-RD-----E-----I-----H-----Q-----R-----RQ-----FR-----V-----T-----LH-----I-----T-----Q-----G-----	103
GM1	-H-----K-----G-----E-----ER-----M-----H-----Q-----R-----LR-----V-----L-----T-----D-----R-----H-----G-----	103
KR	-G-----Q-----L-----R-----EK-----H-----V-----NS-----R-----N-----EST-----N-----G-----Y-----	121
HD5	-----R-----I-----S-----QTV-----Q-----D-----K-----P-----A-----S-----T-----	113
UC2	-C-----K-----E-----E-----YR-----G-----R-----RQ-----LR-----LH-----A-----S-----N-----R-----D-----G-----S	103
CONSENSUS-B	-La...-?D--Kg--LH-----?????-----Kr-GL-I-----R-hsls-s-tEe--L--R--n--veH--M--st???-Ta...QA?tcIPPVMDQLV-RSnPSes?G	113
UC1	-TT...-KD--KG--LH-----T-----KR-GL-I-----R-RSLG-6-TEZ--L-V-R--E--VED--M--ST...-TA...QAFTCIPPVMDQLV-RSNPSSMBG	116
EMOA	-A...-RD--KG--LH--SEYGYTD-----K-GL-I-----R-RPL-S-TEG--L--R--N-I-KD--H--ST...-TA...QA-TCIPPVMDQLV-RSNPSSSQG	122
D205	-TA...-QD--KE--LH-----KR-GL-I-----R-RS-STA--EE--ML--R--N--VED--H--PN-NQS-TT...QAPQCVPVMDQLV-RSAPSGSMG	119
CONSENSUS-SD	-ssh??-E-R-R--H-----?-----C-----N-----?-----Q-----R-----S-----t-T--L--Q--n-A-Esi--na...--?.....Lcdpt-ssr-PQD	103
MM251	-SSHER-E-RKR--H-----S--T--N--Q-----R--Q-----R--S-----T--T--L--Q--N-A-ESI--NT...-A...LCDPTKSSR-PQD	107
MMF11	-SSHER-E-RKR--H-----S--T--N--Q-----R--Q-----R--S-----T--T--L--Q--N-A-ESI--NT...-A...LCDPTKSSR-PQD	107
MM12H	-SSHER-E-RKR--H-----TD--T--N--Q-----R--Q-----R--S-----T--T--L--Q--N-A-ESI--NT...-A...LCDPT-NSR-PQD	108
MM1A11	-SSHER-E-RKR--H-----ID--T--N--Q-----R--Q-----R--S-----T--T--L--Q--N-A-ESI--NT...-A...LCDPTKSSR-PQD	108
MM239	-SNHER-E-RKR--H-----T--N--Q--K--R--Q-----R--S-----T--T--L--Q--N-A-ESI--NT...-A...LCDPT-D5R-PQD	107
MM142	-RSHTG-E-R-R--H-----T--S--S--N--QK-----Q-----R--S-----T--T--L--Q--N-A-ESI--NI...-A...LCDPT-NSR-PQD	107
WNC	-SSH-E-E-R-R--H-----T--N--Q-----R--S-----TWT--LV--Q--N-A-ESI--NI...-I...LHDPT-NPR-PQD	107
SNMFBJ14.441.A	-SSN...-E-R-R--H-----D--N--R-----Q--I-----R--S-----VT--L--Q--R--A--E--N--ASA...-P...LKDA--SPS	100
SNMHP	-GSN...-E-R-R--HF-----S--XX--N--R-----X--I--X--R--S-----T--L--Q--A--E--N--ASA...-P...LKDA--SP	100
SNMNH	-SST...-E-R-R--HF-----N--R-----Q--I-----R--S-----V--T--L--Q--A--E--N--SA...-P...LNDV-RSP	100
SN62A	-SST...-E-RKR--HF-----N--R-----Q--I-----R--S--N--T--L--Q--A--E--S--SA...-P...LNDV-RSS	100
CONSENSUS-9TH	-SDQ...-E-RKR--QF-----TD-----N-----N-----NR--S-----ST--L-V-Q--S--D-----PN...-KD...L-DT--N	100
STM	-SDQ...-E-RKR--QF-----TD-----N-----N-----NR--S-----ST--L-V-Q--S--D-----PN...-KD...L-DT--N	100

CONSENSUS-A	svPQpragYkRDSyERve?LVGg?GTar?QnTraskKDQaggR?cPPVRds?InkeTl\$	159
ISY	-----G-E---S---K-D---T---S-N-----RD-S---	158
KR	-A--CED-E--C---R---N--D-Q--C-----T-----G-G--R----	179
NDS	---L-----CO--DI--ND---CP---TR--DR-----S--G-S-	171
CONSENSUS-B	cgrdsc-?x?dlmgasQ-ag?xnRxdpQEDQ--t\$	144
UC1	CGRDSC-KR-SPHESQKDSQSNRRDPQEDQ--T	150
ERQA	CGRDSC-RGEDLVGSPQ-SGRRDHCNTQEDQ--GS	156
D205	YGRNSC-CARDLMGGSQ-SGESNRRDPQEDQ--T	153

HIV2/SIV VPX

Protein Alignment Summary Table (Vpx)

Name	Accession No.	Region	First Author	Reference
ROD	M15390	complete genome	Guyader,M.	Nature 326 (6114), 662-669 (1987)
NIHZ	J03654	complete genome	Zagury,J.F.	PNAS 85, 5941-5945 (1988)
ISY	J04498	complete genome	Franchini,G.	PNAS 86, 2433-2437 (1989)
ST	M31113	complete genome	Kumar,P.	J. Virol. 64, 890-901 (1990)
BEN	M30502	complete genome	Kirchhoff,F.	Virology 177, 305-311 (1990)
CAM2	D00835	complete genome	Tristem,M.F.	J. Gen. Virol. 72, 721 (1991)
D194	J04542	complete genome	Kuehnel,H.	PNAS 86, 2383-2387 (1989)
GH1	M30895	complete genome	Hasegawa,A.	ARHR 5, 593-604 (1989)
KR	U22047	complete genome	Kraus,G.K.	PNAS 90 (9), 4226-4230 (1993)
MDS	Z48731	gag, pol, vif, tat, rev, nef, env	Becker,M.	Unpublished
UC2	U38293	complete genome	Barnett,S.W.	Unpublished (1997)
UC1	L07625	complete genome	Barnett,S.W.	J. Virol. 67, 1006-1014, (1993)
D205	X61240	complete genome	Dietrich Et Al,U.	ARHR 8, 1619 (1992)
EHOA	U27200	complete genome	Rey-Cuille,M.A.	Virology 202 (1), 471-476 (1994)
MM251	M19499	complete genome	Franchini,G.	Nature 328, 539-543 (1987)
MM1A11	M76764	complete genome	Luciw,P.A.	ARHR 8: 395-402 (1992)
MM32H	D01065	complete genome	Rud,E.W.	J. Gen. Virol. 75, 529 (1994)
MM316S	L22790	vif, vpx, vpr	Kodama,T.	J. Virol. 67, 6522-6534 (1993)
MM239	M33262	complete genome	Regier,D.A.	ARHR 6, 1221-1231 (1990)
MM142	M16403	complete genome	Chakrabarti,L.	Nature 328, 543-547 (1987)
MNE	M32741	complete genome	Benveniste,R.E.	Unpublished (1990)
SMMPBJ14.441A	M31325	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
SMMH9	M80194	complete genome	Courgnaud,V.	J. Virol. 66, 414-419 (1992)
SMMH4	X14307	complete genome	Hirsch,V.M.	Nature 339, 389-391 (1989)
SM62A	U04982	partial env, nef, tat, rev	Hirsch,V.M.	J. Virol. 68, 2649-2661 (1994)
STM	M83293	complete genome	Novembre,F.J.	Virology 186, 783-787 (1992)

	← vif cda		
CONSENSUS-A	McdPRetvPPGNsGGeTigAF?MLQRtVEAlNREAVNHlPRELlPQVMQMSWRYNHDeqGmS?SYTKYRYLclmQKA7f1H?kkgCTClGgghGpgGM7sp?GPPPPPPpGLV?	108	
RCD	-----A-N-----E-----II-VYN-VR-----R-----X	113	
NINZ	-----E-A-----H-----T-----I-M-MYM-A-ADG-----E-SE-----X	112	
ISY	-N-I-----E-D-----R-----V-N-F-----R-E-----S-----X	113	
ST	-AG-I-----E-----I-----M-S-R-----S-----X	113	
BEN	-----R-----E-E-I-L-----A-----I-T-F-R-----W-EDM-RE-E-E-----X	113	
CAM2	-----E-A-----Q-----V-F-R-----S-----X	113	
D194	-A-R-----E-I-L-----A-T-----I-VY-F-----R-----X	113	
GH1	-----R-----E-I-L-----D-----V-F-R-----S-----X	113	
KR	-A-K-----G-V-----A-E-M-----DL-Q-----R-Y-M-V-----S-----X	111	
NDS	-A-----EQ-A-----T-----MY-GR-----R-E-----S	112	
UC2	-----R-----E-I-L-----L-E-----N-L-----S	112	
CONSENSUS-B	-----R-----Q-V-----E-icL-V-----A-RE-----I-----l-M-V-YT-----F-QE-----S-----A-	110	
UC1	-----R-----D-V-----D-E-ITEL-V-----C-A-RE-----S-----L-M-V-YT-----R-QE-----S-----AX	112	
D205	-----R-----D-V-----A-E-ITEL-V-----A-RE-----I-----L-M-V-YT-----R-QE-----S-----AX	112	
ERCA	-----R-----V-E-ET-L-ML-V-----K-A-RE-----I-----M-FA-G-RE-----S-----AS	111	
CONSENSUS-SD	-S---RI-----E-N-E-----E-----I-l-M-C-----R-E-A-----A-	111	
MM281	-S---RI-----E-N-E-----E-----Q-V-----L-M-C-----R-E-A-----AX	113	
MM1A11	-S---RI-----E-N-E-----E-----P-V-----I-L-N-C-----R-E-A-----AX	113	
MM32H	-S---RI-----V-E-N-E-----E-----Q-V-----I-L-M-C-----R-E-A-----AX	113	
MM3169	-S---RI-----E-N-E-----E-----P-V-----I-L-N-C-----R-E-A-----AS	112	
MM239	-S---RI-----E-N-E-----E-----P-V-----I-L-M-C-----R-E-A-----AX	113	
MM142	-S---RI-----E-N-E-----E-----Q-V-----I-L-M-C-----R-E-A-----AX	113	
MNE	-S---RI-----E-N-E-----E-----Q-V-----I-L-M-C-----R-E-A-----AX	113	
SYMPEJ1d.441.A	-S---RI-----D-E-A-----R-E-M-V-----I-M-M-C-----R-E-A-----AS	112	
SMH9	-XX---XI-----XX---D-H-X-B-A-----X---R-X-E-M-V-----X---I-L-M-C-----R-E-A-----AX	113	
SMH4	-S---RI-----K-E-N-E-A-----R-E-M-E-----I-L-V-C-----R-E-E-A-----T-----AX	113	
SM62A	-----E-N-G-A-----R-E-M-E-----I-L-V-C-R-----R-E-EA-----T-----AS	91	
CONSENSUS-STN	-S---RI-----A-E-E-H-----D-----R-E-----PG-V-----I-L-M-S-R-----R-E-A-----AX	113	
STN	-S---RI-----A-E-E-H-----D-----R-E-----PG-V-----I-L-M-S-R-----R-E-A-----AX	113	

HIV2/SIV ENV

Protein Alignment Summary Table (Env)

Name	Accession No.	Region	First Author	Reference
ROD	M15390	complete genome	Guyader,M.	Nature 326 (6114), 662-669 (1987)
NIHZ	J03654	complete genome	Zagury,J.F.	PNAS 85, 5941-5945 (1988)
885	M80207	env gp120 region	Nerrienet,E.	Unpublished (1992)
ALI	L25445	env	Costa Taveira,N.	ARHR 10, 223-224 (1994)
ISY	J04498	complete genome	Franchini,G.	PNAS 86, 2433-2437 (1989)
ST	M31113	complete genome	Kumar,P.	J. Virol. 64, 890-901 (1990)
BEN	M30502	complete genome	Kirchhoff,F.	Virology 177, 305-311 (1990)
D194	J04542	complete genome	Kuehnelt,H.	PNAS 86, 2383-2387 (1989)
CAM2	D00835	complete genome	Tristem,M.F.	J. Gen. Virol. 72, 721 (1991)
CAM2BR	U07105	partial env	Grezt,M.	J. Virol. 68, 2161-2168 (1994)
D1024	U07104	partial env	Grezt,M.	J. Virol. 68, 2161-2168 (1994)
D766	U07106	partial env	Grezt,M.	J. Virol. 68, 2161-2168 (1994)
D808	U07107	partial gp105 env	Grezt,M.	J. Virol. 68, 2161-2168 (1994)
D868	U07108	partial gp105 env	Grezt,M.	J. Virol. 68, 2161-2168 (1994)
GH1	M30895	complete genome	Hasegawa,A.	ARHR 5, 593-604 (1989)
CBL21E	U05350	env gp160 region	Breuer,J.	Unpublished
CBL22E	U05351	env gp160 region	Breuer,J.	Unpublished
CBL23E	U05352	env gp160 region	Breuer,J.	Unpublished
CBL24E	U05353	env gp160 region	Breuer,J.	Unpublished
MDS	Z48731	gag, pol, vif, tat, rev, nef, env	Becker,M.	Unpublished
CAM1E	U05359	env gp160 region	Breuer,J.	Unpublished
CAM3E	U05355	env gp160 region	Breuer,J.	Unpublished
CAM4E	U05356	env gp160 region	Breuer,J.	Unpublished
CAM5E	U05357	env gp160 region	Breuer,J.	Unpublished
CAM6E	U05358	env gp160 region	Breuer,J.	Unpublished
KR	U22047	complete genome	Kraus,G.K.	PNAS 90 (9), 4226-4230 (1993)
D1071	U07105	partial gp105 env	Grezt,M.	J. Virol. 68, 2161-2168 (1994)
UC2	U38293	complete genome	Barnett,S.W.	Unpublished (1997)
UC3	U38294	C2-V4 of HIV2 env	Barnett,S.W.	Unpublished
UC5	U38295	C2-V4 of HIV2 env	Barnett,S.	Unpublished
UC6	U38296	C2-V4 of HIV2 env	Barnett,S.W.	Unpublished

UC7	U38297	C2-V4 of HIV2 env	Barnett,S.	Unpublished
UC8	U38298	C2-V4 of HIV2 env	Barnett,S.W.	Unpublished
UC9	U38299	C2-V4 of HIV2 env	Barnett,S.	Unpublished
UC10	U38300	C2-V4 of HIV2 env	Barnett,S.W.	Unpublished
UC11	U38301	C2-V4 of HIV2 env	Barnett,S.	Unpublished
UC12	U38302	C2-V4 of HIV2 env	Barnett,S.W.	Unpublished
UC14	U38303	C2-V4 of HIV2 env	Barnett,S.	Unpublished
5132E2	L16964	env partial	Gao,F.	ARHR 9, 703-704 (1993)
FA3	L33095	partial env gp41 v3	Gao,F.	J. Virol. 68, 7433-7477 (1994)
GB1222	L10637	env gp41	Ottens,R.A.	AIDS 8 (3), 297-306 (1994)
CI171	X78505	partial gp36 env	Cuoto-Fernandez,J.C.	ARHR 10, 1289-1294 (1994)
HIV2CI9	X78506	env	Cuoto-Fernandez,J.C.	ARHR 10 (9), 1157-1163 (1994)
HIV2ISO53	X78507	env	Cuoto-Fernandez,J.C.	ARHR 10 (9), 1157-1163 (1994)
HIV2VI390	X78510	env	Cuoto-Fernandez,J.C.	ARHR 10 (9), 1157-1163 (1994)
HIV2VI905	X78511	env	Cuoto-Fernandez,J.C.	ARHR 10 (9), 1157-1163 (1994)
VI1056	U67349	env	Nyambi,P.N.	ARHR (1996) In press
VI1415	U67350	env	Nyambi,P.N.	ARHR (1996) In press
VI495	U67351	env	Nyambi,P.N.	ARHR (1996) In press
VI884	U67352	env	Nyambi,P.N.	ARHR (1996) In press
SL93F	U75440	env	Chen,Z.	J. Virol. 71 (5), 3953-3960 (1997)
UC1	L07625	complete genome	Barnett,S.W.	J. Virol. 67, 1006-1014, (1993)
2D205 ALT	X61240	complete genome	Dietrich Et Al,U.	ARHR 8, 1619 (1992)
EHOA	U27200	complete genome	Rey-Cuille,M.A.	Virology 202 (1), 471-476 (1994)
GH2P22	D10458	env	Kawamura,M.	Virology 188, 850-853 (1992)
IA3	L33096	partial env gp41 v3	Gao,F.	J. Virol. 68, 7433-7477 (1994)
ONFT3	L33097	partial env gp41 v3	Gao,F.	J. Virol. 68, 7433-7477 (1994)
VI1011	U67348	env	Nyambi,P.N.	ARHR (1996) In press
22381	L33077	p16/p28 partial gag	Gao,F.	J. Virol. 68, 7433-7477 (1994)
FO784	L33083	gag, p16/p28 partial	Gao,F.	J. Virol. 68, 7433-7477 (1994)
MM251	M19499	complete genome	Franchini,G.	Nature 328, 539-543 (1987)
MM32H	D01065	complete genome	Rod,E.W.	J. Gen. Virol. 75, 529 (1994)
MM1A11	M76764	complete genome	Luciw,P.A.	ARHR 8: 395-402 (1992)
MM316ZQ	L22814	env,nef	Kodama,T.	J. Virol. 67, 6522-6534 (1993)

HIV2/SIV ENV

MM132ZL	L22809	env,nef	Kodama,T.	J. Virol. 67, 6522-6534 (1993)
MM239	M33262	complete genome	Regier,D.A.	ARHR 6, 1221-1231 (1990)
CTL7202	M61062	env	Burns,D.P. W.	J. Virol. 65, 1843-1854 (1991)
MM142	M16403	complete genome	Chakrabarti,L.	Nature 328, 543-547 (1987)
MM18101	U19595	env, partial	Lane,T.E.	Virology 212 (2), 458-465 (1995)
SEGA	L26938	env	Kodama,T.	Unpublished (1994)
SEGE	L26942	env	Kodama,T.	Unpublished (1994)
SEQJ	L26947	env	Kodama,T.	Unpublished (1994)
7FBE2	U18019	env gp120 partial	Zhu,G.W.	J. Virol. (1995) In press
7FLG1	U18023	env gp120 partial	Zhu,G.W.	J. Virol. (1995) In press
7FLN1	U18028	env gp120 partial	Zhu,G.W.	J. Virol. (1995) In press
7FSE1	U18034	env gp120 partial	Zhu,G.W.	J. Virol. (1995) In press
MNE	M32741	complete genome	Bernveniste,R.E.	Unpublished (1990)
MNELN11	U06277	env partial	Chackerian,B.	J. Virol. 68 (6), 4080-4085 (1994)
MNESPL17	U06352	env partial	Chackerian,B.	J. Virol. 68 (6), 4080-4085 (1994)
MNELIV25	U06377	env partial	Chackerian,B.	J. Virol. 68 (6), 4080-4085 (1994)
PHTBE1	U18039	env gp120 complete	Zhu,G.W.	J. Virol. (1995) In press
PH7LG1	U18046	env gp120 complete	Zhu,G.W.	J. Virol. (1995) In press
SMP209	L20008	env	Hynes,N.A.	ARHR 9, 803-806 (1993)
P209C15	L20009	env	Hynes,N.A.	ARHR 9, 803-806 (1993)
SMMPBJ14.15 SMMPBJB	L03295	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
SMMPBJ14.441 SMMPBJA	M31325	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
SMMPBJ6.6	L09212	env polyprotein gag nef	Hirsch,V.M.J.	J. Virol. 67, 2466-2474 (1993)
SIVSMMH9	M80194	complete genome	Courgnaud,V.	J. Virol. 66, 414-419 (1992)
SMMH4	X14307	complete genome	Hirsch,V.M.	Nature 339, 389-391 (1989)
SM62A	U04982	env, nef, tat, rev	Hirsch,V.M.	J. Virol. 68, 2649-2661 (1994)
HUMB670	S67406	env	Khabbaz,R.F.	N. Engl. J. Med. 330, 172-177 (1994)
SIVB670	M90048	env, nef, LTR	Mullins,J.I.	Unpublished (1992)
STM	M83293	complete genome	Novembre,F.J.	Virology 186, 783-787 (1992)
STMAK3	X60668	env	Khan,A.S.	J. Med. Primatol. 20, 167-171 (1991)
HIV7312AG	L33079	p16/p28, partial gag, env	Gao,F.	J. Virol. 68, 7433-7477 (1994)
SMCI2	X78508	env	Cuoto-Fernandez,J.C.	ARHR 10, 1289-1294 (1994)
SMCI8	X78509	env	Cuoto-Fernandez,J.C.	ARHR 10 (9), 1157-1163 (1994)

This page intentionally left blank.

HIV2/SIV ENV

CONSENSUS-A	N?gcnqLlieilltSecIlyC??qVTVfYqIPaWnNAsIPLfC.AtkQRDTMGSTIQCLPDNDYQEI?1NVTFAFDAM?DRVTeQAIEDVWnLPETSikPCVKLTPLCvAh?C????????N?L	110
R0D	...M-----A-----V--T,-----V-T-X-T-----R-----T-----N-----H-----K-SSTESST..G-N-	121
NIH2	-KGSK-----V-A--Y--M-X,.-F-----T-----N-----V-----N-T.....R-M-	118
#05	-----X-----V--S-----Q-----R-----S-----	48
ALI	-MSS-----VT--A-----V--X,-----V--X-----A-----V--R-----I--K-S.....-IS	117
ISY	..S-KI--V-P-----T..X-----V-V-X-----P-----I--V-----T-M.....-AS	115
ST	..C-----PV-S--A-----V,-----V-V-----A-----N-----V--S-----R-----S-----	113
BEN	-EP-----FVV-----Y--S,-----X-----I-----N-----V--H-----N-SRVO....G-T-	121
D194	-EP-----V-----K,-----T-----R-----N-----I-----	116
CAN2	-ER-----A-----R.Q-----V--K-----P-----I-----K-----IS	117
CAN2BR	-GC-----A-----K.Q-----V--K-----P-----I-----K-----IS	117
D1024	..GA-----T-----R,-----S--G-----V-----A-----N-----T-----T-RKSS...C-TS	118
D766	-TRR-A--B-----K,.-H-----T-----T-----N-----T-----N-SKNS...G-TS	121
D608	-APQRT-----S-----E,-----A-----R-----A-----N-----T-----N-SKNY...N-TS	121
D868	-ARE-T--I--T--A--R-Q,-----K-----A-----N-----T-----N-SKDS...S-TS	121
GH1	..C-KSL-CV-S--A--Y-V--T,-----V-V-----K-----T-----Y--S-----S-----S-----	113
CBL21	..CS--L--V-S--A--I-----T,-----V-----A-----V-----S-----T-N-T.TM....K-IS	119
CBL22	..S-KI--V-----V--T,-----V-----LT-----H-----N-TKTR...R-M-	120
CBL23	-MG-----V--A-T-----T..H-----T-----R-----N-----	116
CBL24	..C-----VTT--A-----S-T..K-----R-----A-----V--S-----K-----N-----	115
HDS	-TRKMH-----KPKQ-----E--A-----V--S-----N-SRLGDGASTR-T-	128
CAM1	-AYK-----KRX-----F-----N-----K-D.....G-I-	120
CAM3	-TRNG-----A--Y--R.Q-----T-----S-----N-----S-----K-----N-----	117
CAM4	-AC-KS-----T-----A,-----P-----N--I-----K-RKMT...G-T-	121
CAM5	-TC-----Y-VH-K,-----V-V-----A-----N--D--V-----IK-E-KKL....G-A-	121
CAM6	-GRK-----I-----I--K,-----P-----V-----N-SRSP...S-S-	122
KR	-D.S-----IV-----A,-----K-----R-----E-----N-----V-----V-----Q-E-NSTSTES..S-S-	123
D1071	-----T-SKNS...G-TS	11
UC2	-BP-----AV-----K,-----V-V-----E-----R-----N-NPVT....Q-N-	121
CONSENSUS-B	-Ah?s-?-?-?L--I-iyGf?gk--E-----EV-----L-----V-----p--T--I-----?-vd--S-----nkt-e-...--?	110
UC1	-ANTS-H-F-LL--I-VYGFLGHKKR-----TV-----T-----V-----G--T--SV-I-----N-----VD--S-----R-NMTGT....-T-	123
D205	-AYPSSR-P--L--IGISGFV-K,-----TV--I--T-----V-----G--T--R--I-----Q--VD--R-----N-SKTETMP..G-AS	124
ENQA	-AMVN-Y--VTL--I-IYGVWCK..NF-----K-----R-----V-----T--Q--I-----D--TK--S-----T-K-NKWTSS....	119

CONSENSUS-SD	-GCLG-----L-VYG--t..l-----v-----T-----?-----T-----g--s-la-----s-----7-----q-----s-----Ic-x-NKSEDr.WG.L-	123
MM251	-GCLG-----L-VYG--T..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
MM12H	-GCLG-----L-VYG--T..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
MM1A11	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
MM1320	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----N-----Q-----S-----IT-R-NKSETDR.WG.L-	124
MM1322L	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----N-----Q-----S-----IT-R-NKSETDR.WG.L-	124
MM239	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----N-----Q-----S-----IT-R-NKSETDR.WG.L-	124
CTL7202	-GCLG-----FL-VYG--T..L-----V-----T-----T-----G--s-la-----s-----N-----Q-----S-----IT-R-NKSETDR.WG.L-	124
MM142	-GCLG-----L-VYG--T..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
MM18101	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	27
SEGA	-GCLGI-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----V-----Q-----S-----IT-R-NKSETDR.WG.L-	124
SEGE	-GCLGI-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----V-----Q-----S-----IT-R-NKSETDR.WG.L-	124
SEJ3	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----N-----Q-----S-----IT-R-NKSETDR.WG.L-	124
7FBE2	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----N-----Q-----S-----IT-R-NKSETDR.WG.L-	124
7FLG1	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----N-----Q-----S-----IT-R-NKSETDR.WG.L-	124
7FLN1	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----N-----Q-----S-----IT-R-NKSETDR.WG.L-	124
7FSE1	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----N-----Q-----S-----IT-R-NKSETDR.WG.L-	124
MNE	-GCLG-----FL-VYG--T..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
MMELN11	-GCLG-----FL-VYG--T..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	26
MMESPL1	-GCLG-----FL-VYG--T..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	21
MMELIV2	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	26
PHYSE1	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
PHYLG1	-GCLG-----L-VYG--T..L-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
P209C19	-GCLG-----L-V-LK-V..-----V-----K-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
P209C15	-GCLG-----L-V-LG-V..-----V-----K-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	123
SIVSMBPBJB	-GCLG-----L-L-SG-V..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
SIVSMBPBJA	-GCLG-----L-L-SG-V..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
SIV6P6	-GCLG-----L-L-SG-V..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
SMBH9	-GCLG-----LS-x-SG-V..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
SMBH4	-GCLG-----L-V-VLE-C-V..-----V-----K-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
SMB2A	-GCLG-----L-V-VLE-C-V..-----V-----K-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
HUMB670	-GCLG-----L-L-SG-V..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	14
SMB670	-GCLG-----L-L-SG-V..-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
CONSENSUS-STM	-ACPG-----L-----T-----T-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124
STM	-ACPG-----L-----T-----T-----V-----T-----T-----G--s-la-----s-----E-----Q-----S-----IT-R-NKSETDR.WG.L-	124

HIV2/SIV ENV

CONSENSUS-A	t??7?L??7??7??7??7??...	224
ROD	...SKS...TSTTTTPTD...	224
NIRZ	...DT...QNTI...D...	214
885	...NN...TTTUNT...	147
ALI	...ESTT...SPSPGSLKP...	221
ISY	...ESAVA...TPSPSGP...	215
ST	...AKNT...STPTTTT...	218
SEN	...PHNR...SSSTSRPPT...	224
D194	...SGTTA...PSP...PNTI...	213
CAM2	...SDTTH...IR...TTTPT...	223
CAM38A	...SDTTH...IR...TTTPT...	225
D1024	...GN...ATTSTP...N...	221
D766	...GSTT...PA...AISM...	221
D608	...AGNT...TPT...AESN...	217
D868	...G.NTP...TST...ASV...	221
GH1	...NN...TTTGSTTGM...	215
CBL21	...TARP...TTGI...SNTESILK...	224
CBL22	...IAQSI...TT...AKAV...	220
CBL23	...DARNT...TPTTASPR...	220
CBL24	...AT-PEGBAQATSTTAAT...	225
NDS	...NA-AATTIRATTIAPR...	225
CAM1	...STGNT...TSTRAR...	223
CAM3	...QSNTA...TESPAR...	225
CAM4	...ATTAS...TTL...LNDITLL...	226
CAM5	...LSNK...TTS...STANLA...	218
CAM6	...GMY...TTI...SGTTAATT...	228
KR	...SEGSTVP...EIL...TS...	216
D1071	...IGSTT...TTD...RAMV...	117
UC2	...NA-AKPTAARPTTNP...	227
CONSENSUS-B	...TTT...t??S...11...	185
UC1	...TKPITTPITTTKPS...	221
D205	...STTTTTPITTSR...	224
EHQA	...ASKETTFSSASLSST...	223

CONSENSUS-SD	ks-...-ttaa-tt-tcas??-d-v-...-f-...-t-...-eq-q-is-k-...-k-...-e-...-a-l-q-g-...-est-...-q-...-i-...	218
MM251	KSS.....TTTAAAPTSAVUS..BKIDMV-...-s-aqn--t-...-eq-q-is-k-t-...-k-t-b-...-t-l-q-...-g-s-...-d-esr-...-q-...-ti-...	232
MM328	KSL.....TTTAPTAPTAAAK...IDMV-...-s-th--t-...-eq-q-ic-k-...-k-t-b-...-t-l-q-...-g-s-...-d-esr-...-i-q-...-ti-...	229
MM1A11	KSS.....TTTTAAPKAMSEK...INM-...-s-vh--t-...-eq-q-ic-k-...-k-e-...-a-l-q-...-g-s-...-g-esr-...-q-...-i-...	229
MM1362Q	KSI.....TTTASTTTSTTASA...KIDMV-...-s-aq--a-...-eq-q-is-k-...-n-e-...-a-l-q-...-g-n-...-g-esr-...-q-...-i-...	230
MM1322L	KSI.....TTTASTTTSTTASA...KIDMV-...-s-aq--a-...-eq-q-is-k-...-b-e-...-a-l-q-...-g-n-...-g-esr-...-q-...-i-...	230
MM239	KSI.....TTTASTTTSTTASA...KIDMV-...-s-aq--t-...-eq-q-is-k-...-k-e-...-a-l-q-...-g-n-...-g-esr-...-q-...-i-...	230
CTL7202	KSI.....TTTASTTTSTTASA...KIDMV-...-s-aq--t-...-eq-q-is-k-...-k-e-...-a-l-q-...-g-n-...-g-esr-...-q-...-i-...	230
MM142	KSS.....TTTASTTTTAKSVETADIV-...-vvh--t-...-eq-p-is-k-...-k-e-...-a-l-q-...-g-s-...-g-esr-...-q-c-d-...-i-c-...	233
MM18101	KSS.....TTTASTTTPTTASA...KIDMV-...-s-th--t-...-eq-q-is-k-...-k-e-...-t-l-q-...-g-s-...-d-esr-...-q-...-ti-...	133
SEGA	KSI.....PTTAPTASTTASA...KIDIV-...-s-aq--t-...-eq-q-s-k-...-k-e-...-a-l-q-...-g-np-...-g-esr-...-q-...-i-...	230
SEGE	KSI.....PTTASTTSTTASA...KIDMV-...-s-aq--t-...-eq-q-is-k-...-k-e-...-a-l-q-...-gtn-...-gesr-...-q-...-i-...	230
SEGV	KSI.....PTTAPTSTTASA...KIDMV-...-s-aq--t-...-eq-q-is-k-...-k-e-...-a-l-q-...-g-n-...-d-esr-...-q-...-i-...	230
7FBE2	KSI.....TTTAPTSTTASA...EVIDMV-...-s-aq--t-...-eq-q-is-k-...-k-e-...-a-l-q-...-g-n-...-g-esr-...-q-...-i-...	230
7FLQ1	KSI.....TTTAP...TTATA...SVIDMV-...-s-aq--t-...-eq-q-is-k-...-k-e-...-a-l-q-...-g-d-...-g-esr-...-q-...-i-...	237
7FLN1	KSI.....TTTAPTSTTASA...EVIDMV-...-s-aq--t-...-eq-q-is-k-...-k-e-...-a-l-q-...-g-n-...-g-esr-...-q-...-i-...	230
7FSE1	KSI.....TTTASTTTTASA...EVIDMV-...-s-aq--t-...-eq-q-is-k-...-k-e-...-a-l-q-...-g-n-...-g-esr-...-q-...-i-...	230
MNE	KSS.....TTTAPTAPTAKAE...AIKVV-...-s-nh--t-...-eq-p-is-k-...-k-rbe-...-a-l-q-...-g-s-...-b-esr-...-q-...-i-...	230
MNELN11	KSS.....TTTAPTSTTAKTE...AIKVV-...-s-nh--t-...-eq-p-is-k-...-k-rbe-...-a-l-q-...-gss-...-e-esr-...-q-...-vi-...	132
MNESPL1	KSS.....TTTAPTSTTAKAE...AIKVV-...-s-nh--t-...-eq-p-is-k-...-k-rbe-...-a-l-q-...-g-s-...-e-esr-...-q-...-i-...	127
MNELIV2	KSS.....TTTAPTSTTETE...AIKVV-...-s-nh--t-...-eq-p-is-k-...-k-rbe-...-a-l-q-...-g-s-...-b-esr-...-vq-...-i-...	132
PTTBE1	KSI.....TTTATP...TTASA...KIDMV-...-s-aq--t-...-eq-q-is-k-...-k-g-...-a-l-q-...-g-n-...-g-esr-...-q-...-i-...	228
PTTLQ1	KLI.....TTTAST...TTASA...KIDMV-...-s-aq--t-...-eq-q-is-k-...-k-e-...-a-l-q-...-g-n-...-g-esr-...-q-...-i-...	228
P209C19	RGA.....ETTTATTPAGTPSG...KEKV-...-g-d--rwns-a-...-eq-p-ig-k-...-k-rre-...-r-li-qsan-se-...-esk-q-...-i-...	232
P209C15	RGA.....ETTTATTPAGTPSG...KEKV-...-g-d--rwns-a-...-eq-p-ig-k-...-k-rre-...-r-li-qsan-sa-...-esk-q-...-mq-...-i-...	231
SIVSMMBPJB	GTP...APTITQTTTQASTTTPTSPITAKVV-DS..D--KIN--t-...-eq-p-s-k-...-k-re-...-r-li-q-ns-e-...-dsk-...-q-...-i-...	236
SIVSMMBPJA	GTP...APTITQTTTQASTTTPTSPITAKVV-DS..D--KIN--t-...-eq-p-s-k-...-k-re-...-r-li-q-ns-e-...-dsk-...-q-...-i-...	236
SIV6P6	GTP...APTITQTTTQASTTTPTSPITAKVV-DS..D--KIN--t-...-eq-p-s-k-...-k-re-...-r-li-q-ns-e-...-dsk-...-q-...-i-...	236
SMME9	G.....KPAITTTQSTTTTSPPI...IAKVV-DS..D--RSN--t-...-eq-p-s-k-...-k-re-...-r-li-qns-e-...-dsk-...-q-...-i-x-...	233
SMNH4	GNA.....GTTTTAITTTATPS...VAENV-...-s-nh--kmns-a-...-eq-p-ig-k-...-n-e-...-r-li-q-s-a-es-...-esk-n-...-i-...	231
SM63A	GNA.....RTAKSTTTTTVAPSVTEMV-...-s-nh--kmns-a-...-eq-p-ig-k-...-k-iz-...-r-li-q-s-a-...-esek-n-...-i-...	232
MMMB670	GNA.....ITTTATTAAPSAKTPKUTANVV-...-s-vknn--t-...-eq-p-s-k-...-k-e-...-...-...-...-...-...-...-...-...-...-...-...-...	76
MMMB670	GNV.....PTTTPPTATPSKE...TAMVV-...-s-vknn--t-...-eq-p-s-k-...-k-e-...-...-...-...-...-...-...-...-...-...-...-...-...	229
CONSENSUS-STM	GKT.....VTTVTPTAAAAAT...KPELV-...-s-vknn--t-...-e-sl-q-k-...-k-re-...-s-li-q-...-v-...-gesr-...-r-...-q-...-l-...	229
STM	GKT.....VTTVTPTAAAAAT...KPELV-...-s-vknn--t-...-e-sl-q-k-...-k-re-...-s-li-q-...-v-...-gesr-...-r-...-q-...-l-...	229

CONSENSUS-A

II-B-40
DEC 97

CONSENSUS-SD	Y	M	K	VSS	M	R	L	V	I	D	K	G	K	KD	IK	345				
MM251	Y	M	R	VSS	M	R	L	V	I	D	K	G	K	KD	IK	359				
MM32N	Y	M	R	VSS	M	R	L	V	I	D	K	G	N	KD	IK	356				
MM1A11	Y	M	K	VSS	M	R	L	V	I	D	K	R	K	KD	IK	356				
MM316ZQ	Y	M	R	VSS	M	R	L	V	I	D	K	G	K	KD	IK	357				
MM1322L	Y	M	R	VSS	M	R	L	V	I	D	K	R	G	K	KD	IK	357			
MM239	Y	M	R	VSS	M	R	L	V	I	D	K	G	K	KD	IK	357				
CTL7202	Y	M	R	VSS	M	R	L	V	I	D	K	R	X	KD	IK	357				
MM142	Y	M	R	VSS	M	R	L	V	I	D	K	R	G	N	K	IK	360			
MM18101	Y	M	R	VSS	M	R	L	V	I	D	K	G	N	KD	IK	260				
SEGA	Y	M	R	VSS	M	R	L	V	I	D	K	E	K	KD	IK	357				
SEGB	Y	M	R	VSS	M	R	L	V	I	D	K	E	K	KD	IK	357				
SEGC	Y	I	M	E	VSS	M	R	L	V	I	D	K	R	G	K	KD	IK	358		
7PBE2	Y	S	D	M	K	VSS	S	M	R	L	V	I	D	K	R	G	K	KD	IK	355
7FLO1	Y	M	R	VSS	M	R	L	V	I	D	K	G	K	KD	IK	354				
7FLN1	Y	M	R	VSS	M	R	L	V	I	D	K	G	K	KD	IK	357				
7FSE1	Y	M	R	VSS	M	R	L	V	I	D	K	G	K	KD	IK	357				
MNE	Y	M	R	VSS	M	R	L	V	I	D	K	R	E	N	K	IK	357			
MMELN11	Y	M	R	VSS	M	R	L	V	I	D	K	R	E	N	K	IK	259			
MMESPL1	Y	M	R	VSS	M	R	L	V	I	D	K	R	E	N	K	IK	254			
MMELIV3	Y	M	R	VSS	M	R	L	V	I	D	K	R	E	N	K	IK	259			
PHTBE1	Y	M	R	VSS	M	R	L	V	I	D	K	G	K	KD	IK	355				
PHTLG1	Y	M	R	VSS	M	R	L	V	I	D	K	G	K	KD	IK	355				
P209C19	Y	S	A	K	VSS	K	S	M	R	L	V	I	D	K	R	G	N	S	I	359
P209C15	Y	S	A	K	VSS	K	S	M	R	L	V	I	D	K	R	G	N	S	I	358
SIVSMMPB3B	Y	S	A	T	VSS	S	M	R	L	V	I	D	K	G	K	KD	IK	363		
SIVSMMPB3A	Y	S	A	T	VSS	S	M	R	L	V	I	D	K	G	K	KD	IK	363		
SIV6P6	Y	S	A	T	VSS	S	M	R	L	V	I	D	K	G	K	KD	IK	363		
SMH9	Y	S	A	T	VSS	S	M	R	L	V	I	D	K	G	K	KD	IK	360		
SMH4	Y	S	L	A	VSS	K	S	M	R	L	V	I	D	K	R	G	N	S	I	358
SM62A	Y	S	A	K	VSS	K	S	M	R	L	V	I	D	K	R	G	N	S	I	359
HUNB670																			76	
SMNB670	Y	A	VSS	A	B	P	R	R	I	V	I	D	K	R	G	N	S	I	356	
CONSENSUS-STM	Y	T	A	VSS	M	R	L	V	I	D	K	R	G	N	S	I	356			
STM	Y	T	A	VSS	M	R	L	V	I	D	K	R	G	N	S	I	356			

HIV2/SIV ENV

CONSENSUS-A	kgTl?kHPRYkg...tndt?777inf?agpgkgSUpEvayNMWnC...rGE...ELYCNWtw/LmwVE...nrt?n?C...?hNYvpChIkQilNcWkVgknvYLPPrAGeLcCnStVTSiiANid	414
ROD	-E--A-----R-----R--M-S-A-----I-----R-----HR--A-----R-----S-----	456
NINZ	-E--A-----RSRTEN-K-K--R-----T-----S-----GGR-----QR--A--R-R-----R--L-----	451
B85	-I-----K--K-Y-TK-----I-----HS--A-----L-S-----V-----	379
ALI	-E--V-----N--Q--TK--R--A-EV-----R--GGE-----Q--A-----L-----	456
ISY	-V-----N--R--T--D-----EV-----K--GQ-----Q-----E-----S-E-----	450
ST	-L--A-----E--K-R-I--ER-----Q-----Q-----Q-----	453
BEN	-VQ-----I--G--K--TK--A--P-----DK--Q-----RR-----A-E-----	456
D194	-A--G-----G--K--TK--I--T-----K--Q-----HQ--A--R-----T-----	449
CAN2	-AE-----KNIT--D-T-K--ER-----T--S-----F-----KP--T-----KR--A--R-----	459
CAN3BR	-AE-----KNIT--D-T-K--ER-----T--S-----F-----KP--T-----KR--A--R-----	461
D1024	-V-----E--K-E--D-H-A-----A-----V--I-----N-K-----W-----H-----	457
D765	-V-----R-----E-R--K--T-----A-----I-----V-R-----W-----H-----	457
D808	-E--T-----R-----G--K--A-----A-----K-A-Q-----W-----H-----	453
D868	R--IAE-----R--N-G--M-K-V-----A-----IAE-----I--K-E-A-----W-----H-----	457
GM1	-I-----X--M--TK--R-----P--Q-----Q--A--R-----Q-----	450
CSL21	-V-----R-----K-E--M-T-AG-E-----T--Q-----Q-----R-----L-S-----	460
CSL22	-A-----N-E--S-T-T--R-----P-----K--GQ-----QR-----S-----L-----	456
CSL23	-E--A-----E--M--T-----I-----K-----HR--R-----	453
CSL24	-A-----S--K--IG-----F-----T-----W-----S-----	460
MDS	-E-II-----ENIT--M-T-T-----DKP-----W-----R-----V-----	470
CAM1	-V-----R-----G--M--T-----F-----K--GE-----K-----	457
CAM3	-E--A-----K--K--A--R-----E-R-----W-----R-----	461
CAM4	-LE-----ATNEN--N-RLAE-----GQP-----Q-----E-----V-----L-----	461
CAM5	-VV-----R-----EPNKT--V--I-----F-----K--GQE-----Q--A--A--E-R-----I-----E	452
CAM6	-KE-----K--F-V-N--N-T-E--R-----Q-----QR-----R-----R-----M	463
KR	-V-----R-----Q--K--TQ-----A--V-----R--I-----A--KP-----QR--A--R-----R--Q-I-----V-----	451
D1071	-A-----N-G--K--T--S--A--I-----S-----D--A--R-----	323
UC2	-AG-----S--K--VK--V-----T-----F-----SCK-----QR--A--R-----QY-----	463
UC3	-E-IAD-----T-----KNIS--Q--LTENG5-----RP-----G.I-----ER--A--V-----S--S--L-----	178
UC5	-A-----R--N--IK--R-----QE-----K--A-----M-----	178
UC6	-E-IA-----T-----E-IT--K-KLAENG5V-----R--M-----L-----DP-----ER-----V-----M--T--N--L-----	177
UC7	-A-----G--K--TK-----F-----I-----RP--Q-----Q-----	178
UC9	-E-ITS-----INQIT--K-RLAENG5V-----R-----D-A--T-----OK-----V-----H-----A--I--N--L-----	178
UC9	-E-IM-----T--S--LTENSP-----R-----Q-P--TI-----KK--A--V-----R-----V-----L-----	179
UC10	-V-----N--K--K--x-----AE-----E-----DQxK-----PR-----R-----R-----S--T--M--L-----	179
UC11	-II-----H-----IG--R--T-----I-----Q-----Q-----A-----R-----Q-----	178
UC12	-AE-----X--A--T-----S-----I-----Q-----QX-----X-----S-----	178
UC14	-VG-----E--F--AK-----G--RP-----QR-----	179
CONSENSUS-B	-E-II-----?--?Ni-----rl-eh7?-----F-----?--?C-----?rY--?--Y--?--?--?--?--S-----l-----?	404
UC1	-E-II-----N-E--R-RLVG-SA-----RH-----F-----GT-----CK--T-----V-----Y-----T--S--S--L-----	458
D205	-R-II-----G-AKNITPVKL.VSEN-----TT-----K--T-----RR--A--R--R-----Y-----S-----L-----M	461
EMOA	-E-IKN-----S--TNIS--Q-RLAENGARS-----R-----F-----G-----LKR--AS--R--V-----I--R-----S--L-----L-----	456
GM2	-----N-----K--KS-----RR-----R-----R--I-----L-S-----M	66

CONSENSUS-SD	--IV--T--N-D..K-LT--G--TF--K-N--DR-7-7n...7777qsp?eqhkr--T--D--L--	455
MM251	--IV--T--N-D..K-LT--G--TF--K-N--DR-DV-T...QRPKERHRR--R--D--L--	473
MM32R	--IV--T--N-D..K-LT--G--TF--K-N--DR-DV-N...QRPKERHRR--R--D--L--	470
MM1A11	--IV--T--N-D..K-LT--G--TF--K-N--DR--TAN...QKPKQKHRR--R--D--L--	470
MM316ZQ	--IV--T--N-D..K-LT--G--N-ATF--K-N--DR--TAN...QKPKQKHRR--R--D--L--	471
MM131ZL	--IV--T--N-D..K-LT--G--N-ATF--K-N--DR--TAN...QKPKQKHRR--R--D--L--	471
MM239	--IV--T--N-D..K-LT--G--TF--K-N--DR--TAN...QKPKQKHRR--R--D--L--	471
CTLT202	--IV--T--N-D..K-LT--G--TF--K-N--DR--TAN...QKPKQKHRR--R--D--L--	470
MM142	--IV--T--N-D..K-LT--G--TF--K-N--DR-SL-T...QRPKERHRR--R--D--L--N	474
MM18101	--IV--T--N-D..K-LT--G--TF--K-N--DR--T-N...QKPKQKHRR--R--D--L--	346
SEGA	--IV--T--N-D..K-LT--G--VF--K-N--DR--T-N...QKPKQKHRR--R--D--L--	471
SEGE	--IV--T--I-MID..K-LT--G--TF--K-N--DR--T-N...QKPKQKHRR--R--D--L--	471
SEGF	--IV--T--N-D..K-LT--G--TF--K-N--DR--T-N...QKPKQKHRR--R--D--L--	472
7FBE2	--IV--T--N-D..K-LT--G--TF--KIN--KMR--TAN...QKPKQKHRR--R--D--L--	463
7FLG1	--IV--T--N-D..K-LT--G--TF--K-N--DR--TAN...QKPKQKHRR--R--H--D--L--D--	468
7FLN1	--IV--T--N-D..K-LT--G--TF--K-N--DR--TAN...QKPKQKHRR--R--D--L--	471
7FSE1	--IV--T--N-D..K-LT--G--TF--K-N--GR--TAN...QKPKQKHRR--R--D--L--	466
MNE	--IV--T--N-D..K-LT--G--TF--K-N--DR--L-G...TTQKPKQKHRR--R--R--D--L--	473
MNELN11	--IV--T--N-D..K-LT--G--TF--K-N--DR--L-G...TTQKPKQKHRR--R--R--D--L--	375
MNESPL1	--IVN--T--N-D..K-LT--G--TF--K-N--DR--L-G...TTQKPKQKHRR--R--R--D--L--N	370
MNELIV2	--IV--T--N-D..K-LT--G--TF--K-N--DR--L-G...TTQKPKQKHRR--R--R--D--L--	375
PMTSE1	--IV--T--N-D..K-LT--G--TF--K-N--DR--TAN...QKPKQKHRR--R--D--L--	469
PMTLS1	--IV--T--N-D..K-LT--G--N-ATF--K-N--DR--TAN...QKPKQKHRR--R--D--L--	464
P209C19	--E-V--T--K..K-LT--G--TF--K-S--DRDQSS...RWKQKQKPEQKRR--R--D--L--E--	479
P209C15	--E-V--T--K..K-LT--G--TF--L-K-N--DRDQSS...RWKQKQKPEQKRR--R--D--L--E--	477
SIVSMMPEJB	--E-V--T--K-E..Q-KLT--G--TF--K-N--DIQ-GS...RWTSQKQKPEQKRR--R--D--L--E--	481
SIVSMMPEJA	--E-V--T--K-E..Q-KLT--G--TF--K-N--NIQ-GS...RWTSQKQKPEQKRR--R--D--L--E--	481
SIV6P6	--E-V--T--K-E..Q-KLT--G--TF--K-N--DIQ-GS...RWTSQKQKPEQKRR--R--D--L--E--	481
SMH9	--XX-V--T--K-E..Q-KLT--G--TF--K-N--KIK-GS...RWTSQKQKPEQKRR--R--D--L--E--	478
SMH4	--E-V--T--K..K-LT--G--TF--K-N--DRDQSS...RWKQKQKPEQKRR--R--D--L--E--	477
SM62A	--E-V--T--K..K-LT--AR--TF--K-N--DRDQSS...RWKQKQKPEQKRR--R--D--L--E--	478
HMB670		76
SM670	--E-IV--T--AE..R-LT--G--TF--K-N--DRNINGS...RWKQKQKPEQKRR--R--D--L--	475
CONSENSUS-STM	--E-V--T--A..K-RIV--G--TF--K-N--7...ST?EM?WNN?KE?Q-?--R-V--?--?--D--	455
STM	--E-V--T--A..K-RIV--G--TF--K-N--I...STSEMDNNNNKKEQKRR--R-V--Q--D--	475
BTMAK3	K-N-----STLEMGWNNNNKKEQKRR--R-V--R--D--	86

HIV2/SIV ENV

	CONSENSUS-A	533
ROD	WQNN--	584
NIMZ	AGD--	575
885	VFGN-G-	426
ALI	TDG--	561
15Y	VDGD-R-	576
ST	GGE--	578
BEW	IDK-R-HT	583
D194	SDG--	574
CAN2	ERD--	585
CAN3BR	ERD--	587
D1024	GG--	507
D766	GG--	508
D808	GG--	504
D868	GG--	509
GH1	VNS--	575
CBL21	TYG--	585
CBL22	VNG--	581
CBL23	ANG--	579
CBL24	VKN--	585
MD8	DKD--	595
CAN1	TDM-G-ET	584
CAN3	VEA--	586
CAN4	MR-DS-T	586
CAN5	TDMGKT--	579
CAN6	VD--	588
KR	MFD--	576
UC2	TDG--	588
5132E2		29
FA		29
GB1222		73
CI171		29
CI9		29
IS053		29
VI390		29
VI905		29
VI1056		29
VI1415		29
VI495		29
VI884		29
SL93FE		20
CONSENSUS-B	--sd--?	523
UC1	VYYDG-D-ET	587
D205	SD-S-T	586
EHQA	WIDK-L-	582
GE2	SAN-A-	191
JA		29
OR		29
VI1011		29
CONSENSUS-C		29
2238		29

HIV2/SIV ENV

CONSENSUS-D
FD784

-----L-----T----- 29
-----L-----T----- 29

CONSENSUS-SD	.widg---.---M-----1---V---LTGGTS-N-----L-----T-----	581
MM251	.WTDG---.S-M-----L---DV---TTGGTS-N-----T-----L-----T-----	600
MM32H	.WTDG---.---M-----L---DV---TTGGTS-N-----T-----L-----T-----	597
MM1A11	.WTDG---.---M-----L---NVR---TTGGTS-N-----L-----L-----T-----	597
MM31620	.WTDG---.---M-----L---DV---TTGGTS-N-----T-----L-----T-----	598
MM1323L	.WTDG---.---M-----L---DV---TTGGTS-N-----T-----L-----T-----	598
MM239	.WTDG---.---M-----L---DV---TTGGTS-N-----T-----L-----T-----	598
CTL7202	.WTDG---.---M-----L---DV---TTGGTS-N-----T-----L-----T-----	524
MM142	.WTDG---.S-M-----L---NV---TTGGTS-N-----VT-----L-----T-S---	601
SEGA	.WTDG---.---SM-----	489
SEGE	.WTDG---.---SM-----	489
SEQ3	.WTDG---.---M-----	490
7FBE2	.SIDG---.---M-----K-----L---NV---ATGGTS-N-----	518
7FLG1	.WTDG---.---SM-----L---DV---TTGGTS-N-----	524
7FLM1	.WTDG---.---M-----L---NV---ATGGTS-N-----	527
7FSE1	.WTDG---.---M-----L---NV---TTGGTS-N-----	522
MNE	.WTDG---.---M-----L---NV---TTGGTS-N-----T-----L-----T-----	600
MNELNL1	.WTDG---.---X-----	385
MNESPL1	.WTDG---.---M-----	389
MNELIV2	.WTDG---X-----	382
PHTBE1	.WTDG---.---M-----L---NV---TTGGTS-N-----	525
PHTLQ1	.WTDG---.---K-----L---SV---ATGGTS-N-----	520
P209C19	.WTDG---.---M-----L---DVR---TTGGTS-N-----V-----L-----T-----	605
P209C15	.WTDG---.---M-----L---SVR---TTGGTS-N-----V-----L-----D-----T-----	604
SIVSMNPBJB	.WTDG---.---N-----A---SV---TTGGTS-N-----V-----L-----A-----T-----	608
SIVSMNPBJA	.WTDG---.---M-----A---SV---TTGGTS-N-----S---V-----L-----A-----T-----	608
SIV6P6	.WTDG---.---M-----SV---TTGGTS-N-----V-----L-----A-----T-----	608
SMMH9	.WTDG---.---M-----SV---TTGGTS-N-----L-----L-----T-----	605
SMMH4	.WTDG---.---M-----I-----L---SVR---TTGGTS-N-----V-----L-----T-----	604
SM62A	.WTDG---.---M-----L---SVR---TTGGTS-N-----V-----L-----T-----	605
SMMH670	.WTDG---.---M-----L---NVR---TTGGTS-N-----L-----L-----T-----	602

CONSENSUS-STW	.WTDG---.---A-----L---NV---TT,STG-T-----T-----T-----T-----	589
STM	.WTDG---.---A-----L---NV---TT,STG-T-----T-----T-----T-----	601
STNAK3	.WTDG---.---A-----L---NV---TT,STG-T-----T-----T-----T-----	212

CONSENSUS-A	-----	29
BA_7312A	-----	29

CONSENSUS-U	-----	29
SMC12	-----	29
SMC18	-----	29

HIV2/SIV ENV

CONSENSUS-A	Y L K Q A Q I N S K G C A F R Q V C H T T V P W V N D S I T F D M N N H T W Q E W S Q V R Y L S A N I S ? S L E Q A Q I Q Q E K M Y E L Q K I N S W D V E G N W E D I L S W I K Y I Q Y G V Y I V V S I I A L R I V I Y V V Q M I S R I R K G Y R P V F S S P	662
ROD	Q R A D K K I A V L I A V	714
NIH2	S T K T L P V I V I	705
ALI	R N K D E Q T T L A V I V	711
ISY	A R T E H K I F E M V	706
ST	T R I N B V	708
BEN	H S K K Q E I L V H A L F	713
D194	K R H Q Q G A I L	704
CAM2	A E K Q E L N T S V	715
CAN2BR	A E K Q E L N T S V	717
GM1	S K Q Q V V A I L	705
CBL21	T Q R I T R I F P E	715
CBL22	R D I Q Q F I V V	711
CBL23	T E Q K I D Q Q I I V V I	708
CBL24	R Q H Q E L A V I I V	715
WDS	S E E K Q E Q I V	725
CAM1	L A K Q R V I	714
CAM3	K H E N I R V A I F	716
CAM4	V N L R T D Q L S R V V I I	716
CAM5	R Y K I D D S R F I I V V	709
CAM6	K Q V C A V	718
KR	R R Y L E N I V T D R V I I L	706
UC2	T Y L E N I V R K Q E A R L F	718
5132B2	D E R Q A V A I	150
FA	Q I E V V I L R	150
GB1222	H V K V E R L V H V V V	203
CI171	L K M Q S S F V V R	149
CI9	A E E I L I V V A I L	151
ISO53	Y T Q N I E Q A V V A G R	151
VI390	A K F Q Q A V A F	151
VI905	T N Q Q I A S I Q C L I G R	150
VI1056	T V K D Q I E R I G R	151
VI1415	T K D Q I E R	151
VI495	T Q Q I T G C L I G R	151
VI694	I I N D T L I V V I F F	151
SL93FE	A R P E T N K E K D F T D T D L I R I P D S T W L V V L V R L X	135
CONSENSUS-B	l D E T ? Q ? ? D f d T L E Q f M A Y L L V A I V I ? A	648
UC1	L F E T E Q K R M F D T A L E R F M A H L L V A L V I M A	717
D205	E E T N Q K H F D T A L E I H L L V A L V V V I A	716
ENOA	E K Q R F D T K L E Q I S F M A R L L I V V A I I A	712
GM2	P T K Q Q K M F T A L E H L L V A L V V I A	321
JA	L P E T N N Q Q K N D T Q L E Q F M A K L L V A L V I I A K	150
ON	L P E T I E Q A K Q F D T Q L E Q S F M A H L R L L V A V I M A	150
VI1011	L P E T H L R Q R E M D T R L E Q F M A R L L L V I I A K	151
CONSENSUS-C	K M Q E Q Q K I A F D T E L L V Y L F A V V A M	151
2238	K M Q E Q Q K I A F D T E L L V Y L F A V V A M	151
CONSENSUS-D	E P S T R D F T Q L E N I F L I A R S	150
F0784	E P S T R D F T Q L E N I F L I A R S	150

HIV2/SIV ENV

CONSENSUS-SD	-----P-?-?-?-?-----Lk-df-----T-l-E-----?-----l-l-V-g-----A-Q-----	707
MM251	-----A-----P-A-----D-----RK-DF-E-TAL-E-----A-----I-V-V-L-----I-AR-Q-----	730
MM32H	-----A-----P-A-----D-----RK-DF-E-TAL-E-----A-----V-L-----I-AR-Q-----	727
MM1A11	-----A-----P-A-----K-E-----RK-DF-E-TAL-E-----A-----V-L-----I-AR-Q-----	727
MM3162Q	-----A-----S-P-A-----K-E-----S-RK-DF-E-TAL-E-----S-----A-----V-L-----I-AR-Q-----	727
MM1322L	-----A-----P-A-----K-E-----RK-DF-E-TAL-E-----A-----V-L-----I-AR-Q-----	728
MM239	-----A-----P-A-----K-E-----RK-DF-E-TAL-E-----A-----V-L-----I-AR-Q-----	728
MM142	-----A-----P-A-----E-----RK-DF-E-TAL-E-----I-I-V-L-----I-AR-Q-----	731
MNE	-----A-----P-AN-N-E-----RK-DF-E-TAL-E-----A-R-----V-L-----I-AR-Q-----	730
P209C19	-----P-----V-I-D-----Q-R-DS-----TQL-E-----F-----I-----R-----L-L-VVG-----A-Q-----	735
P209C15	-----P-----V-I-D-----Q-R-DS-----TQL-E-----F-----I-----R-----L-L-VVG-----A-Q-----	734
SIVSMNPBJR	-----P-----T-N-----K-NF-----TQL-E-----T-----I-----L-L-V-G-----A-Q-----	738
SIVSMNPBJA	-----P-----T-N-----K-NF-----TQL-E-----T-----I-----L-L-V-G-----A-Q-----	738
SIV6P6	-----P-----T-N-----K-NF-----TQL-E-----T-----I-----L-L-V-G-----A-Q-----	738
SMNH9	-----XX-----X-----P-T-X-X-X-X-NF-----TXX-E-----X-----XX-XX-X-X-----L-L-V-G-----A-Q-----	735
SMNH4	-----P-----ET-V-N-----R-DF-----TQL-E-----I-----R-----L-L-V-G-----A-Q-----	734
SM62A	-----P-----ET-V-N-S-----R-DF-----TQL-E-----I-----R-----L-L-VVG-----A-Q-----	735
SM6B670	-----P-----T-V-----RK-DS-----TQH-E-V-----I-----R-----L-L-VVG-----A-Q-----	732
CONSENSUS-STM	-----P-----V-----RK-DF-----TQL-E-V-----?-VR-----L-I-?VN-VA-??-?-A-----	714
STM	-----P-----V-----RK-DF-----TQL-E-V-----VR-----L-I-LVN-VA-?M-L-A-----	731
STRAK3	-----P-----V-----RK-DF-----TQL-E-V-----A-VR-----L-I-VN-VA-----A-----	342
CONSENSUS-A	-----D-----Q-K-I-D-----E-----A-V-----V-VI-----IQ--R-----	150
BA_7312A	-----D-----Q-K-I-D-----E-----A-V-----V-VI-----IG--R-----	150
CONSENSUS-U	-----R-?-?-?-?-?-Z-?T-Q-----IKI-D-----E?-?-?-?-?-?-?-F-V-YI-F-V-----G-?-V-LI-L-CK?--	143
SMC12	-----R-A-----E-NY-Q-----IKI-D-----E-E-----R-F-V-YI-F-V-----G-IV-LI-L-CK--	150
SMC18	-----R-----E-T-Q-----IKI-D-----EK-N-----F-V-YI-F-V-----G-V-LI-L-CK--	150

	/ early stop	<- tat cda	* nsf start -> *	<- rev cda	*
CONSENSUS-A	PgYIQQIHhKd7eQpaaeeTeeDvg7n7Gd7PwPwplayihfllrlrlrltgiynlCrdLlSr7F7tllqp17qPl7777c77Rdw)Rl7AAYLQYQcchIQaaqfaar7tREtIagaw7lwg7l7r				772
ROD	RG--N--Q-S-Q-RY--Q--R-S--S-L-L-Y-N--RT-F--A--A--C-G--RV-E-				837
NRH2	Q--T-G--S-Q-RS--L--SIPI--P-S-QRAL-AI--K--L--T--G-D--RA-Q-				835
ALI	Q--T-G--D-V-RS--L--H-A-I--S--I-SI--P-S-QRAL-TI--K--L--TWT--D--AMQW				841
TSY	N--D--NDV-SRS--E--R-S--YL-L-P--K--L--V--TS-G-B--A-C-				825
ST	A-F--R--NSV--NM--R--Q--NR--S-Q--L-S-S-RRAL-AV--FNT--Q--R--A-G--TN--GF-T-QQ				838
BEN	L--L--RG--M-G--GDS-YDL--N-VQ--K-T--I--KNSP-RRL-S-S--AI--K-Q--T--GM--EARR-				839
D194	L--T-RG--N--A-DDS-PGL--LN-Q--H-T--S-G--KNSP-RRL-S-S--AI--K--R--TA--I--G-CEAAQ-				830
CAM2	P--S--G-G-RS--Q-E--Q--N--I-Q--L-N--PKV-F--A--AA--C-DV--M-Q-				838
CAM2BR	P--S--G-G-RS--Q-E--Q--N--I-Q--L-N--PKV-F--A--AA--C-DV--M-Q-				822
GH1	L--T-RG--N-G--DRDD-YDLF--N--H-T--K--TNSP-RRL-S-N--AI--K--G--KT--S--GG-CAAVQ-				830
CBL21	A-L-V--HRG-PS--DSV-SS--S--I--T-I-KS-Q--S-G-QRAL-AI--FG--L--A--TSV-MFC-TMQQ				844
CBL22	N--V--W-D-G--D-A-SSS--S--R--Q-V--W-SS--S-LI--A--K--R--KP--D--C-G-				829
CBL23	L--W--N-V--SLF--R--HQ--N--ISL-R-VF-E-QRAL-AI--TD--G--A--T-D--A-Q-				837
CBL24	L--RHW--D-G--D-V-TSL--I--SSLI--L-S-QRAL-AI--E--L--LT-A--N--A-Q-				845
HDE	R--D--G--G-RS--Q-E--Q--I--L--F-N--Q--RI-F--V--TS-C-G--T-Q-				848
CAM1	L--L--S-V-RF--R--S--I-L--L-N--AST-F--Q--AAK--TS-C-S--RT-G-				837
CAM3	Q--D--I-G-RS--T--R--R--S--S-L-F-N--RT-F--RQ--A--TS-C-G--RT-DN				839
CAM4	L--T-K--Q--D-V-SL--L--R--S--S-P--RA--L-AI--K--TA--N--EAWQ-				839
CAM5	L--W--RG--G-M-A-NS--Q-R--Q--S--N--KS-L-L--NT-F--A--TS-C-S--T-G-				828
CAM6	L--N-G--N-Q-YRS--Q--R--S--S-L-L-N--RT-F--FA--B-C-S--T-Q-				841
XR	W--D--D-A-NSI--SS--Q--Q--SV-K--S-F--L-F-E-QRAL-TI--TI--VL-VL--T--S--D--ANG-				836
UC2	L--T-RG--N--G-A-DAS-YDF--N--QL--H-T--S--KNSP-RRL-S-N--AI--K--I--TA--G-CKAVQ-				844
GB1222	Q--T--S-V-RF--S--I--S--N--I-L--F-N--P--RT-F--A--A--C-G--T-G-				326
CONSENSUS-B	S-T--P-R--RG--N--pQ-nQp-YRS--Q-E--7-Q-Rh--IN--g--t--LKT-Q?--Q--P--pV--Is-F--7?A-a?--e-g?c--aa-R-				760
UC1	S-T--P-R--HRG--N--DEG-NEGAYRS--Q-E-A--Q--NN--IN--G--N--LKTQI--A--Q--P--SL--IS-F--I--AT-AA--MTQ-A--YA-R-				836
D205	S-T--P-R--RG--N--QG-NDDYRS--Q-E--L-Q--NN--IN--G--T--LKT-QI--HQ--STM-Q--P--PV--IS-F--LA-A--A-G--S-GET--EA-R-				838
EHQA	S-T--P-R--RG--N--GG-N-EYRS--Q-E--P--Q--ND--IN--SG--T--KT-Q--V--Q--P--PP--R--IS-F--I--A--AAG--S-A-TS--V-R-				831
GH2	S-T--P-R--RG--N--GG-EDQ-YRS--Q-E--L-QPRN--IN--S--BT--LKT-R--Q--P--LV--ID-P--LR-A--V--S-GETF--EI-R-				460
CONSENSUS-SD	A-V--7p-Q7gq-l-Tk-Ge-g-G--rG-nrs--Q-E--Q--N-FSN--t?l--vQI--vl-r-s-tLQR--EV?2ait--msyF--v--7w--a?--s--qD--Et-g-				825
MM251	S-F-ST-TQQ-PAL-T--GK-G-G-EGG-NSS--Q-E--Q--N-FSN--T--AVQI--L-R-SATLARI-EV--TELT--MSYFH--V--GN-SAT--D--ET-R-				859
MM32H	S-F-ST-TQQ-PAL-T--GK-G-G-EGG-NSS--Q-E--Q--N-FSN--T--AVQI--L-R-SATLARI-EV--TELT--MSYFH--V--VQM-SAT--GD--ET-R-				857
MM1A11	S-F-ST-TQQ-PAL-T--GK-G-G-EGG-NSS--Q-E--Q--N-FSN--T--VVOI--L-R-SATLARI-EV--TELT--MSYFH--V--VQM-SAT--GD--EA-R-				855
MM3162Q	S-F-ST-TQQ-PAL-T--GK-G-G-EGG-NSS--Q-E--Q--N-FSN--T--VVOI--L-R-SATLARI-EV--TELT--MSYFH--V--VW-SAT--GD--ET-R-				856
MM1322L	S-F-ST-TQQ-PAL-T--GK-G-G-EGG-NSS--Q-E--Q--N-FSN--T--VVOI--L-R-SATLARI-EV--TELT--MSYFH--V--VW-SAT--GV--ET-R-				858
MM239	S-F-ST-TQQ-PAL-T--GK-R-G-EGG-NSS--Q-E--Q--N-FSN--T--VVOI--L-R-SATLARI-EV--TELT--MSYFH--V--VW-SAT--GD--ET-R-				858
MM142	S-F-ST-TQQ-PAL-TK-GKGG-G-GSG-NSS--Q-E--Q--N-FSN--T--AVQI--P-R-SATLARI-EV--ELT--MSYFH--V--AQ-SAT--GE--EA-Q-				860
MM5	S-F-ST-TQQ-PAL-TK-G-G-G-EGG-NSS--Q-E--Q--N-FSN--T--AVQI--F-R-STTLORI-EV--TELT--MSYFH--V--VW-SAT--GD--ET-G-				859
P209C19	A-V--P--GQ-P-TK-GE-G-EGG-NSS--Q-E--Q--N-FSS--W-L-SYQI--VL-R-STTLORI-EVI--IEIT--MSYF--V--VW-SAT--S--DI--ET-G-				865
P209C15	A-V--PV--GQ-P-TK-GE-G-EGG-NSS--Q-E--Q--N-FSS--W-L-SYQI--VL-R-STTLORI-EVI--IEIT--MSYF--V--VW-SAT--S--DI--ET-G-				864
SIVSMHPEJB	A-V--VP--QTGG-L-TK-GE-G-G-GRG-NRS--Q-E--Q--N-FSS--W-L-NCOQ--VL-S-SRTLORI-EVI--VOI--MRYL--A--VW-KFY--S--D--ET-G-				868
SIVSMHPEJA	A-V--P--QTGG-L-TK-GE-G-G-GRG-NRS--Q-E--Q--N-FSS--W-L-NCOQ--VL-S-SRTLORI-EVI--VOI--MRYL--A--VW-KFY--S--D--ET-G-				868
SIV6P6	A-V--P--QTGG-L-TK-GE-G-G-GRG-NRS--Q-E--Q--N-FSS--W-L-NCOQ--VL-S-SRTLORI-EVI--VOI--MRYL--A--VW-KFY--S--D--ET-G-				869
SNH39	AXVX--P--XTGG-L-TK-GE-G-G-GRG-NRS--Q-E--Q--N-FSS--W-L-NCOQ--VL-S-SRTLORI-EVI--VEIT--MRYL--A--VW-KFY--S--D--ET-G-				865
SNH44	A-V--P--GQ-P-TK-GE-G-G-GRG-NRS--Q-E--Q--N-FSS--W-L-SYQI--VL-S-STTLORI-EVI--IEIT--MSYF--V--VW-KFA--S--DI--ET-G-				864
SN62A	A-V--P--GQ-P-TK-GE-G-GRG-NRS--Q-E--Q--N-FSS--W-L-SYQI--VL-S-STTLORI-EVI--IEIT--MSYF--V--VW-KFA--S--DI--ET-G-				865
SNB670	A-V--P--GQ-P-TK-GE-G-G-GRG-NRS--Q-E--Q--N-FSN--TC-L-SYQI--VF-R-FSILORI-EV-TEIS--R-F--V--GN-ANG--S--GD--ET-G-				862
CONSENSUS-STM	S7?--2--GQ--7TK-G--GG-DRG-INS--Q-E-T--Q-V--W--NF-AC--IYQ--H-TF-RISRILORI-EVV--G--I--A--AW-ANG--S-G-D--ET-G-				940
STM	SCR--P--GQ--TK-G--GG-DRG-INS--Q-E-T--Q-V--W--NF-AC--IYQ--H-TF-RISRILORI-EVV--G--I--A--AW-ANG--S-G-D--ET-G-				961
STMH3	S-H--L--GQ--LTK-G--G				366

HIV2/SIV ENV

CONSENSUS-A	IGRGILavPRIRQGAEIALLS	793
RDP	-----L---	858
MLKZ	-----L---	856
ALI	V--R-----	882
ISV	-----	846
ST	-----	859
BEN	-----L---	860
D194	-----	851
CAN2	-----L---	859
CAN2BR	-----V-----L---	843
GM1	V-----	851
CBL21	-----I-----L---	865
CBL22	-----E-----	850
CBL23	-----	858
CBL24	-----L---	866
NDS	-----	869
CAN1	-----L---	856
CAN3	F-----SI-----	860
CAN4	L-----L---	860
CAN5	-----L---	849
CAN6	-----I-----	862
KR	-----L---	857
UC2	-----	865
GB1222	-----T---	247
CONSENSUS-B	aLga-i-I-----l-L---	781
UC1	TAER-I-I-----L-L---	857
D205	AA-A-I-I-----L-LT--	859
EMDA	AAAE-I-I-----L---	852
GH2	MAGH-V-I-----RL-L---	461
CONSENSUS-SD	v--w--I-----l-lt---	846
MN251	G--W--I-----L-LT--	880
MN32H	G--W--I-----L-LT--	878
MN1A11	G--W--I-----L-LT--	876
MN3162Q	G--W--I-----L-LT--	877
MN1322L	G--W--I-----L-X	876
MN239	G--W--I-----L-LT--	879
MN142	G--W--I-----L-LT--	881
MNE	V--W--I-----L-LT--	880
P209C19	V-----I--V--L-L---	886
P209C15	V-----I--V--L-L---	885
SIVSMMFBTB	V-----I-----L-LT--	889
SIVSMMFBJA	V-----I-----L-LT--	889
SIVSP6	V-----I-----L-LT--	889
SMNH9	V-----LXX-----L-LT--	886
SHNH4	V-----I-----L-LT--	885
SH62A	V-----I-----F-LT--	886
SHNB670	V--WV--I-----L-L---	883
CONSENSUS-STN	V--R-G-I-----L-LT--	861
STN	V--R-G-I-----L-LT--	883

HIV2/SIV NEF

Protein Alignment Summary Table (Nef)

Name	Accession No.	Region	First Author	Reference
ROD	M15390	complete genome	Guyader,M.	Nature 326 (6114), 662-669 (1987)
NIHZ	J03654	complete genome	Zagury,J.F.	PNAS 85, 5941-5945 (1988)
ISY	J04498	complete genome	Franchini,G.	PNAS 86, 2433-2437 (1989)
ST	M31113	complete genome	Kumar,P.	J. Virol. 64, 890-901 (1990)
BEN	M30502	complete genome	Kirchhoff,F.	Virology 177, 305-311 (1990)
CAM2	D00835	complete genome	Tristem,M.F.	J. Gen. Virol. 72, 721 (1991)
D194	J04542	complete genome	Kuehnel,H.	PNAS 86, 2383-2387 (1989)
GH1	M30895	complete genome	Hasegawa,A.	ARHR 5, 593-604 (1989)
KR	U22047	complete genome	Kraus,G.K.	PNAS 90 (9), 4226-4230 (1993)
MDS	Z48731	gag, pol, vif, tat, rev, nef, env	Becker,M.	Unpublished
UC2	U38293	complete genome	Barnett,S.W.	Unpublished (1997)
UC1	L07625	complete genome	Barnett,S.W.	J. Virol. 67, 1006-1014, (1993)
EHOA	U27200	complete genome	Rey-Cuille,M.A.	Virology 202 (1), 471-476 (1994)
D205	X61240	complete genome	Dietrich Et Al,U.	ARHR 8, 1619 (1992)
FO7844 FO7EC ENVA13	M87069	partial env	Gao,F.	Nature 358, 495-499 (1992)
MM251	M19499	complete genome	Franchini,G.	Nature 328, 539-543 (1987)
MMP11	M72323	tat, rev, nef	Colombini,S.	PNAS 86, 4813-4817 (1989)
MM32H	D01065	env, gag, nef	Rud,E.W.	J. Gen. Virol. 75, 529 (1994)
MM1A11	M76764	env, gag, nef, pol, rev	Luciw,P.A.	ARHR 8: 395-402 (1992)
MM239	M33262	complete genome	Regier,D.A.	ARHR 6, 1221-1231 (1990)
MM142	M16403	complete genome	Chakrabarti,L.	Nature 328, 543-547 (1987)
MNE	M32741	complete genome	Benveniste,R.E.	Unpublished (1990)
MM155	L28171	gp120, gp41, tat, rev, nef partial	Kodama,T.	Unpublished (1994)
MMW25	X86727	partial env, nef	Whatmore,A.M.	J. Virol. 69 (8), 5117-5123 (1995)
NEFW61	X90853	env, nef	Whatmore,A.M.E.	Unpublished
SMMPBJ	M31325	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
SMMPBJB	L03295	complete genome	Dewhurst,S.	Nature 345, 636-640 (1990)
PBJ6P6	L09212	env, gag, nef	Hirsch,V.M.J.	J. Virol. 67, 2466-2474 (1993)
STM	M83293	complete genome	Novembre,F.J.	Virology 186, 783-787 (1992)

HIV2/SIV NEF

	<- env cds	LTR ->	
CONSENSUS-A	MGASGSKrS?psqulrERLLrargp?cggp???spGaysqsgqsgsgrcQkspSCBggqgqGdfmNtPwrtPA?apak???YkQGMDDVdSdddlvgv.pvtrp.rvPlR?MtyklAvDmShfikekgG		118
ROD	-----H-R-PR-Q-----A-A---YNNE-----RF-----D-----R-----KD-A-R-NL-R-----Q-R-B---K---P-HR-I---L-TR--		128
NIH2	-----K-L---Q---Q---ET---RCNE---G-L-H---N---R---V-----A-R-EL-----L---Q-F---P---F-----P---F-----		128
ISV	-----R-R---Q---A---LWDE-E-G---FN-----L---R---R-----T-K-ES-R-----SD-S---A---RM---DL-D---		128
ST	-----E-R-----QTP-EAS---HMDKL---L-----G---R---R-----A---GS-----E---R-R---L---		127
BEN	-----L-KH-R-----DGY-KQRDA-----F-E---N---EY---S---N-T-RQ-DL-R-----I---R-E---I---		128
CAM2	-----R-L---Q---A-T---ECYNALB---SLR-----D---N-L---R-----A---A-K-NA-R---I---A---T---		128
D194	-----EH-----GYVQRNA---S-----E-----AI-Q-NS-----M---E---I---		128
GH1	-----H-KH-R-----H-GYVQCHNA-----KG-----R-----I-Q-XL-----I-S-----A---F---		126
KR	-----C-RSL-----ET---QMDG-A---L-F-----G-NL---R-----AGR-GTL---A-N-N-I---A---I---LN---		128
NDS	-----R-R---Q---A-A---HMDL---R---DNG-R-----T---DS-R-----Y-S---T---A---		128
UC2	-----L-XQ-----GYVRQCSA---S---G-----R---Y-----N-Q-ES-----M-Q---S-S---P---		128
CONSENSUS-B	--SA---q-kqQ?-----??e?Py-?Isgetreqs-r-PC-DNDLN-----ns?gAE,.....GGC-DY-es-E-NE,---?y-R-??--p-----		102
UC1	--SA---Q-KQQ-----TQEEPY-KLSEQRREQS-R-PC-DNDLN-----RMAPRAE,.....GGC-DY-S-E-NE,---Y-R-NR---S---M-----		110
ENOA	--SA---Q-KQQ-----R-PR-ESSEQRREQS-L-YPC-DNDLN-----D-KTLGAE,.....GGGK-DS-ED-E-NE,---R-R-G---P-F-----L---E		110
D205	--SA---BRQ---K---VPERPY-RLSGERRREQS-R-PC-DNDLN-----MARGAE,.....GGC-DA-ES-S-E,---A#Y-R-#I---P-----Q---		110
CONSENSUS-D		X----	5
PC7941		X----	5
CONSENSUS-SD	--gai--?-k?7gd-q-----etY-cl-evEgq---lg-lkgls-l---k-n-qy-----M-e-r-La-Rk-----ee-----,--?,---s---i-----		121
MM251	--GAI-NR--K-AGD-QK-----ETY-RLGGEVEDGS---LG-L-KGLS-R---K-M-QY-----N-E-K-LA-RK---I-EE-----S-R-K---A---I-----		128
MMF11	--GAI-NR--K-TGD-QK-----ETY-RLGGEVEDGS---LG-L-KGLS-R---K-M-QY-----N-E-K-LA-RK---I-EE-----S-R-K---A---I-----		128
MM32H	--GAI-NR--K-AGD-QK-----ETY-RLGGEVEDGS---LG-L-KGLS-L---K-M-QY-----N-E-R-LA-RK---EE-----M---T-S-----		128
MM1A11	--GTI-NR--RSTGD-Q-----ETY-RLGGEVEDGS---LG-L-KGLS-L---K-M-QY-----N-E-R-LA-RK---I-EE-----S-R-K---T-S---I-----		128
MM239	--GAI-NR--R-GD-Q-----ETY-RLGGEVEDGS---PG-LDRGLS-L---K-M-QY-----N-E-R-LA-RK---I-EE-----S-R-K---T-S---I-----		127
MM142	--GAI--K-FRD-Q-----ETY-RLGGEVEDGS---LG-LDRGLS-L---K-M-QY-----N-E-R-LA-RK---I-EE-----I---EA---T-S---I-----		128
MM1E	--GAT--R--KSPGD-Q-----ETY-RLGGEVEDGS---LG-LDRGLS-L---K-M-QY-----N-E-R-LA-RK---I-EE-----I---EA---T-S---I-----		128
MM155	--GAI-NR--R-GD-Q-----ETY-RLGGEVEDGS---PG-LDRGLS-L---K-M-QY-----N-E-R-LA-RK---I-EE-----S-R-K---T-S---I-----		128
MMW25	--GAI-NR--R-GD-Q-----ETY-RLGGEVEDGS---PG-LDRGLS-L---K-M-QY-----N-E-R-LA-RK---I-EE-----S-R-K---T-S---I-----		123
NEFW61	--GAI-NR--R-GD-Q-----ETY-RLGGEVEDGS---PG-LDRGLS-L---K-M-QY-----N-E-R-LA-RK---I-EE-----S-R-K---T-S---I-----		116
SHMPBJA	--GVT---QRKRGGN-Y---Q---ETY-RLWEGLE---DA-KGLS-L---P-K-CE-Q-----N-T-RA-LD-R-----SA-C---S---V-I-----I-----		126
SHMPBJB	--GVT---QRKRGGN-Y---Q---ETY-RLWEGLE---DA-KGLS-L---P-K-CE-Q-----N-T-RA-LD-R-----SA-C---S---V-I-----I-----		126
PEJ6P6	--GVT---QRKRGGN-Y---Q---ETY-RLWEGLE---DA-KGLS-L---P-K-CE-Q-----N-T-RA-LD-R-----SA-C---S---V-I-----I-----		126
SEM9	--GVT---QRKRGGN-x---Q---ETY-RLWEGLE---DA-KGLS-L---P-K-CE-Q-----N-T-RA-LD-R-----SA-C---S---V-I-----I-----		126
SMGH4	--GAI---QYKRGGN-----Q---ETY-RLWEGLE---LGA-KGLS-L---P-K-SE-QY-----N-T-RA-LD-R-----DE---I---S-H---A---I-----		128
SM62A	--GAI---QYKRGGN-----Q---ETY-RLWEGLE---LGA-KGLS-L---P-K-SE-QY-----N-T-RA-LD-R-----DE---I---S-H---A---I-----		128
SM670	--A---KQGG-----Q---ETY-RLWEGLE---LGA-KGLS-L---P-K-SE-QY-----N-T-RA-LD-R-----DE---I---S-H---A---I-----		128
CONSENSUS-STM	-----QRKQGE-----ETY-KLLEGL-EGSGP---GA-DNDLN-H---P-R-NE-Q-----KN---A-SA-LE-R-----EE---H---A-H---E---I-L---S---		128
STM	-----QRKQGE-----ETY-KLLEGL-EGSGP---GA-DNDLN-H---P-R-NE-Q-----KN---A-SA-LE-R-----EE---H---A-H---E---I-L---S---		128

		-- max HIV-1 nef similarity--	
CONSENSUS-A	LeGmEYS?rRhrilLdiyleKKEGIlpDhQNYThGpGvRYPMfFENLWKLvPvDvQe7ed7a7??7thcLvHpaQcSr7DD7MGETLvwrfQpmlay7V7aF7lyPEEPChkSGlpB7aMKAHLKARGIP	234	
ROD	---E---R---N---A---P---G---T---...---XF---P---E---L---S---E---IR---E---R---R---R---	254	
N1H2	---L---Q---L---D---...---T---N---R---R---	176	
ISY	---Y---E---L---T---P---G---T---...---L---M---S---V---F---P---R---KE---TP---I---ME---DO---	254	
ST	---LY---D---R---V---G---K---P---QD---S---...---F---P---T---PS---E---IR---Y---D---R---	253	
BEW	---Q---R---A---ED---...---AN---H---E---Q---E---N---K---T---...---K---	254	
CAN2	---L---E---A---S---TS---G---T---TDTL---L---M---N---R---S---LK---E---TR---D---	256	
D194	---I---RE---LF---T---Y---IS---A---SV---...---H---Y---E---S---K---I---H---K---	254	
OH1	---D---RD---L---C---S---A---D---...---HY---T---H---E---L---T---D---K---I---H---R---	252	
KR	---D---Y---E---M---K---P---GE---...---L---GS---P---M---R---E---T---MR---Y---E---	251	
MDS	---E---K---T---K---LP---D---NT---...---I---KF---Q---SL---E---T---K---Q---E---R---	253	
UC2	---RD---L---V---C---MS---A---D---...---N---M---Y---I---NS---E---K---T---K---	254	
CONSENSUS-B	---iy---E---T---N---VSG---i---k?---??pa---r7a---...---i---SW---i---?---Q---SL---D---V---nrf---YQ---k---?---?	120	
UC1	---IY---E---T---N---VSG---Y---I---RT---...---S---I---SW---I---A---L---ND---Y---MR---YQ---K---X	226	
EHQA	---I---E---X---T---N---VSG---K---INNIA---P---E---...---AW---P---E---Q---SL---D---Y---SRF---YQ---M---X---R---	236	
D205	---Y---E---T---F---N---VSG---I---XY---E---FAATREE---...---E---M---I---SW---I---I---Q---SL---D---V---MRF---YQ---E---R---	237	
CONSENSUS-D	---IY---NH---QN---S---E---L---MY---PD---AQGD---...---R---YQS---PSE---V---SR---R---N---E---IK---S---D---VRRR---T---LY	126	
F07B41	---IY---NH---QN---S---E---L---NY---PD---AQGD---...---R---YQS---PSE---V---SR---R---N---E---IK---S---D---VRRR---T---LY	126	
CONSENSUS-SD	---IY---A---N---D---S---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	243	
NCG51	---IY---A---N---D---S---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	247	
NC11	---IY---A---N---D---S---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	241	
NC12H	---IY---A---N---D---S---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	254	
MM1A11	---IY---A---N---D---S---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	254	
MM239	---IY---A---N---D---S---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	253	
MM142	---IY---A---N---D---S---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	252	
MNE	---IY---E---K---M---H---S---P---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	254	
NC155	---IY---E---K---M---H---S---P---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	254	
MM25	---IY---A---N---D---S---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	249	
MEFW61	---IY---A---N---D---S---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	242	
SMNPBJA	---IY---D---K---L---A---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	252	
SMNPBJB	---IY---D---K---L---A---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	252	
PBJ6P6	---IY---D---K---L---A---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	252	
SMNP	---IY---D---K---L---A---I---KT---N---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	252	
SMNH4	---IY---NE---M---S---I---HY---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	253	
SM62A	---IY---NE---M---S---I---HY---SD---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	254	
SHB670	---IY---DN---S---P---P---E---SN---AQED---...---R---Y---MQ---KW---PW---V---A---K---T---T---E---YVR---S---S---E---VRRR---T---LL	254	
CONSENSUS-SYM	---IY---E---N---V---A---I---KQ---NSN---AQEDD---...---G---Y---HQM---PW---V---A---K---L---HT---E---VRH---S---KE---VERR---T---LL	255	
SYN	---IY---E---N---V---A---I---KQ---NSN---AQEDD---...---G---Y---HQM---PW---V---A---K---L---HT---E---VRH---S---KE---VERR---T---LL	255	

CONSENSUS-A	E&??	236
RQD	--X	257
NIW2	--X	181
ISY	--X	257
ST	--X	256
REN	Y-Ex	258
CAM2	--X	261
D194	Y-Ex	258
GR1	Y-X	255
KR	--\$	253
NDS	-N\$	255
UC2	Y-E\$	257
CONSENSUS-B	??-	221
ENOA	TES	238
D205	TIX	240
CONSENSUS-D	KTADKKETG\$	137
F07841	KTADKKETG\$	137
CONSENSUS-SD	KNADKKETa7	252
MM11	NNADKKETRx	261
MM32H	NNADKKETRx	264
MM1A11	NNADKKETRx	264
MM239	NNADKKETRx	263
MM142	EMADKKETRx	262
MVE	KNADKKETRx	264
MM153	NNADKKETRx	263
NSW25	NNADKKETRx	258
NEFW61	NNADKKETRx	251
SMHPBJA	KNADKKET\$	261
SMHPBJB	KNADKKET\$	261
PBJ6P6	KNADKKET\$	261
SMH9	KNADKKET\$	261
SMH4	KNADKKET\$	262
SM62A	KNADKKET\$	263
SM670	KNADKKET\$	263
CONSENSUS-STN	KNADKKET\$X	265
STN	KNADKKET\$X	265

This page intentionally left blank.

Protein Alignment Summary Table (Gag)

Name	Accession No.	Region	First Author	Reference
AGMTYO	X07805	complete genome	Fukasawa, M.	Nature 333, 457-461 (1988)
AGM155	M29975	complete genome	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)
AGM3	M30931	complete genome	Baier, M.	Virol. 176, 216-221 (1990)
SIVREV	L40990	complete genome	Hirsch, V.M.	J. Virol. 69, 955-967 (1995)
2010G	U10899	gag	Jin, M.J.	J. Virol. 68, 8454 (1994)
AGM677	M66437	complete genome	Fomsgaard, A.	Virol. 182, 397-402 (1991)
AGMT9	L19254	gag, pol	Hirsch, V.M.	Virol. 197, 426-430 (1993)
AGMT17	L19250	gag, pol	Hirsch, V.M.	Virol. 197, 426-430 (1993)
AGMT27	L19251	gag, pol	Hirsch, V.M.	Virol. 197, 426-430 (1993)
AGMT40	L19252	gag, pol	Hirsch, V.M.	Virol. 197, 426-430 (1993)
AGMT49	L19253	gag, pol	Hirsch, V.M.	Virol. 197, 426-430 (1993)
TAN1	U58991	complete genome	Soares, M.A.	Virol. 228, 394-399 (1997)
SAB1C	U04005	complete genome	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SYK	L06042	complete genome	Hirsch, V.M.	J. Virol. 67, 1517-28 (1993)
AGM692	M29974	gag	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)

AGM GAG

VERVET.COM	MGAATSA, LnrRQLdQF57IRLRPaGKKYQIKHLLIMAGKHeRPGHxLLStEeGCKKIIEVlyPLEPTGSGIKSLfMLVCVlyCvHke7KVKDTESAVALVRQoCHLveKEk?A.....??	114
AGM770	-----K-----R-----L-----Q-----H-----S-----	117
AGM155	-----E-M-----R-D-----R-----S-----M-----Y-----L-----I-----D-----T-----	117
AGM3	-----K-H-----T-----S-----F-----OKE-----I-----RN-----ER	119
SIVREV	-----KG-----K-----D-----D-----F-----I-R-Q-----R-----	117
2010G	-----K-----S-----	17
GRIVET.COM	--GGH--.-SG-S-T-K-----K-----K-----Q-----T-----A-----C--IW-I-A-Q-----V--K-HY--D-NEK-----	117
AGM577	--GGH--.-SG-S-T-K-----R-----K-----Q-----T-----A-----C--IW-I-A-Q-----V--K-HY--D-NEK-----	117
TANTALUS.COM	--GH--.-sg-N-t-k-----L-----?k-----K-----Q-----a-----C--IW-I-A-Q-----vi-k-h-----L-----	116
AGM9	--VGH--.-SG-N-A-K-----R-----KL-----R-----K-----K-----Q-----Y-----S-----Y--IW-I-A-Q-----I--K-QS-----T-----	117
AGM17	--GH--.-S-N-T-K-----L-----S-----R-----K-----Q-----A-----C--IW-I-A-Q-----VI-R-H-----T-----	116
AGM27	--SCH-T.-SG-N-T-R-----L-----DK-----K-----Q-----A-----C--IW-I-A-Q-----II-K-H-----T-----	117
AGM40	--GH--.-SG-N-T-K-----P-----L-----DK-----K-----Q-----A-----C--IW-I-A-Q-----H-----9T-----	117
AGM49	--GH--.-TG-N-T-K-----L-----D-----K-----Q-----A-----C--IW-I-A-Q-----VI-K-R-----A-----T-----	117
TAN1	--GH--.-SG-N-T-K-----L-----R-----R-----K-----Q-----S-----C--IW-I-A-Q-----VL-K-R-----T-----	117
SABAEUS.COM	--SN-V.-SG-K-A-SV-----KLR-V-S-LD-S-SAN--K-VV-LS-L-V--N-IA--C--A-I-A-I-----K-K-KEEV.PA-MTES-----	116
SAB1C	--BN-V.-SG-K-A-SV-----KLR-V-S-LD-S-SAN--R-VV-LS-L-V--N-IA--C--A-I-A-I-----K-K-KEEV.PA-MTES-----	116
SYKES.COM	---G-I-TG-E-RY-K---K---R-LVR-V-K-LD---SDQ-M-SK---S-LT-L-L-AN--N---GIIS-VWA--AKKE-E---Q-KQK-KEA-N.WKDDPP-TSGGQSENSSQW	129
SYK	---G-I-TG-E-RY-K---K---R-LVR-V-K-LD---SDQ-M-SK---S-LT-L-L-AN--N---GIIS-VWA--AKKE-E---Q-KQK-KEA-N.WKDDPP-TSGGQSENSSQW	129
TANTALUS.COM	--GH--.-sg-N-t-k-----L-----dk-----K-----Q-----a-----C--IW-I-A-Q-----fi-k-h-----L-----	116
TV_AGM692	--SGS--.-SG-K-----H-----L-----K-----K-----L-----T-----IC-I-Q-A-----IRIR-Q-----D-GEN-----	117
VERVET.COM	??tctssGq.....kKnd?g?tagpGg.SQNPPAQOQ.GhaNvHVPLSPRTLNAWVXAVEERKMFAGIEIVPMFQALSEGCTPYDINQNLNVLGDMQALQIVKEIINEEAAQWdvthPpPAGPLPaGQL	231
AGM770	-----K-IA-----C-----L-----	236
AGM155	..VTFPG.....Q--WT-G-T-----L-----	236
AGM3	MT-----K-V-V-----I-----IA-----	240
SIVREV	..P--E.....SNRET-----P-----	236
2010G	..A-----R.....Q--RSTATSS-Q-----T-----IQ-----	136
GRIVET.COM	AK.....-E..T-----E-R-Y-VVN-.N-----Q-----C--RM--V-----LS--V-----I-----L--V-----E--R--R-----	230
AGM677	AK.....-E..T-----E-R-Y-VVN-.N-----Q-----C--RM--V-----LS--V-----I-----L--V-----E--R--R-----	230
TANTALUS.COM	..aaP-g-.....?Q-yn.ta-----rhG-Y-VV-.N-Q--t-----t-----R-----LS-----I-----M--I--V-----?-----	233
AGM9	..AAP-G-.....QQ-YN.LA-----RHG-Y-V-.N-Q-----T-----R-----LS-----I-----M--I--V-----N--LA-----V-----	236
AGM17	..ATP-G-.....Q-YN.TA-----NHQ-Y-VV-.N-Q--T-----T-----R-----LS-----I-----M--I--V-----S--L-----	234
AGM27	..AAP-G-.....QQ-YN.LA-----RDG-Y-V-.N-Q--Y--M-----C--R--M-----VIS--V--I-----M--I--V-----I-----	236
AGM40	..P-----QHSR.BD--SR-QHG-Y-VV-.N-Q--T-----T-----R-----LS-----I-----M--I--V-----L-----	236
AGM49	..ATP-G-.....Q-YN.T-----QHG-Y-VV-.N-Q--T-----T-----R-----LS-----I-----M--I--V-----I-----V-----	236
TAN1	..AAP-G-.....QQ-YN.TA-T-S-RHQ-Y-VV-.N-Q--T-----T-----R-----LS-----I-----M--I--V--D-----I-----	236
SABAEUS.COM	..-A-----TRELOAK---E..P-VT-S-G-R-Y-IVSV.N-Q--Q--Q-----VI-----S--V-----S--A--AI-----AV-E-----DV-----D--LR---QQ-PAQ-V-	241
SAB1C	..-A-----TRELOAK---E..P-VT-S-G-R-Y-IVSV.N-Q--Q--Q-----VI-----S--V-----S--A--AI-----AV-E-----DV-----D--LR---QQ-PAQ-V-	241
SYKES.COM	MAS-----KVVDQEQ.....AA-P--R....-Y-LLRNPG-Q-I-TGVFV--KT--E--NS--D-S--L--T--FI--L-G--AV-----M-VI-D---G-A--LE--Q-QQ-A-QAG-	253
SYK	MAS-----KVVDQEQ.....AA-P--R....-Y-LLRNPG-Q-I-TGVFV--RT--E--NS--D-S--L--T--FI--L-G--AV-----M-VI-D---G-A--LE--Q-QQ-A-QAG-	253
TANTALUS.COM	..aP-g-.....Q-yn.-a-----?hG-Y-VV-.N-Q--t-----t-----R-----LS-----I-----M--I--V-----l-----	232
TV_AGM692	AKQIDK..T-PT-S-R--Y--.N-V-----VI-----V-----V-----I-----	230

AGM GAG[illegible]

AGM GAG

VERVE7.COM	Pa.....??PYDPANKLLQOYAeKGRQ1ReQ??n.pPa?NPDWtEGYSLMSLPGEDQ\$	505
AGN170	---GTT-----RR----M-----	519
AGN155	-N.....MST-----M-N-MR----N---N-----	520
AGN3	---S-----D-----RKK----V-----	521
SIVREV	---T-----GKR.T-PT---A-----	519
GRIVET.COM	E.....TA-----QR---ERBQ.TRQKEKEV-DV--S---G--	513
AGN677	E.....TA-----QR---ERBQ.TRQKEKEV-DV--S---G--	513
TANTALUS.COM	-AH.....gfpGpp?aGA-----r-e--k-d--k-KE.....KEL-D--S-----	516
AGN9	-AH.....ETSUGPT.AAGA-----R-E--K-D--K-KE.....KEL-D--S-----	522
AGN17	AAH.....GPPTGPP.VAGA-----R-G--K-D--R-KE.....KEL-D--S--	513
AGN27	-AH.....VPPTGLP.AAGA-----RR-E--K-ED--KK-KE.....KEL-D--S-----	524
AGN40	-TH.....GPPTGPP.AQGA-----R-E--K-D--R-KE.....KEL-D--S-----	522
AGN49	-TH.....GVPTGPSWAGA-----R-E--TK-E--K-KE.....KEL-D--S-----	525
TAN1	-AH.....GPPTGSP.AAGA-----R-B--K-D--K-KE.....KEL-D--S---G--	523
SABAEUS.COM	ERDYSRPEENMYADRPPTRGPG-D---TA--K--VQ--RQKQ-WQ-HS-QQS-YEBAYS--R-----	554
SAB1C	ERDYSRPEENMYADRPPTRGPG-D---TA--K--VQ--RQKQ-WQ-HS-QQS-YEBAYS--R-----	554
SYKES.COM	ED...IEDGFWLTWBAQMSQQAQAK-QN6PSKKPPTNREVLSPE..SSCKESTKSLYP--S-----	554
SYK	ED...IEDGFWLTWBAQMSQQAQAK-QN6PSKKPPTNREVLSPE..SSCKESTKSLYP--S-----	554
TANTALUS.COM	-AH.....gfpGpp?a-QA-----R-e--k-d--k-KE.....KEL-D--S-----	517
TV_AGN692	-PH.....HFGA--E-TR--EK-TQE-A-Q-RKVE..KESQAGREE-D---K-----	513

Protein Alignment Summary Table (Pol)

Name	Accession No.	Region	First Author	Reference
AGMTYO	X07805	complete genome	Fukasawa, M.	Nature 333, 457-461 (1988)
AGM3	M30931	complete genome	Baier, M.	Viol. 176, 216-221 (1990)
AGM155	M29975	complete genome	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)
SIVREV	L40990	complete genome	Hirsch, V.M.	J. Virol. 69, 955-967 (1995)
AGM677	M66437	complete genome	Fomsgaard, A.	Viol. 182, 397-402 (1991)
AGMT9	L19254	gag, pol	Hirsch, V.M.	Viol. 197, 426-430 (1993)
AGMT17	L19250	gag, pol	Hirsch, V.M.	Viol. 197, 426-430 (1993)
AGMT27	L19251	gag, pol	Hirsch, V.M.	Viol. 197, 426-430 (1993)
AGMT40	L19252	gag, pol	Hirsch, V.M.	Viol. 197, 426-430 (1993)
AGMT49	L19253	gag, pol	Hirsch, V.M.	Viol. 197, 426-430 (1993)
TAN1	U58991	complete genome	Soares, M.A.	Viol. 228, 394-399 (1997)
SAB1C	U04005	complete genome	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SABD37P	U04018	pol	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
AGM266	M22974	pol	Li, Y.	J. Virol. 63, 1800-2 (1989)
AGM385	M21311	complete genome	Daniel, M.D.	J. Virol. 62, 4123-4128 (1988)
AGM381	M73537	complete genome	Allan, J.S.	J. Virol. 65, 2816-2828 (1991)
SYK	L06042	complete genome	Hirsch, V.M.	J. Virol. 67, 1517-28 (1993)

AGM POL

VERVET.COM	LMKIMFYARTMSGTXXXXMMXXSKRLQ.GTGEFFRVW.XVDQXKTEKF...SRP:SMXGTx.CAPSTExox...TIRPSKEAP	80
AGM3	--KDE-LE-ET-AP-----R-K-----P--SE-Q-----Y--G-AN-----S--C----	84
AGMTYO	-----ENK-SKV-XIGTP-----,T--Q-----Y--S--E--S--RHH.....P-----	86
AGM155	--KA-AL-----P--D--K-----SH--G--K-----QLY.....L-----	89
SIVREV	--EA-TMG-----,T--D--L-----CH--S--E.R--D-----	86
GRIVET.COM	VRQNWPGY---EW--K-----,PLGRSE-K--CAIQ--H--S--N.SP-MGN.....SL-S-----	60
AGM677	VRQNWPGY---EW--K-----,PLGRSE-K--CAIQ--H--S--N.SP-MGN.....SL-S-----	60
TANTALUS.COM	SNAXHARDHNGRSRTKRBQ??R??FLLQM-KD-?-L-RL-KSRTQ??Q--??-C??--?T--Q--EHA?TN-fqaAQ-.FRAR-SR-DGSPDA?-YipHRsas?cgs.....?--Q--C-	102
AGMT9	SNAXHARDHNGRSRTKRBQGRRTYFLLQM-TNR-HSERLPQORTKQVSO--XT-TL--G--DRA-K--KND.TN-IEAAQ-.FRAR-SR-DGSPDA--ROLRRSHS.TGS.....QTTV	117
AGMT17	RR--Q--KEMARTLRPQAAQ-.FRAR-SR-DGSPORR-WIPHRsas.CRS.....L-S-Q-T-	56
AGMT27	M-PO-T-T-RL-KSRTNEVPQV-ET-AMQ---KW--Q--ENANPN-FQ-AQ-.FRAR-TR-DGSKAR-RIPHRsas.CGS.....L--Q--	80
AGMT40	RRTMK-3Q--KA-THG---VRA-K--END.TN-IEAAQ-.FRAR-SR-DGS-HAN-RIPHRSSS.TGS.....Q-II	75
AGMT49	---KNANTN-FQAAQ-.FRAR-SR-DGSPAN-VGSHRSEVRLAGS.....L--Q--	52
TANI	QRQRTFLLQM-TN-A-T-RL-KSRTKQ-PR--EA-TL---PQW--Q--EHA.NL-N-AAQ-.FRAR-SR-DGSPDA--RIPHRFAS.CGS.....Q-TA	98
SABAEUG.COM	-----,PLGRSE-GE-.PSDLHQTNSSPNQTLQQAQGLVCRQTSQDQTRARRSSNSPVK-V	61
SABLC	-----,PLGRSE-GE-.PSDLHQTNSSPNQTLQQAQGLVCRQTSQDQTRARRSSNSPVK-V	61
SYKES.COM	---ERMGQLQEAPSQLPSEERAIC-PS-GHRRMAHAN....MYSTDEP.....TST-E.GT	51
SYK	---ERMGQLQEAPSQLPSEERAIC-PS-GHRRMAHAN....MYSTDEP.....TST-E.GT	51
VERVET.COM	AAxCRExETxEG..TxxXTSNEERLDGFI.FFELPLMRRPXTVXIEVUPxXANLDTGADOTIIEKTDQLXG..PHRPKIIxXIQGJLNVxYxx....xEVxKEXILRGTxxGMPPIHIXGRNL	200
AGM3	--I--QG-AV---,KEK---S-----K--Y--G--IR-L-----A--S--T-K---GG-----E-SD....R--RL-D-----ILL-S-----I--I	184
AGMTYO	--I--R--T---,AKEES-G--G-----K--Y--G--IK-L-----N--S--GG-----E-ND....R--I-D-----ILL-----I--	200
AGM155	--V--R--N-K..SEQKPP-EQ--E---R-CI-GGTAVK-L-----D--R--S-----VGQ-----E-DN....V--QL-D-----VLI-----I--F	179
SIVREV	--V--GK-V---,REEN-T-K--QGR..-----K--I--G--VWVL-----I--T--V--GG-----E-SN....V--I-D-----VLI-----I--	176
GRIVET.COM	P-V---GTAP-R..GERTOKET-GEREG-C.-L-----MKR-I--GT-VQ-L-----Q-K--HPPPHK---S-VVGQ---IH--E-QG....VQ-QL-D--IT-SILI-S-----I--I	182
AGM677	P-V---GTAP-R..GERTOKET-GEREG-C.-L-----MKR-I--GT-VQ-L-----Q-K--HPPPHK---S-VVGQ---IH--E-QG....VQ-QL-D--IT-SILI-S-----I--I	182
TANTALUS.COM	g?V-kkgy.....p-?44-gegtG-L.-l-F-----?KIIY-QGQ-VQVL-----E-IM-E-R..---VVGQ---IS--Q-DN....IL-MIDG-ET--NVLI-P-----V--I	211
AGMT9	G-V-KKQG.....PASEA-OBGTG-L.-L-F-----X	149
AGMT17	G-V-KKRR.....P-KKAKREGTG-L.-L-Fx	82
AGMT27	GVI-KKGG.....SVKEA-CKRAG-L.-I-F-----	121
AGMT40	GTV-E-QG.....P-EKT-REGIG-L.-L-F-----x	107
AGMT49	RTVHE-RO.....T-KET-GEOTG-L.-L-F-----x	84
TANI	G-V-QKGR.....SIEKA-OBGTG-L.-L-F-----MSKIY-QGQ-VQVL-----E-IM-E-R..---VVGQ---IS--Q-DN....IL-MIDG-ET--NVLI-P-----V--I	211
SABAEUG.COM	CCSGETA--AVAKPLA.T-EPLR...G-LQLPQVS-----MK--Y--GQKVT-L-----SV-QG?E-GDN....K-R--GG---CI--?A-HN....Q--?I--?7CKA-ILV--?--V--I--V	171
SABLC	CCSGETA--AVAKPLA.T-EPLR...G-LQLPQVS-----MK--Y--GQKVT-L-----SV-QGIE-GDN....K-R--GG---CI--?A-HN....Q--?I--D7CKA-ILV-E--V--I--V	177
SABD37P	SV-QGLE-GDN....K-R--GG---CI--RA-HN....Q--EI-Q-CKA-ILV-D--V--I--V	59
SYKES.COM	ELSLB-TPHQO-SALABEQ.W-OGDQESLSLSQ-S-----MIE-DV-GDLVQNLV-----R-E-I--HQP..-S--VGGL--NIT-RQ-RNIRFTVVVKPSQKR-QVE--LLV-P--V--L--I	178
SYK	ELSLB-TPHQO-SALABEQ.W-OGDQESLSLSQ-S-----MIE-DV-GDLVQNLV-----R-E-I--HQP..-S--VGGL--NIT-RQ-RNIRFTVVVKPSQKR-QVE--LLV-P--V--L--I	178

VERVET.COM	xxxAGAXLdGQLSEXPITPVKXEGAGPXLKQWPLSKXKINALXUICXQLEKEOKISXGGENAYNTpXfXlxxxDKSxWbXVxxxSxxKATQxPFEKQLxIxPAGLxOxITxLDxGDATYSI	330
AGN3	LAP--K-V-----Q-----LR--R-F-----E-----Q-----Q-----KI-----V-C-RKK--Q-R-L-DPR-LN---D--V--G-PH-S-FEK-TZ--V--I-----	314
AGNTVO	LAP-VPR-V-----K-V---LK--R-CVR-----E--E-Q--S--Q-----RV-----I-C-RKK--Q-R-L-DPR-LN---D--V--G-PH---KK-RQ--V--V-----	330
AGN155	LAQ--K-V---QT-----RLK--R-R-----E--I-Q--KT--E--L-RV--D-----V-C-RKK--Q-R-L-DPR-LN---D--V--G-PH---KK-KQ--II-V-----	309
SIVREV	LAP--R-V-----Q-----LRD--R-HI-----E-----T--E-----KI-----I-C-RKK--TQ-R-L-DPR-LN-V--D--V--G-PH---QK-KQ--V--I-----	306
GRIVET.COM	LAQ--MK-V--V--SQ-EE-K-QLK--KD--K-----RE--E--T--K-N-E--L-RI---P---V-A-RKK--TQ-R-L-DPR-LN---D--V--G-PH---QKXKQ--VI-I-----	312
AGN677	LAQ--MK-V--V--SQ-EE-K-QLK--KD--K-----RE--E--T--K-N-E--L-RI---P---V-A-RKK--TQ-R-L-DPR-LN---D--V--G-PH---QKXKQ--VZ-I-----	312
TANTALUS.COM	LAPL--R-I---N-K---X-QLKP-MS--K-----EE-----T--TOA-----KV---P---V-C-RKK--TQ-R-LIDPRALN---D--V--G-PH---KKXKQ--V--I-----	341
TAN1	LAPL--R-I---N-K---X-QLKP-MS--K-----EE-----T--TOA-----KV---P---V-C-RKK--TQ-R-LIDPRALN---D--V--G-PH---KKXKQ--V--I-----	341
SABAEUS.COM	LAQL-VT-NL??..??-?FIK-?LEP-?D--R?R-----E--E--??-E?--?-HLERI-P--P---V-A-RKK--TQ-R-LIDPR?LN--?-D-Q-V--G-PH---QOREQ--V--I?--F-C	282
SAB1C	LAQL-VT-NL?..RE-?FIK-NLKP-QU--RIR-----E--E--KA--ED--Q-HLERI-P--P---V-A-RKK--TQ-RILNDPRQLN-S--D-Q-V--G-PH---QOREQ--V--I--F-C	305
SABD37P	LAQL-VT-NLVA..KT-LPIK-QLKP-KD--R-R-----E--E--RT--E-----HLERI-P--P---V-A-RKK--Q-R-LNDPR-LN---D-Q-V--G-PH---QOREQ--V--IA--F-C	187
SYKES.COM	LTKL-VK-V-V..TALEPVK-SLRPKEL-R-----VE-LE--KA-VEINL-A-QLEKASPT-P---V-V-RKK--KK-R-LIDPRKLNE--D--V--G-PH-G--K-RQKL-II-LK-----V	305
SYK	LTKL-VK-V-V..TALEPVK-SLRPKEL-R-----VE-LE--KA-VEINL-A-QLEKASPT-P---V-V-RKK--KK-R-LIDPRKLNE--D--V--G-PH-G--K-RQKL-II-LK-----V	305
VERVET.COM	PxDPEPKXtAFTxPxVWxGPGIXYQfXoLxQGIXGxPxixxxTAAxIxXELKxELKPxTIVxTMDxWVGSQEDExXxLVExXKXKxXGULExPExKVQKxPPYEWxGYKxWxHxGxLxSIELxX	460
AGN3	-L-----Y-----I-S--Q---TR--NC-P--R-S-T-FQN--S-LE--K---L--Q---DL-----YT-DR--QL-N--SA---T--K---K-----P---QIS---ED	444
AGNTVO	-L--N---Y---I-T--Q---R--NC-P--R-S-T-FQN--S-LE--KN-PAL--Q---DL-----N-HT-DR--QL-T--QA---T--EK-M--E-----P---E-SR-Q-E-	460
AGN155	-L-----Y---I-T--E--R--NC-P--R-S-T-FQN--SK-LE--K---QL--Q---DL-----EGPK-DQ--QTL-NR-QE---T-EK--RE--F-----P--K-Q---EK	439
SIVREV	-L-----Y---I-S--Q---VR--NC-P--R-S-T-FQN--X-LEK--E---L--Q---DL-----YT-DR--QL-V--ET--T--K---K-----PQ--Q-N--H-E-	436
GRIVET.COM	-LCK---Y---I-S--T--R--NC-P--K-S-T-FQN--N-LE--RHTGLE--Q---DL-LA-DN--TR-MQ-DIV-KM-LEK--T--K--RE--W-----HPN--TINK--PP	442
AGN677	-LCK---Y---I-S--T--R--NC-P--K-S-T-FQN--N-LE--RHTGLE--Q---DL-LA-DN--TR-MQ-DIV-KM-LEK--T--K--RE--W-----HPN--TINK--PP	442
TANTALUS.COM	-L-----PY---I-Q--V--R--NC-P--K-S-T-FQH--QRVLK--GRKVQULE-IQ---DM-I--DLP-KE-DEK--EV-KL-LLR-F-T--K---R--D-----YPSH-S-NK--IP-	471
TAN1	-L-----PY---I-Q--V--R--NC-P--K-S-T-FQH--QRVLK--GRKVQULE-IQ---DM-I--DLP-KE-DEK--EV-KL-LLR-F-T--K---R--D-----YPSH-S-NK--IP-	471
SABAEUS.COM	-L-?D-Q-Y---I-S---?E--?RZ-YKV-P--K-S-T-FQT--NK-LQ-FRQKNPDVD-YQ---DMLIA-DRPKAE-LVM-QQL-DY-ET--FKT-EK-F--D--L---E-YPK--Q-QE-T-P-	408
SAB1C	-L--D-Q-Y---I-S---RE--R--YKV-P--K-S-T-FQT--NK-LQ-FRQKNPDVD-VQ---DMLIA-DRPKAE-LVM-QQL-DY-ET--FKT-EK-F--D--L---E-YPK--Q-QE-T-P-	435
SABD37P	-L-ED-Q-Y---I-S---KE--HRF-YKV	217
SYKES.COM	-L-K---PY---V-SI--AS--ER--TV-P--K-S-T-FQS--INQ-LQFFRKYSOL-LIQ---DLI-TDRS-KA-QEI-QQIVTA-LKV-FKV-KB-W-DQY-MQ-L--T-HPD--Q-QK--FN	435
SYK	-L-K---PY---V-SI--AS--ER--TV-P--K-S-T-FQS--INQ-LQFFRKYSOL-LIQ---DLI-TDRS-KA-QEI-QQIVTA-LKV-FKV-KB-W-DQY-MQ-L--T-HPD--Q-QK--FN	435

VERVET.COM	.XEDVYXNDIXKKVGONMAAXXXGXRFXKIKLWGXKKNLE/VX/XTEAEAYAZEMDKDRTXEQYXXHPOKXXXAAVQKLENGQWSYQFYxR..GooxKVQXYXX...XNNXNXXXXAGLV	584
AGM3	--E--V---QL--L---QLY--L---MLC--I-.,---L--T-T-E-----E-----IL--Q--T-YK--RPFR-----G-----KO--,--QVL---K-TK.,--QK-T-T-EPRVL----	567
AGMTYO	.-DE--V--Q-L--L---QLY--LK-R.-C--ITG----L--L-A-P-----E-----A-IL--Q--T-YK--IPFR-----G-----KO--,--QVL---K-TK.,--QK-T-T-ELRTL----	583
AGM155	--Q--Y--LQ-I--L---QLY--L--N-C-L-.,---L-D-E-P-----E-----IL--Q--T-YA-E-PLR-----GD-----KO--,--KIL---KFAK.,--QRAT-T-ELRVL-V-	562
SIVREV	.-Q--V---Q-L--L---QLY--L---Q-CR-I-.,---L-E-Q-E-----E-----IL--Q--T-YA--PLR-----D-----KO--,--KIL---K-AK.,--QK-T-T-ELRTL----	559
GRIVET.COM	LBGE--Y-K-Q-V--VL---SQIY--IK--HTCAML-,---L--EIV-E-----E-KN-QGIVQSTQ--T-YD-L-ELI-T---QGE---T---TQ.,--AVL---R-AK.,--QRET-T-DLRFL-H--	566
AGM677	LBGE--V-K-Q-V--VL---SQIY--IK--HTCAML-,---L-EIV-E-----E-KN-QGIVQSTQ--T-YD-L-ELI-T---QGE---T---TQ.,--AVL---R-AK.,--QRET-T-DLRFL-H--	566
TANTALUS.COM	.Q-H--V---Q-L--VL---SQLY--IK--ALC--L-.,--AQDLTA--Q--EQ--VEFRQ-Q-IL-E-QQ-A-YD-S-SLR-TIT--DD---G---KQ.,--KVL--T-K-TK.,--SK-T-T-EPRVL--N-	594
TAN1	.Q-H--V---Q-L--VL---SQLY--IK--ALC--L-.,--AQDLTA--Q--EQ--VEFRQ-Q-IL-E-QQ-A-YD-S-SLR-TIT--DD---G---KQ.,--KVL--T-K-TK.,--SK-T-T-EPRVL--N-	594
SABABUS.COM	.R-E--V---Q-L--L---SQIYT-IX--HLCR-I-.,--ARPLT-I-Q-E---LELE--RQILRQKQQ-Q-YD-ALPLR-K-L-GD---G--IYQPE.NKIL---K-AK.,--IKTA-T-ELRNL---	532
SABIC	.R-E--V---Q-L--L---SQIYT-IX--HLCR-I-.,--ARPLT-I-Q-E---LELE--RQILRQKQQ-Q-YD-ALPLR-K-L-GD---G--IYQPE.NKIL---K-AK.,--IKTA-T-ELRNL---	532
SYKES.COM	IDDEX-V-QLQ-LI-VL---SQIYS-IX--ELC-CI-.,-T-QLT--LT-L-EA---ELE--RQIL-E-QR-S-YD-K-PLE-HIT--GSQ--G-MIKQ-QK-PPLIT-KTAK.,--TFAA-S-DYQSL-Q-L	561
SYK	IDDEX-V-QLQ-LI-VL---SQIYS-IX--ELC-CI-.,-T-QLT--LT-L-EA---ELE--RQIL-E-QA-S-YD-K-PLE-HIT--GSQ--G-MIKQ-QK-PPLIT-KTAK.,--TFAA-S-DYQSL-Q-L	561
VERVET.COM	XKICKXALVINGELExXELPxxREVNENNAIXWXWHIPMFVSXPXVKXNXLTXEPFpxEDVYVDGACNRNSKHQKAGTITQYGKXRVKELEMKTXQQAEXXAIKMALKDSGPXMNIYVDSQYA	714
AGM3	Q-L--E3-----VL--IE-----D-Q-S--D---T-PL--L-Y--K-----K-----R-----Q-----T-N--LM-----E--N-----	697
AGMTYO	Q--E-----I--VL--IE-----D-Q-S--D---T-PLL--L-Y--K-----K-----S-----Q-----T-T-N--KLt-----E--N-----	712
AGM155	Q--G-B-----Q--?F--VE-DT-----D-Q-S--D---V-PL-PL-Y--K-----G-----Q-----Q-Q--QT-T-N--LT-----E--X-----	692
SIVREV	Q--E-----IF--VE-----D-Q-S--D---V-PL--L-Y--K-----G-----Q-----Q-----QT-T-N--LV-----E--EK-----	689
GRIVET.COM	Q-----E--T--R--RVQ--VDKKT-DM--QD--Q-S-----T-LL--L-Y-S-VK--KG-----ASKVT-L-----LSER--S-IRS--T-N-----LT-V-----E--EN-----V	696
AGM677	Q-----E--T--R--RVQ--VDKET-DM--QD--Q-S-----T-LL--L-Y-S-VK--KG-----ASKVT-L-----LSER--S-IRS--T-N-----LT-V-----E--EN-----V	696
TANTALUS.COM	QRV--E-ITF--K--IN-V-VE-----SD--Q-T-L--I--I-T-PLI-L-YN-VKD--KAK--C-----L--FL-NT-ESQ-IE--DT-N--LQ-VLL--KH-KSK-----V	724
TAN1	QRV--E-ITF--K--IN-V-VE-----SD--Q-T-L--I--I-T-PLI-L-YN-VKD--KAK--C-----L--FL-NT-ESQ-IE--DT-N--LQ-VLL--KH-KSK-----V	724
SABABUS.COM	Q--G-E??---??-?-?-VE-?-?-?-?D--Q-S-----?-?-pl?l-Y--K-----G-----I-----?-Q-----T-N--L?-?-?-?-?-?	643
SAB1C	Q--G-ESI---QI-IN--VE--L-----SD--Q-T-----N-T-QUIRL-VK-VKD--Q-A-----A-----L-DR-DQK-VA--T-N-K--LE--LL--R--SK--I-----	689
AGM266	xQD--QHS--D-T--V-PL-PL-Y--K-----G-----R-----Q-----Q-----T-N--LT-----E--R-----	100
AGM395	-E-----TL--VE-----D-Q-S--D---T-PNL-L-Y--K--LAK-----R-----Q-----Q-----T-N--	100
SYKES.COM	M--GIQS-WY--KV-TFH--VK--E--K--TD--QAT-V--VK-I-T-PL-RMYVN-VP-----EAVTF-----A-D--T-N--VASD-TQ--QY--QT-N--LEOLL--Q--KEN--V--S	691
STR	N--GIOS-WY--KV-TFH--VK--E--K--TD--QAT-V--VK-I-T-PL-RMYVN-VP-----EAVTF-----A-D--T-N--VASD-TQ--QY--OT-N--LEOLL--O--KEN--V--S	691

VERVET.COM	NGILTAxPTxSDxPLVQIIA=NVxKQxIXxQWVxAXxGLIXxNEIXxKLVxGIxRFLxWxxxIEBAQIEHExVHxNDxxxADThxxxPUIVAKxIVxMCFKQIXGEFxxQVDAxPxxVWxMDxTHxEGKx	844
AGM3	-----Q--Q--S--I-----L--Q--BQ--YL--P--D--GG--E--D--Q--M--KI--FLEK-----R--N--RML--YGL-----E--A-----V--Q-----G--Q--C--L--V	827
AGM7YD	-----Q--Q--S--I-----L--IQ--Q--YL--P--H--GG--E--D--K--V--FLEK-----K--R--N--KML--YGL-----E--A-----V--G-----GT--Q--C--L--K--V	842
AGM155	-----Q--Q--S--I-----Q--Q--EA--YL--P--H--GG--E--D--K--V--I--FIOR-----DR--S--RNL--FGL-----E--A-----V--I--G-----G--Q--C--I--I	822
SIVREV	-----Q--Q--S--I-----L--K--RA--YL--P--H--GG--E--D--Q--V--I--FIOR-----DR--S--KML--FGL-----E--A--NAK-----I--G-----G--Q--C--V--I	819
GRIVET.COM	-N---C--QE--NS-----QALMK--RQVYL--P--H--GG--TE--D--K--QI--FLDR-----DD--AK--N--RSMVQEFGL--N---E--AA-----R--K--G-----IET--Q--C--L--V	826
AGM677	-N---C--QE--NS-----QALMK--RQVYL--P--H--GG--TE--D--K--QI--FLDR-----DD--AK--N--RSMVQEFGL--N---E--AA-----R--K--G-----IET--Q--C--L--V	826
TANTALUS.COM	-----Q--Q--SV--N--EE--IN--DA--YLT--P--H--GG--QE--D--Q--QV--FLDR-----S--DK--T--QFIR--AFGI--AL--E--AA-----R--I--G-----VG--Q--C--L--I	854
TANI	-----Q--Q--SV--N--EE--IN--DA--YLT--P--H--GG--QE--D--Q--QV--FLDR-----S--DK--T--QFIR--AFGI--AL--E--AA-----R--I--G-----VG--Q--C--L--I	854
SABAEUS.COM	---IAGE--E--NNI--Q---EELIX--EAVYIA--P--H--VGG--E--D--Q--QV--FLDR-----DK--A--RSMVQEFGL--A---E--AA-----SV--G-----G--Q--C--L--I	773
SABIC	---IAGE--E--NNI--Q---EELIX--EAVYIA--P--H--VGG--E--D--Q--QV--FLDR-----DK--A--RSMVQEFGL--A---E--AA-----SV--G-----G--Q--C--L--I	819
SYKES.COM	Y---HTC--NTEH--I---QEAIK--EA--YVT--P--H--GG--AVD--K--KI--FLER--PQ--D--R--S--MEYLRQEFHL--RQ---A--IQQ-----NR--K--G---VDIYN--Q--C--E--V	821
SYK	Y---HTC--NTEH--I---QEAIK--EA--YVT--P--H--GG--AVD--K--KI--FLER--PQ--D--R--S--MEYLRQEFHL--RQ---A--IQQ-----NR--K--G---VDIYN--Q--C--E--V	821
VERVET.COM	VIVxVHVASQPIEAIXIPRETGKxTASCALKIXSERPITOxTXQGFHFTxQEVAAACWGXIEHxTGVVYxPxSxGxiEhMKNQLKxiIXKIRxCOYTETAVLHACHxHMFKXKGOIOGXTPAERLIN	974
AGM3	I--A-----V-----E--KFL--L-----LH--D--S-----R--T-----N--Q--Q--S--S-----E--G--D-----A-----IL-----Q--S-----	957
AGM7YD	---A-----V-----E--KFL--L-----LH--D--S-----I--X--T--I--N--Q--Q--S--S-----E--G--D-----A-----IL-----Q--S-----	972
AGM155	---A-----V-----E--KFL--L-----LH--D--S-----I--X--T--I--N--Q--Q--S--S-----E--G--D-----A-----IL-----Q--S-----	952
SIVREV	---A-----V-----RE--KFL--L-----V--LH--D--S-----I--X--T--I--N--Q--Q--SV--S--S-----E--G--D-----A-----IL-----Q--S-----	949
GRIVET.COM	I--A-----V-----E--HFL--LLA--VKHLH--D-----S--N--V-----N--T--I--N--Q--Q--SV--S--R--E--SQ--D--ERL--Q--T--I-----ISS--V--	956
AGM677	I--A-----V-----E--HFL--LLA--VKHLH--D-----S--N--V-----N--T--I--N--Q--Q--SV--S--R--E--SQ--D--ERL--Q--T--I-----ISS--V--	956
TANTALUS.COM	I--A-----L--V--A-----E--KFL--I-----KLH--D--VS--QTI--QV--T--I--N--Q--Q--SV--S--R--ST--E--ED--APL-----I-----M-----	984
TANI	I--A-----L--V--A-----E--KFL--I-----KLH--D--VS--QTI--QV--T--I--N--Q--Q--SV--S--R--ST--E--ED--APL-----I-----M-----	984
SABAEUS.COM	I--A-----?--V--Q--A--HFL--LA-----KLH--D--T--S--Q--I--K--T?--N--Q--Q--VV--S--E--GQ--DAERL--?--V--I-----YSA-----	900
SABIC	I--A-----V--Q--A--HFL--LA-----LH--D--T--S--Q--I--K--TP--N--Q--Q--VV--S--E--GQ--DAERL--I--V--I-----YSA-----	949
AGM181	x--V--Q--A--QFL--LA-----AKLH--D--S--Q--I--K--TP--N--Q--Q--VV--S--E--GQ--DAERL-----x	94
SYKES.COM	IC--A--NT--Y--TKILK--DE--LFLMQ--A-----K--IH--D--VSDKFK--A--CG--T--I--N--Q--Q--IV--SK--RY--EA--SQ--DVTHLQ--A--TFIL-----IS--G--Y--	951
SYK	IC--A--NT--Y--TKILR--DE--LFLMQ--A-----K--IH--D--VSDKFR--A--CG--T--I--N--Q--Q--IV--SK--RY--EA--SQ--DVTHLQ--A--TFIL-----IS--G--Y--	951

AGM POL

VERVET.COM	NITTOLEIOxxOTRIQKILNFRVYYREGRDxVWxGPAQLIWKGEGA.VVxExEZEExVPRAYAKKIXDYBPRKRMGDExNxBGAXGxINQOGxSQTLED\$	1074
AGM3	-----L-HL-----P-K-G-----,--IKGGV-LKEY-----I-K-----S-L--G-A--	1046
AGMYYO	I-----HL-----P-K-----,--LKGGSDLV-----I-K-----KQ-V-N-GOV--TR-S--	1061
AGM155	-----NTL-----P-K--R-----,--LK-G--LV-----I-K-----TL--THL--G-S-K--GD-	1047
SIVREV	-----L-TL-----P-K-----,--IK-G--LKI-----I-K-----G-N--LR-P-----RN-----	1049
GRIVET.COM	-L-----LNTL-NQ-----K-----P-K--R-----,--IK-G-DIKV-----I-K--GE--T-DS-GSN--VREANK--EGD-DLQ-QE	1057
AGM677	-L-----LNTL-NQ-----K-----P-K--R-----,--IK-G-DIKV-----I-K--GE--T-DS-GSK--VREANK--EGD-DLQ-QE	1057
TANTALUS.COM	-----HI--QQ--S--K-----EP-K--T-L-----,--IK-G-LKV-----I-KE-GE--TV--KT-M--	1067
TAN1	-----HI--QQ--S--K-----EP-K--T-L-----,--IK-G-LKV-----I-KE-GE--TV--KT-M--	1067
SABAEUS.COM	I-H-E--TKTL-Q--S--Q-----P-K--K-----,--IQ-QG-LKTI-----I-K--G.,--ALDSQAPL--NORTAGEVD	990
SABIC	I-H-E--TKTL-Q--S--Q-----P-K--K-----,--IQ-QG-LKTI-----I-K--G.,--ALDSQAPL--NORTAGEVD	1039
SYKES.COM	-LY-E-QL-.QN-TSP-FS-----Q-KNE,-K--R-L-----V--QT-EGDIFA-----I-T-HG.,E--DSGSHV-NDEKT-	1036
SYR	-LY-E-QL-.QN-TSP-FS-----Q-KNE,-K--R-L-----V--QT-EGDIFA-----I-T-HG.,E--DSGSHV-NDEKT-	1036

Protein Alignment Summary Table (Vif)

Name	Accession No.	Region	First Author	Reference
AGMTYO	X07805	complete genome	Fukasawa, M.	Nature 333, 457-461 (1988)
AGM155	M29975	complete genome	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)
AGM3	M30931	complete genome	Baier, M.	Virol. 176, 216-221 (1990)
SIVREV	L40990	complete genome	Hirsch, V.M.	J. Virol. 69, 955-967 (1995)
AGM677	M66437	complete genome	Fomsgaard, A.	Virol. 182, 397-402 (1991)
SAB1C	U04005	complete genome	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SYK	L06042	complete genome	Hirsch, V.M.	J. Virol. 67, 1517-28 (1993)
TAN1	U58991	complete genome	Soares, M.A.	Virol. 228, 394-399 (1997)

AGM VIF

VERVET.COM	MnqeKwVMrvTWKvp??1ITKNOGIVzYNNr.kRnLk.N?Y?HKYQICWANYTNRyVIF?g?hGeIhvdLYMHLLPeqBMLSTTavGIGY?snv??k.yRTELDPaTADmaIH?HYFcCfTeRAIq?h	116
AGMTYO	--PM-----GD-----Q-----N-Y-----I-X-----C-----V-LES-----I--G--N--K--Q-	127
AGM155	-S--H--L--QREV-----N-----E-K-----LPGS-----I--A-K-----V-L-ND-----C-----D--Q-	127
AGM3	-----BE-----T-K-D--K-R-----E-L-Q--S-----K-----E--L--RDPW-----L--T-----RX-	127
SIVREV	-----SGD-----K--T-----Q-H--N--C--I--I-LE--QI-C-----R-----I--LQA-GM..P-----G--L--S-----K-	126
GRIVET.COM	-ER--Q-IV--V-R-SBRQ-SR-R--T-KI-.WKQ-P.-E-RH-W-VQ-QFM-Y-QFI--LSDDY-E-MI-HN---R---SHQ--LS-Y..HQMG.-K--V--Q--R--LY--M--D---Q-	125
AGM677	-ER--Q-IV--V-R-SBRQ-SR-R--T-KI-.WKQ-P.-E-RH-W-VQ-QFM-Y-QFI--LSDDY-E-MI-HN---R---SHQ--LS-Y..HQMG.-K--V--Q--R--LY--M--D---Q-	125
SABAEUS.COM	...-H-IV-PL---TGGQER-TSL-K-H-KVSKQCVH-R-TP-TX-R-M--SYQEW---LKQAL-K-TN-----K--E--T--G...SKGE.WF---W--HI--WS--P--D--V-Q-	124
SAB1C	...-H-IV-PL---TGGQER-TSL-K-H-KVSKQCVH-R-TP-TX-R-M--SYQEW---LKQAL-K-TN-----K--E--T--G...SKGE.WF---W--HI--WS--P--D--V-Q-	124
SYKES.COM	...---IVVP--RMTFRQ-DRL-H-IXTHKYKSKE-EKAT-KH---E-Q--YQWT--V.-D-T-WITP-HN-A--R--HMQGR---...QWQNM-D-T-V-RL--NF--P--A--VNO-	123
SYK	...---IVVP--RMTFRQ-DRL-H-IXTHKYKSKE-EKAT-KH---E-Q--YQWT--V.-D-T-WITP-HN-A--R--HMQGR---...QWQNM-D-T-V-RL--NF--P--A--VNO-	123
TANTALUS.COM	-ER--L--T-L--R-SGEH-D-X--K-H--N--QD-T-L---CG---C--FL--L-GE-K-V--C-----A-SF.E-W.QNT-K-VT-DV--H--C--P--D---Q-	126
TAN1	-ER--L--T-L--R-SGEH-D-X--K-H--N--QD-T-L---CG---C--FL--L-GE-K-V--C-----A-SF.E-W.QNT-K-VT-DV--H--C--P--D---Q-	126
VERVET.COM	1rQ7RFv.PCQPPeGHE?..TQqVP6LQTLALLAQ...NG.1RaRSqRgktir?RniGekQgav...GrMakr?Yr.....RsQsgsq?afw?R?PvP...SmELLSQOr?tg?shdgkG16	212
AGMTYO	--M-----S-----K-----S-----Q-----Y-T-----P-GEA--E-T-----K-WY-----	232
AGM155	--M--I-----G--L-----K-----E--T-----Q-----Q-TG-----N-OR--T--P-T-I-----GE--RT-S--I	232
AGM3	-L-Q--T-----K-----Q--S--OGT--M-FE-----HA-----Y--D--A-A-----G-KESH--AR--	232
SIVREV	I--E-----T...--K-----I-----KQA--S-D--RTH...--GH-----IT--Q-A-L...--L-----E--SPN--E--	231
GRIVET.COM	I--EKYT.W-T-K-----Q--L--V-YT...--I-K-K-TF--MAG---R--N...--T-HAQ.....G-KRR--K-L-NEHAN...--CROCKET	219
AGM677	I--EKYT.W-T-K-----Q--L--V-YT...--I-K-K-TF--MAG---R--N...--T-HAQ.....G-KRR--K-L-NEHAN...--CROCKET	219
SABAEUS.COM	I--EKYL.W-XHQV--QF...--RVYT...--RVAPT..S--GSSQ--P-ESQRDT--RNMGFAQRAVRMAPRNVTPQ-RGPV-L-KESFFFS-VEYCGR-EKSWANQAGR	242
SAB1C	I--EKYL.W-XHQV--QF...--RVYT...--RVAPT..S--GSSQ--P-ESQRDT--RNMGFAQRAVRMAPRNVTPQ-RGPV-L-KESFFFS-VEYCGR-SH\$SWANQAGR	242
SYKES.COM	V--ELLTSH-WT...-TD...--QVYL...KD-GGFLQ-LPACA-NTMW-H--KCR-DPKRDQCHCKG-...TQSDR-IQ--YSSRNIN...-L-SILKR-GRD	221
SYK	V--ELLTSH-WT...-TD...--QVYL...KD-GGFLQ-LPACA-NTMW-H--KCR-DPKRDQCHCKG-...TQSDR-IQ--YSSRNIN...-L-SILKR-GRD	221
TANTALUS.COM	I--ES-L.W-TYK--VAENNM--R--P--TVYTDPLR--R-K-P-GK--MV-----Q-----I--HGS.....T--TTP--E-T-L--R-GXEMGTN-R--	238
TAN1	I--ES-L.W-TYK--VAENNM--R--P--TVYTDPLR--R-K-P-GK--MV-----Q-----I--HGS.....T--TTP--E-T-L--R-GXEMGTN-R--	238

Protein Alignment Summary Table (Vpr)

Name	Accession No.	Region	First Author	Reference
AGMTYO	X07805	complete genome	Fokasawa, M.	Nature 333, 457-461 (1988)
AGM155	M29975	complete genome	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)
AGM3	M30931	complete genome	Baier, M.	Virology 176, 216-221 (1990)
SIVREV	L40990	complete genome	Hirsch, V.M.	J. Virol. 69, 955-967 (1995)
AGM677	M66437	complete genome	Fomsgaard, A.	Virology 182, 397-402 (1991)
SAB1C	U04005	complete genome	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
U20965	U20965	vpr	Jubier-Maurin, V.	J. Virol. 69, 7349-7353 (1995)
U20967	U20967	vpr	Jubier-Maurin, V.	J. Virol. 69, 7349-7353 (1995)
TAN1	U58991	complete genome	Soares, M.A.	Virology 228, 394-399 (1997)

AGM VPR

VERVET.CON	MASG.....RQPRBARPGa?EIMDLAREPMdWLRDM1ed?NqEAXIHFGRELLIQWNYCQSE.GER?QcPn?ERAYKYVY:LWQXALPVHFRCCRRRQPFEPY...EERRdGQGGGR	106
AGNTYO	-----V-----K-----Q-L--R-----H--H-----	110
AGN155	-----E--GL-----V-EI-N-----Y-----Q-R-IA-----N-----	110
AGN3	-----L-----I--K-----NR--L--K-----	110
SIVREV	-----G--H--V--N-----L-----R-A--T--R-----	110
GRIVET.CON	-----PL--ML--D-----Q--R-L-E--RR--MM-IR--V--R-HN--WN-IG--I--SM-----G--S-----N-----	109
AGN677	-----PL--ML--D-----Q--R-L-E--RR--MM-IR--V--R-HN--WN-IG--I--SM-----G--S-----N-----	109
SABAEUS.CON	----GMLPPVGGDPPKDPKKN--EI--ML-T--?--F-----Q-L--QC--P-N--RL--WNIV--PAIDR-Q?RI-GW----IL-----MKGR-CKPKTHPA-GPGaGgppP-L--ap	127
SAB1C	----GMLPPVGGDPPKDPKKN--EI--ML-T--P--P-----Q-L-S--QC--P-N--RL--WNIV--PAIDH-Q?RI-GW--C-IL-----MKGR-CKPKTHPA-GPGAGGppP-L--AS	130
U20965	----GMLPPTGGDPPKDPKKN--EV--ML-T--D--F-----Q-L--QY--P-N--RL--WNIV--PAIDR-QSRL-GW----VL-----MKGR-CKPKTHPA-GPG.GG.PP-L--AP	128
U20967	----GMLPFMGDPPKDPKKN--EI--ML-T--A--F-----Q-L--QC--P-N--RL--WNIV--PAIDR-Q-RI-GW----IL--Y--MKGR-CKPKTHPA-GPGAGGppP-L--TP	130
CONSENSUS.TAN	--E-.....S--R--ML-----S--VAEL--QR-----F--M-A-I--R-----T-----N-V----	109
TAN1	--E-.....S--R--ML-----S--VAEL--QR-----F--M-A-I--R-----T-----N-V----	109
VERVET.CON	a?R...VPPGLdS	114
AGNTYO	-N...-----E	119
AGN155	PG...-----	119
AGN3	-G...-----	119
SIVREV	-N...A-----	119
GRIVET.CON	...APPP----	117
AGN677	...APPP----	117
SABAEUS.CON	GG.AaaAa---\$	137
SAB1C	GG.AASAA---\$	140
U20965	GG.AAAA---\$	138
U20967	GG.ASAAA---\$	140
CONSENSUS.TAN	DG..REP----A-	119
TAN1	DG..REP----A-	119

Protein Alignment Summary Table (Tat)

Name	Accession No.	Region	First Author	Reference
AGMTYO	X07805	complete genome	Fukasawa, M.	Nature 333, 457-461 (1988)
AGM155	M29975	complete genome	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)
AGM3	M30931	complete genome	Baier, M.	Virology 176, 216-221 (1990)
STVREV	L40990	complete genome	Hirsch, V.M.	J. Virol. 69, 955-967 (1995)
AGM677	M66437	complete genome	Fomsgaard, A.	Virology 182, 397-402 (1991)
SAB1C	U04005	complete genome	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
U20812	U20812	tat	Jubier-Maurin, V.	J. Virol. 69, 7349-7353 (1995)
U20814	U20814	tat	Jubier-Maurin, V.	J. Virol. 69, 7349-7353 (1995)
SYK	L06042	complete genome	Hirsch, V.M.	J. Virol. 67, 1517-28 (1993)
TAN1	U58991	complete genome	Soares, M.A.	Virology 228, 394-399 (1997)

AGM TAT

VERVET.COM	MDKGBeeg?.vHQDLieqykkPL?tCtMKCECKKCCYBQCQLCFLqKGLGvTYNaphrR.RKKiasl?r?p?q?Q.SlStr?GBD?QeTgesQ?KVEt?A7t??1gRKnl777r?avGA?AS	100
AGMTYO	-----A--I,-----S-D-Q--Q--K-----,---R--MLA-L-H,---KH--Q---PT--E---T-QEN	100
AGM155	-----D-D,-----K--R--E-----F--R---Y---F---,-----AD-I-VFQ-,---I-,---S-----K--EQ-KAHLRTS---GDET-QP---G	118
AGM1	-----D-G.AY-----L-A--KR-----Y--C-----I-,-----B-D-F-E-K,-----S---KG-E---S-R-APB-----AQSGR-T--S	118
SIVREV	-----RT,-L-----R-----T--R-----,---SVQPN-LSQ-D-,-----,---Q-A---X---RETT-AQI-----D-ERDK-E---N--	119
GRIVET.COM	---E-,-PH.PLL-----HR--QP-----Y-----E-----R--VS-K-,---TSTQONQD-IRQ-,---V.Q-NQ---E-GKTE--	88
AGM677	---E-,-PH.PLL-----HR--QP-----Y-----E-----R--VS-K-,---TSTQONQD-IRQ-,---V.Q-NQ---E-GKTE--	88
SABAEUS.COM	--Qe?-aRp.QVWEE-Q-EL?R--QA-d-t-y--?-f--?-?-A--IC-YV--?-R?n-KIQNMKV-VHNS,----.WT--S---KKK-?T?-Q?-?TAD-T---E-?-D-CVPO\$Q?GSG7\$	101
SAB1C	--QEQ-ARp.QVWEE-Q-ELHR--QA-D-T---V--P--I---HK-A--IR-YV--P-RAS-XI-HNOVSLHN	73
U20812	-AQE-.AR..QVWEE-Q-ELQR--QA-D-P-Y-----S---TR-A--IC-YV--K-R-A-RIQNMKV-VHNS,----.WT--S---KKK-KTA-QSTKTADPT---E-SADTCVPO\$QIGSGT\$	116
U20814	--QVR-GRP.QVWEE-Q-ELRR--QA-N-T-Y--A--P--V--N--A--IC-YV---RHS-KIQNMKV-VHNS,----.WT--S---KKK-T--G-P-TADHT---E--SDICVPO\$QFSGIS	119
SYKES.COM	MSSTDQICQTRVFP.SPLEGTFLS.-G-PTP-,-----N-----I--AR--K-AARS-SEDDSA-TGTL.PRAG-TQAMP--KKKA,----EKDSTSLPSAE--	110
SYK	MSSTDQICQTRVFP.SPLEGTFLS.-G-PTP-,-----N-----I--AR--K-AARS-SEDDSA-TGTL.PRAG-TQAMP--KKKA,----EKDSTSLPSAE--	110
TANTALUS.COM	MES,....BDDQABSL--H...R--TP-----R-----V-----I---VS-I-,--P-KNHBNHONLVS-Q---A.N-GMS-----EXT-IPAA-E-SRRFQS	101
TANL	MES,....BDDQABSL--H...R--TP-----R-----V-----I---VS-I-,--P-KNHBNHONLVS-Q---A.N-GMS-----EXT-IPAA-E-SRRFQS	101

Protein Alignment Summary Table (Rev)

Name	Accession No.	Region	First Author	Reference
AGMTYO	X07805	complete genome	Fukasawa, M.	Nature 333, 457-461 (1988)
AGM155	M29975	complete genome	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)
AGM3	M30931	complete genome	Baier, M.	Viol. 176, 216-221 (1990)
SIVREV	L40990	complete genome	Hirsch, V.M.	J. Virol. 69, 955-967 (1995)
AGM677	M66437	complete genome	Fomsgaard, A.	Viol. 182, 397-402 (1991)
SAB1C	U04005	complete genome	Jim, M.J.	EMBO J. 13, 2935-2947 (1994)
U20966	U20966	rev	Jubier-Maurin, V.	J. Virol. 69, 7349-7353 (1995)
U20968	U20968	rev	Jubier-Maurin, V.	J. Virol. 69, 7349-7353 (1995)
SYK	L06042	complete genome	Hirsch, V.M.	J. Virol. 67, 1517-28 (1993)
TAN1	U58991	complete genome	Soares, M.A.	Viol. 228, 394-399 (1997)

AGM REV

VERVET.COM	MF.LGpEE..Rr1lrLia{L?ytNPYPp?VEGTARQRRRAARRMtQXQ?Q?rLaERiw??rQE?Q1vqai.DqL7ld?QHLai...QqL....POPPss??\$	80
AGMTVO	--,-----,--FV--NL-YS-----SG-----Q-D-I-V-V--L,-----,--Y-V,--R-ADEA-----,-----H-A	84
AGM155	--,-----,--Y-S-----S,-----KX--K-IY-----GT--D-----V--T--VT,-----QA	89
AGM3	--,--S,-----NKN-----A-E-L-----MS-V-E-----,--V-Q-----,-----S	89
SIVREV	--,-----,--KVFS--Y-N-----,-----E--R-L-F-----NETK-R--E-TL,---A--V-----,--H,-----N-B-	89
GRIVET.COM	-S,--K--.KQA-KI-KT-YGS---Q.FS-----Q--KQ-Q-IDKI-G-VLNTFEDQ---AQL.QE-Q-ENKD-VL...-R,-----HIH	89
AGM677	-S,--K--.KQA-KI-KT-YGS---Q.FS-----Q--KQ-Q-IDKI-G-VLNTFEDQ---AQL.QE-Q-ENKD-VL...-H,-----HIH	89
SABAEUS.COM	-S,--?--7L--F,--I-KY-YT-----,GQ-----Q--AKQRQ-IIM---1....QTP-SQ--?--AQEFDO-VLDM--P,-----pl-p.ghpss??SS??\$	91
SABIC	-S,--Q--LL--F,--I-KY-YT-----,GQ-----Q--AKQRQ-VIH---L....ETP-SQ--H--AQEFDO-VLDM--P,-----SL-P.GHPTENOTANSSS\$	99
U20966	-S,--K--EL--F,--I-KY-YT-----,GQ-----Q--AKQRQ-ILM---L....QTP-SQ--H--AQELDO-VLDM--P,-----PL-P.GHPSESXULESSA\$	99
U20968	-S,--DIL--F,--I-KY-YT-----,GQ-----Q--AKQRQ-IIM---C....QTS-SQ--S--AQEFDO-VLDM--P,-----PEDQ.AVQSS\$	92
SYKES.COM	--DQ-S-QL.AVF--M--H..QE--G.F--PRQT---R--Q---RT-RLY-QQ--F.EAIFGSRTA-LE-S,-----QISD	77
SYK	--DQ-S-QL.AVF--M--H..QE--G.F--PRQT---R--Q---RT-RLY-QQ--F.EAIFGSRTA-LE-S,-----QISD	77
TANTALUS.COM	-S,--S-DL,--IIQI-RI-YMS-Q---G-----R-Q-FQ-Q-R-VA--S---FIAQ-RRDPSSGE..S-AAAFDQ-VLDM,---VIETL---QEPHDSSTAS\$	103
TANI	-S,--S-DL,--IIQI-RI-YMS-Q---G-----R-Q-FQ-Q-R-VA--S---FIAQ-RRDPSSGE..S-AAAFDQ-VLDM,---VIETL---QEPHDSSTAS\$	103

Protein Alignment Summary Table (Env)

Name	Accession No.	Region	First Author	Reference
AGMTYO	X07805	complete genome	Fukasawa, M.	Nature 333, 457-461 (1988)
AGM3	M30931	complete genome	Baier, M.	Viol. 176, 216-221 (1990)
AGM155	M29975	complete genome	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)
SIVREV	L40990	complete genome	Hirsch, V.M.	J. Virol. 69, 955-967 (1995)
VER1E	U04003	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
VER2E	U04004	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
VER266E	U10896	env	Jin, M.J.	J. Virol. 68 (12), 8454-8460 (1994)
2010E	U10897	env	Jin, M.J.	J. Virol. 68, 8454 (1994)
VER385E	U10898	env	Jin, M.J.	J. Virol. 68 (12), 8454-8460 (1994)
AGM677	M66437	complete genome	Fomsgaard, A.	Viol. 182, 397-402 (1991)
GRI3E	U03994	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
GRI2E	U03995	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
TAN1	U58991	complete genome	Soares, M.A.	Viol. 228, 394-399 (1997)
AGMB05	M81070	env V3	Mueller, M.	J. Virol. 67, 1227-35 (1993)
AGMB30	M81071	env V3	Mueller, M.	J. Virol. 67, 1227-35 (1993)
AGMB53	M81072	env V3	Mueller, M.	J. Virol. 67, 1227-35 (1993)
AGMB14	M80208	env V3	Mueller, M.	J. Virol. 67, 1227-35 (1993)
TAN1E	U03999	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
TAN17E	U04000	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
TAN40E	U04001	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
TAN49E	U04002	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB1C	U04005	complete genome	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
AGMSD29	M81073	env V3	Muller, M.	J. Virol. 67, 1227-35 (1993)
AGMSD30	M81074	env V3	Muller, M.	J. Virol. 67, 1227-35 (1993)
AGMSD37	M81075	env V3	Muller, M.	J. Virol. 67, 1227-35 (1993)
AGMSD45	M81076	env V3	Muller, M.	J. Virol. 67, 1227-35 (1993)
SAB2E	U03996	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB3E	U03997	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB4E	U03998	env	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SYK	L06042	complete genome	Hirsch, V.M.	J. Virol. 67, 1517-28 (1993)

```

VERVET.COM M47t.....lg1?g11?IG1g1v1s?kqQW?TVEYQ?FVWKNSSVQAPCMTPTT:LNATTFNCIPDCHDYP...EVEL.NIT.EPFEAMADANPLVAQAgBNHLLFEQT1KPCVKLSPLCIRM7CVELn 112
AGNTYQ -RY-.....IITL--IV-----I-----I-----S-----G-----A-----M-----N----- 117
AGN3 -L-.....L.I--L--V--NTR--V-----V-----S-----,-----,-----S----- 118
AGN155 -TKP-.....-PIV-G-----QI-T-----I-----V-----,-----,-----N-----K----- 119
SIVREV -PL-.....-VL-AI--SSL--ER--V-----I-----S-----,-----,-----S----- 118
GRIVET.COM -GRL-.....-IKIL-IA--S-GIG..NLV--I-----T-----N-NM-----N-----,-----,-----A-----,-----K-----E-----MR-----I-----S----- 115
AGN677 -GRL-.....-IKIL-IA--S-GIG..NLV--I-----T-----N-NM-----N-----,-----,-----A-----,-----K-----E-----MR-----I-----S----- 115
TANTALUS.COM -GPL-...RGRGVLLV--GLSLIGL-YG.T-YI--I-----I-----N-N-----N-----,-----Q-,-VS,-K-----K-----E-----S-----T-N--?--N-T?-? 116
TAN1 -GPL-...RGRGVLLV--GLSLIGL-YG.T-YI--I-----I-----N-N-----N-----,-----Q-,-VS,-K-----K-----E-----S-----T-N--N-TK-T 121
AGN14 -LL-.....TVLLMLSGCWSLVW...V-YV-----I-----X-----N-N-----N-----,-----Q-,-VS,-K-----K-----E-----S-----T-N--R-N-TR-P 90
SABABUS.COM -LL-.....TVLLMLSGCWSLVW...V-YV-----I-----K--N-N--S-----EPEGTIA--IP--,-K-D--KN-----G--E-----S-----M-----N-YR-E 120
SAB1C -LL-.....TVLLMLSGCWSLVW...V-YV-----I-----K--N-N--S-----EPEGTIA--IP--,-K-D--KN-----G--E-----S-----M-----N-YR-E 120
SYKES.COM -AAF-.....RTYIVC-FSL-SLGFME--YV-----I-K-EDAYAPL--T-SHKQ--K-V-,$A-QI...-RV,-G-Y-P--NSSHNIRO-ILEKSA-L-ANR-----A-M--R-L-TLD- 119
SYK -AAF-.....RTYIVC-FSL-SLGFME--YV-----I-K-EDAYAPL--T-SHKQ--K-V-,$A-QI...-RV,-G-Y-P--NSSHNIRO-ILEKSA-L-ANR-----A-M--R-L-TLD- 119

VERVET.COM s4r??ttptstts?tt??g1?Cv?nkt??nLqsCna?iIEKE???E7ASNCIFMAGY?ADQKNn?YVYVWDABELyCKn?7?7M?7?KE..CYMHNCMSVKEACdRTYWDelRLRYCAPAG 214
AGNTYQ -T-ERA-TPT--PKS-.....-G.PTSGR-----S--R-MED-P-----V-----Y-----K-,-TS-.....-Q----- 235
AGN3 -EPT--K--S--NITASTTT--Q-----STV-E--ET-----LNE-P-----V-----N-----,RG,-,N9NR----- 241
AGN155 -ATS-PA--G-K-.....-R--DS-----DT-----MND-A-----I-----,-----F--BS-SR-GT----- 236
SIVREV -RMP--ASTKA...PKT-D--IKSTNNV--P--SL--E-LBE-A-----I-----,-----SSSS--ST-----E----- 241
VER1E ----- 11
VER2E ----- 11
VER266E ----- 11
VER385E ----- 11
GRIVET.COM GTATTKAT-TA-TTW-.....T-----N-STEQ-G-MAE-P-----I--Q--V--,-MT-Y-Q-LV-M-K-GSEKGS-D-.....T--V----- 224
AGN677 GTATTKAT-TA-TTW-.....T-----N-STEQ-G-MAE-P-----I--Q--V--,-MT-Y-Q-LV-M-K-GSEKGS-D-.....T--V----- 224
GRI3E -----T--V----- 11
GRI2E -----T----- 11
TANTALUS.COM -??P-SS??T??2-??.....-P-T??...??-?-LVTNSH--NS-?-?-?-?-?-NST-Y-?-L?-E?E?-?-?-RQ.....E-----T----- 194
TAN1 -TAPTSSTPTSSST-D.....-PNTDE...S-----TLVNTSMHY-NS-1-S-----R--V--K-.-NST-Y-Q-LV-ERENWT.TGTNG...E-----T----- 233
AGN14 -PTPSSSE-TR-TRN.....-EGTN...DE--SHLVTNSMZF-NS-M-----M-----T-.-NST-Y--LM-EPESKK.NS.RQ...E-----T-----A 190
TAN1E -----T----- 11
TAN17E -----T----- 11
TAN40E -----T----- 11
TAN49E -----T----- 11
SABABUS.COM GGAAT--SP--STA.....R-E-VSVGF.....-DSV--Q-MEK-Q-M-S-----R--V--,-T--D-Q-VV-BEGREKSNATHTVG-----T-----TF----- 232
SAB1C GGAAT--SP--STA.....R-E-VSVGF.....-DSV--Q-MEK-Q-M-S-----R--V--,-T--D-Q-VV-BEGREKSNATHTVG-----T-----TF----- 232
SAB2E -----TF----- 11
SAB3E -----TF----- 11
SAB4E -----TF----- 11
SYKES.COM -PATS-PT--PPTT.....P-----,ET.WMGDNST-PRF--S-NLT-GFK-K-QQ.-RAFFYKIDLH...KEBNSBY...Y-LL--T--SA--E-QTFQPPFIQ---P- 216
SYK -PATS-PT--PPTT.....P-----,ET.WMGDNST-PRF--S-NLT-GFK-K-QQ.-RAFFYKIDLH...KEBNSBY...Y-LL--T--SA--E-QTFQPPFIQ---P- 216

```

```

VERVET.COM 2ALLKcD7dy7Gf6qpcSVSVVHCTgIaMTTVcTGLLLNGSYsBRrTqIHqKHrvsHdsV1i17MkhyNlUvcChRPONKTVLPUTIHAG1VFHSQkYntRLR.....QAWCHFqgnHrgAWkEvk? 133
AGMTYO Y-----E--N-Y-----H-----N-MT--F-----S--R-----MK-----E-----R--Q 359
AGM3 P-----Y--A--T-----N-I-----V-F-----R-----N----- 365
AGM155 Y-----N--A--PE-----T-----Q-----L-----K-----QE 360
SIVREV F-----R--N--M-Y-----S-----N--IV-F--Y-----K-----R--E-----E 365
VER1E Y-----RN-T--APR-R-----S-----T--E--K--N--I-RI-----IR-V-----K-----K-D-----R-I-Q 135
VER2E Y-----H--N-Y-----H-----R-Q--T--I--AF-----R-----E--IQ 135
VER266E F-----A--D--T-----N-----T--A--L-----L-----PS--K-----E 135
2010E S-----K-----N-----L-----I-----N-----E--R 90
VER385E F-----Y--D-----QI-----I-----H-----I-Q 135
GRIVET.COM Y-----K--a--APK-K-----R-I--i--IG--R7--E--ekGN--T-I-KL--F-----R-R-----K-----d-7-----e6 346
AGM677 Y-----K--R--APK-K-----R-I--I--IG--R-----E--GND--T-I-KL--F-----R-R-----K-----D-K-----RE 348
GR13E Y-----K--N--APK-K-----R-I--I--IG--RV--E--EKGN--T-I-KL--F-----R-R-----I-----R-----X-K-----S--L-----RE 135
GR12E F-----R--L--XN--N--APS-K--T-T--R-I--S--IG--RA--E--EKGN--MT--I-KL--F-----R-R-----K-----D-K-----Q 135
TANTALOS.COM F-i--k--TN--t--G.i-r--S-----SSAPG--QA--E--g--I-KL--k?iv-r-----E-----K-----Q-K-----Q-K-----E 316
TAN1 F-i--K--TN--T--G.V-R--S-----SSAPGI--QA--E--Q--N--I-KL--K-KIV-R-----Q--E-----D-K--R--RK 356
AGMB05 --G-----I-KL--Q--IV-R--K-----N-----Q--K-----D-K-----RE 71
AGMB10 T-K-----K-----S-D-K--Q--RQ 34
AGMB53 IH--K-----S-K--KE--E--E 26
AGMB14 V-I--K--TN--T--G.T-R--S-----SSAPG--QA--E--G--I-KL--IL-R-----M--K-----D-K-----RE 313
TAN1E F-i--K--TN--T--G.V-R--S-----SSAPGI--QA--E--G--N--I-KL--K-KIV-R-----Q--K-----D-K--R--RK 134
TAN17E F-i--K--TN--T--G.I-R--S-----SSAPG--QA--E--G--I-KL--F--IX-R-----K--KE--RE 134
TAN40E F-----R--TN--A--G.I-N--S-----ISSAPG--QA--E--GN--I-KL--Q-KIK--I-----K-----D-K-----RE 134
TAN49E F-i--K--TN--T--VG.I-R--S-----SSAPG--QA--E--IN--I-KL--Y-K-Q--R-----K-----H-D-K-----RE 134
SABAEUS.COM Y--R--a--t--s-H--a-k--t--SA--R-I--S--IGI--VA--E--NGN--?--I-RL-RF--IR-R-----N-G-----Q--E 353
SAB1C Y--R--A--T--S-H--A-R--T--SA--R-I--S--IGI--VA--E--NGN--I-RL-RF--IR-R-----K-----W-G-----E 355
AGMSD29 -RI-----A--I-----V-----R-G-----A-E 49
AGMSD30 -R-----K-----W-G--N--S--Q--E 49
AGMSD37 N-----KF-----W-G--N--Q--E 30
AGMSD45 -----W-G--T-----RE 47
SAB2E Y--R--A--T--S-H--A-K--T--SA--R-I--S--IGI--VA--E--NGMA--HT--I-RL-R-F--IR-R-----K-----W-G--D--Q--E 134
SAB3E Y--R--A--T--S-H--K-K--T--SA--R-I--S--IGI--VA--E--NGN--I-RL-R-F--IR-R-----K-----W-G-----Q--E 134
SAB4E Y--R--R--S--R-H--Q--SA--R-I--S--IGI--VA--E--ENG--MTAI--RL--RV--IR-R-----W-G-K-----Q--A--E 134
SYKES.COM YS-----TNFE--DD.V-T--TA-S--QEF--LAS--WFO--T-KAKDKVRFI--QKDK--E--I--VPEALR--QII--E--ESTIKNIQL--AGY--S--S--KFK--KTORDAKR--F--RVT--TEFF--Q--HE 341
SYK YS-----TNFE--DD.V-T--TA-S--QEF--LAS--WFO--T-KAKDKVRFI--QKDK--E--I--VPEALR--QII--E--ESTIKNIQL--AGY--S--S--KFK--KTORDAKR--F--RVT--TEFF--Q--HE 341

```



```

VERVET.COM  MVELNyn77nRT,NVTLSPQIE?IMaELqRYKLVETPIGFAPTeVRRYTQghzQKrv.PFVIGFLqPLGAAGTAMGAaAcALPVSQhLLAGILQOOONLLAAV+AQOQMLVLTINGVENLHARVT  570
AGMTYO      -----SK-----S-----SS-----R-----  605
AGM3        -S-----SK-----T-----D-T-----  612
AGM155      -----NQ-----T-----Q-----Q-----  607
SIVREV      -----NM-----C-----N-----QD-S-----M-----  612
VER1E       -A-----IHK-----A-S-----P-----P.S-E-----  372
VER2E       -----SR-----S-----D-----Q-----T-----  375
VER266E     -----IPK-----S-----D-----A-----D-----  379
2Q10E       -----MQS-----S-----Q-----A-----D-----  302
VER385E     -----IND-----T-----SP-I-----  375
GRIVET.COM  LY--I--NKS GPI-----VRS--Y-D-----D-----P.T-E-----R-----Q-L-----  580
AGM677      LY-AID--NKS GPI-----VRS--Y-D-----D-----P.T-E-----T-----R-----Q-L-----  591
GR1E        LY--I--NKS GPI-----S-VRS--Y-D-----D-----P.S-E-----R-----V-Q-L-----  371
GR1E        LY--I--NKS GPI-----RS--Y-D-----D-----P.T-E-----R-----R-Q-L-----  367
TANTALUS.COM  -M-S-D--SK-M-----TANL-N-Y-----I-K-----V-P.T-E-----A-----T-----Q-----Q-----  565
TAN1        -M-S-D--SKE-----TANL-N-Y-----I-E-----I-V-P.T-E-----A-----T-----Q-----Q-----  607
AGMB05      -M-S-D--SK-M-----TANL-N-Y-----IK-K-----  234
AGMB30      -M-S-D--SK-M-----TANL-N-Y-----I-K-----E-P.T-E-----  219
AGMB53      IM-S-D--TKM-----TANL-N-Y-----N-M-----  185
AGMB14      -L-S-D--DK-M-----TANL-N-Y-----I-K-----X-----  488
TAN1E       -M-S-D--SKE-----TANL-N-Y-----I-E-----I-V-P.T-E-----A-----T-----Q-----M-----Q-----  373
TAN17E      -M-S-D--VNK-M-----TANL-N-Y-----II-K-----D-----V-P.T-E-----S-----R-A-----T-E-----Q-----Q-----  371
TAN40E      IAS-D--NK-M-----TAN--MLY-Y-----I-K-----V-P.T-E-----R-A-----T-E-----Q-----Q-----  373
TAN49E      -M-S-D--SK-M-----TANL-N-Y-----I-K-----V-P.T-E-----R-A-----TE-E-----Q-----Q-----  379
SABAEUS.COM  1Y-----sk-----rs--d-----K-----P-----A-----Q-----Q-----  591
SAB1C       LY-----SK-----S--N--D-----K-----X-----P-----A-----Q-----Q-----  596
AGMSD29     LY-----SK-----H-RS--N-----  192
AGMSD30     -Y-----LK-----L--RS--N--DF--K-----K-----X--P-----  223
AGMSD37     LY-----SK-G-----  162
AGMSD45     LY-----SM-----RS--SY--S-----  198
SAB2E       LY--L--SK-----H--S--N--D-----K-----P-----A-----Q-----Q-----Q-----  369
SAB3E       LY-----SK-----S--N--D-----K-----P-----A-----Q-----Q-----Q-----  368
SAB4E       LY-----SK-----RS--N--D-----K-----P-----A-----Q-----Q-----Q-----  368
SYKES.COM  VLT0IE-YPGSTL--F-PTANV-DV-R-D-FN--IQ-K-----DQ--ELPNT-E-AA-LA--L-S-----G-----L--T--V--QX--E--HL-G--V-----L-  568
SYK         VLT0IE-YPGSTL--F-PTANV-DV-R-D-FN--IQ-R-----DQ--ELPNT-E-AA-LA--L-S-----G-----L--T--V--QX--E--HL-G--V-----L-  568

```

```

VERVET.COM  ALEKYLdQARLNaMGCANRQVCHTTPW.....qmPnrtPdW?NNTWLEWEXQI??PLBGNIT?QL?kareQBERKMLDAYQKL?awsdFWSWDFP SKWLNLKIGFLVvvGiiGLRLLYTtY+ciRV  686
AGMTYO      -----S-----P-T-----Q-----AD-S--G-V-----T-----M-----I-----G-V-----  728
AGM3        -----N-----N-----SY-----T-EE-A-----S-----D-L-----  735
AGM155      -----A-----T--N--E-N-----K--EQ-----K--EQ-----SD--S-----A-I-V-----L-T-----  729
SIVREV      -----V-----Q-M--N-Q-----QE-----E-V-----R-T--N-----G-V-----  735
GRIVET.COM  -----S-----KY--N--K-D-----NA-----QL-EE-QN--S--L--DO--G--SL-T--GYV--I-I-L--PAMVING--RNI  702
AGM677      -----S-----KY--N--K-D-----NA-----QL-EE-QN--S--L--DO--G--SL-T--GYV--I-I-L--PAMVING--RNI  713
TANTALUS.COM  -----T--L--P-----TF-N--D--Q--S--TA-----STT-V-Y-Q-M-T--QD-TSN-NI--V-S-FWMI-N--YI-I-L-LF-MANVING--  687
TAN1        -----T--L--F-----TF-N--D--Q--S--TA-----STT-V-Y-Q-M-T--QD-TSN-NI--V-S-FWMI-N--YI-I-L-LF-MANVING--  724
SABAEUS.COM  -----I--FR-----L-----KY--N--E--Q-----EKY-A--SRI-EQ-K--Q--S--V-----LT--FGMI-AIM-IA--VA-V-LVIIGILRKP  713
SAB1C       -----I--FR-----L-----KY--N--E--Q-----EKY-A--SRI-EQ-K--Q--S--V-----LT--FGMI-AIM-IA--VA-V-LVIIGILRKP  718
SYKES.COM  ---T--R--I-SN--F--I--A-T-EKACGNENFCPK-Q-K--NR--QEVN-TDH-DGL-RE-Q-Q-R-VHDLT--QE-DSL--L--FFY--Y-IGALVL--VSFVGI-KNL  698
SYK         ---T--R--I-SN--F--I--A-T-EKACGNENFCPK-Q-K--NR--QEVN-TDH-DGL-RE-Q-Q-R-VHDLT--QE-DSL--L--FFY--Y-IGALVL--VSFVGI-KNL  698

```

```

VERVET.COM  RQGYsPLSPQIHINw?KG?AdNag.....eP?BGGd??k??seawqK?Sgt??7?k?wckrL?Twls.???tWLY?scLTLLlibLR?A?QYI?QYGLgELW?caAQEa??7a1Ar?AQNaghQIW 784
AGM1Y0      ---V-----QVG--R---D.....-G---MSRIKL--S-D-.....-R--MQ--A--T.RLN-----Q--K---L---A---Q---ILQT--GV---C--- 838
AGM3        -----Q---E.....G-G---KR-NS--P--E--AEW--N-----C-...IS-----V--S-----A-----VV-----L-----Y--- 849
AGM155      -----Q---E---RTG-SK-TM 768
SIVREV      -----R-G-----E---BRDND-S-D---Q---RQKRG-S-.....TFS---TF-----Q-W-----RL-V--P----- 851
GRIVET.COM  ---N--.....S.SAER--G-.....GQR--ESSSSKLIRL-EE-S-PSRINN-WLNFK.SC-LRIR--C--I-----FI-T-VG-L---QQ-QE--TGLAQ---A-RE-WGRIG 817
AGM677      ---N--.....S.SAER--G-.....GQR--ESSSSKLIRL-EE-S-PSRINN-WLNFK.SC-LRIR--C--I-----FI-T-VG-L---QQ-QE--TGLAQ---A-RE-WGRIG 828
TANTALUS.COM ---F-----N-RL.GREQ---QEDKDSSSSRDKS-PSVKESLLPNRQOI-ABERAWRQHLTYWCLTISSWLLRLYQI-RR-LT--QL--QSC-----WQPF-EQ-ARSPF--SA--S-SRTL- 816
TAN1        ---F-----N-RL.GREQ---QEDKDSSSSRDKS-PSVKESLLPNRGGI-ABERAWRQHLTYWCLTISSWLLRLYQI-RR-LT--QL--QSC-----WQPF-EQ-ARSPF--SA--S-SRTL- 893
SABAEUS.COM -K--A---SLPSS-$ 727
SAB1C       -K--A---SLPSS-$ 732
SYKES.COM  LG--V-ILQNPQGR...-DPQKP-D.....-BEGS--REGLNVTFSRE-LRQSLAQOQLWR-.VC-.....-FRS-IRQ-TITWGF-S--FN---I--ASLGREIRDWAAIQA-Y 804
SYK         LG--V-ILQNPQGR...-DPQKP-D.....-BEGS--REGLNVTFSRE-LRQSLAQOQLWR-.VC-.....-FRS-IRQ-TITWGF-S--FN---I--ASLGREIRDWAAIQA-Y 804

```

```

VERVET.COM  LACRSeYR?IinsPRVRQGLBilN$ 809
AGM1Y0      -----N-V----- 863
AGM3        -----A-----G-- 875
SIVREV      ---T--H--S----- 877
GRIVET.COM  AIV-----AV-----KV-G 843
AGM677      AIV-----AV-----KV-G 854
TANTALUS.COM N-----A-LEH---H--E--RWF-- 842
TAN1        N-----A-LEH---H--E--RWF-- 879
SYKES.COM  A-T-RVVEAVAAL--L-----IY-- 830
SYK         A-T-RVVEAVAAL--L-----IY-- 830

```

Protein Alignment Summary Table (Nef)

Name	Accession No.	Region	First Author	Reference
AGMTYO	X07805	complete genome	Fukasawa, M.	Nature 333, 457-461 (1988)
AGM3	M30931	complete genome	Baier, M.	Virology 176, 216-221 (1990)
AGMI55	M29975	complete genome	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)
SIVREV	L40990	complete genome	Hirsch, V.M.	J. Virol. 69, 955-967 (1995)
VER2L	U04017	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
AGM691	M33719	nef partial, LTR	Johnson, P.R.	J. Virol. 64, 1086-92 (1990)
AGTYOAB	M63387	nef partial, LTR	Sakuragi, J.-I.	Virology 185, 455-459 (1991)
AGM677	M66437	complete genome	Fomsgaard, A.	Virology 182, 397-402 (1991)
GRI3L	U04006	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
GRI2L	U04007	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
AGTYOAA	M63386	nef partial, LTR	Sakuragi, J.-I.	Virology 185, 455-459 (1991)
U04016	U04016	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
TAN1	U58991	complete genome	Soares, M.A.	Virology 228, 394-399 (1997)
SAB1C	U04005	complete genome	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
U20811	U20811	nef	Jubier-Maurin, V.	J. Virol. 69, 7349-7353 (1995)
U20813	U20813	nef	Jubier-Maurin, V.	J. Virol. 69, 7349-7353 (1995)
U21093	U21093	3' nef partial, LTR	Jubier-Maurin, V.	J. Virol. 69, 7349-7353 (1995)
U26297	U26297	nef partial	Bibollet-Ruche, F.	Unpublished (1995)
U26298	U26298	nef partial	Bibollet-Ruche, F.	Unpublished (1995)
U26299	U26299	nef partial	Bibollet-Ruche, F.	Unpublished (1995)
SAB1L15	U04008	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB1L14	U04009	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB2L16	U04010	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB2L17	U04011	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB3L18	U04012	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB3L19	U04013	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB4L10	U04014	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SAB4L4	U04015	nef partial, LTR	Jin, M.J.	EMBO J. 13, 2935-2947 (1994)
SYK	L06042	complete genome	Hirsch, V.M.	J. Virol. 67, 1517-28 (1993)

AGM NEF

VERVET.COM	MGLgnSKPqHKKqls1MhALHKTvTrYGLLADPLIG?Sst?QEECDK?LrkSLIRKtNGaMTpeg.RrlQeGD?WdENSD..?EdS...VGPFVrPRVPLRQMTYKLAVDPSHFLKZKGG1DGIYYSd	118
AGNTYO	--SQ---A---YSK--Q---H-----T---V-----A-----Q-----E-E...D-E...T-E...D-E...	123
AGN3	-----A-----Q---L-----A-EE-----K-----K-----R-----E-----	123
AGN155	--S-----TI-R--A--M-----Q---L-----G-----D--Q-----E-----I-----	123
SIVREV	-----HV-----Q--K-----T--AV---Y--G-----D-----E-----D-----X-----E	123
AGN691		8
AGTYOAB		8
GRIVET.COM	--SS---R-QQGL-K--RG-RGKPGADWV--S-----Q---V---G-A-K--W...GK-K--D-----TF---D...D-E...--Q-----S-----E	120
AGN677	--SS---R-QQGL-K--RG-RGKPGADWV--S-----Q---V---G-A-K--W...GK-K--D-----TF---D...D-E...--Q-----S-----E	120
GRI3L		1
GRI2L		1
TANTALUS.COM	--GS---REQGL-R--R--R-AP-V--M-----Q--MI-----NNGGS...TNRGKSTPE.G-KLAA-D..T-D-WEP-E-E...--7---W-P	120
SIU04015		7
TAM1	--GS---REQGL-R--R--R-AP-V--M-----Q--MI-----NNGGS...TARGKSTPE.G-KLAA-D..T-D-WEP-E-E...--E---W-P	121
SABAENS.COM	--GKS--Q-QQcS-W--sk-RqPeI--DM-----L-Q--HI---A-S-?QG--?QGDSSR-E--ik?k??-rQP.S-Y...-DE-E...--?--?--I--G-----E---E	116
SAB1C	--GKS--Q-QRHS-W--SK-RQAP-IQ-DM-----L-Q--HI---A-S-DG--QGDSSR-E--VKKHQ-RQP.S-Y...EDE-E...--CL--A-----I--G-----E---E	124
SIU20811	--GKS--Q-QERS-W--AK-RQVPEIQ-DM-----L-Q--HI---A-S-KDG--KQGDSSR-E--VKKHQ-RQP.S-Y...DDE-T...--Q--H--T--F--G-----E---F-E	124
SIU20813	--AKS--Q-RQORS-W--SK-REAPQI--DM-----L-Q--HI---AES-EDG--KQGDSSR-ED-ITMRHP-RQP.S-Y...DDE-E...--I-----P--I--G-----E---F-E	124
SIU21093		8
SIU26297	--GKS--Q-QQRS-W--SN-RQVPEI--DM-----Q--HI---A-S-SDG--QGDSSR-E--IKVKGf-SQP.S-Y...E-E-E...--S--K--A-----I--	107
SIU26298	--GKS--Q-QQRS-W--SK-RQAPQI--DM-----Q--HI---A-S-SDG--QGDSSR-E--IKVKGf-NQP.S-Y...D-E-E...--E-----I--	107
SIU26299	--GKS--Q-QQHS-W--SK-RQAPQI--DM-----L-Q--MI---A-S-DGS-KQGDSSR-E--IKYKQf-RQP.S-Y...EDE-E...--T--F-----	107
SYRES.COM	--STS--S-QLRSEKRYAIGWRLFGKQ--TP-P-E-SRFLQPCRGGF--AW-ST-TEPIDPHQPDRL..DWGHS-GQKFSPG-IVQDEGIVPGL----C-QT--TL-----I-L---I-N---Q-MN-CE	128
SYR	--STS--S-QLRSEKRYAIGWRLFGKQ--TP-P-E-SRFLQPCRGGF--AW-ST-TEPIDPHQPDRL..DWGHS-GQKFSPG-IVQDEGIVPGL----C-QT--TL-----I-L---I-N---Q-MN-CE	128

VERVET.COM	xRnkILMLYALNEMGIIDDMaWp7GPGIRyPrCF0FCFALVFPVdHEEAotCCERMcLVHPAQ??EDPdg1?Hge7LWVKF0PMLAvtydP7x7yftDmh??v??k\$	213
AGMTYO	--K-----F-R-----Q-----MQ-----S-----I-----N-E-----GL-.KR-	229
AGM3	--Q-----E-----VR-----N-V-----N-X-L-----DLG.KR-	229
AGM155	--K-----A-----LN-----M-I-A-----S-E-----LYST-GTGN	230
SIVREV	--R-----T-V-----EM-----Y-A-----IX-----S-T-----N-Q-E-----AI-.KR-	229
VER2L	x-----K-----K-----KH-----VQ-----N-V-----T-E-----ST-GTGN-	107
AGM691	--E-----Q-YTP-----E-----S-QG-----LQ-----W-T-----R-N-----CKAM-GV..-N--AT-.G--	111
AGTYOAB	--A-----Q-YTP-----R-----V-----MR-----DPE-N-V-----CKAS-GV..-N-.K--	110
GRIVET.COM	--E-----QYVlp-----?c-----?--XN-----l-----aG7---?7---v-----K---a??-WM..?k-----a.kr\$	211
AGM677	--E-----Q-Y-P-----V-----RN-----M-----MQ-----D-----V-----K---E-R-DM..-K---EHA.KR	223
GRI3L	--E-M-----X-----QGYTP-----LT-----T-S-----QM-D-----L-----AGG-G-PGQ-----T-----K---EPK-WM..HK---GmE.KK\$	104
GRI2L	--EA-----QGYTR-----LT-----I-S-----AM-Y-----L-----AGGSE-PGQ-----V-----K-T-EPK-WM..HR---EHA..\$\$	101
AGTYOAA	x-E---I---\$-----Q-YTP-----T-----RN-----IL-----LQ-----D-----V-----RK-----R-EM..-----GmH.KR\$	102
TANTALUS.COM	K-BQ-----?-----Q-Y?P---T-K-----E-----VSQ--Q.D-----L-----IENES-PW..K-T-----N-L-----N-DS..-K---GL-.KR--	219
SIU04016	K-BQ-----\$-----Q-Y?P---T-K-----E-----VSQ--Q.D-----L-----IENES-PW..K-T-----N-L-----N-DS..-K---GL-.KR--	106
TAN1	K-BQ-----Q-Y-P---T-K-----E-----VSQ--Q.D-----L-----IENES-PW..K-T-----N-L-----N-DS..-K---GL-.KR--	222
SABAEUS.COM	?-X---D-----V-G-QNYTD---t---K7-w-----s---?n?--?--L---IayED-PW..K-t-----?---D-VAM-L.HPEQVPSa-s-\$	211
SAB1C	--X---D-----V-G-QNYTD---T---K---N-----S---KNS-N---L---VAYED-AW..K-T-----L---D-VAM-L.HPEQVPSAQG\$	226
SIU20811	--X---D-----V-G-QNYTD---T-F-KM-NL-----NSD--HDSOT---L---IAYED-PW..K-V-----D-VAM-L.-PEQVPSGLSQ\$	226
SIU20813	--X---D-----V-G-QNYTD---T---KQ-N-----MYD--HMD---L---IAYED-QW..K-T---R-----D-VAM-L.VPEQVPSGLSAS	226
SIU21093	--X---D---M-----V-G-QNYTD---A---KM-NL-----NSD--HDSOT---L---IAYED-PW..K-V-----D-VAM-L.-PEQVPSGLSQ\$	110
SAB1L15	x-X---D-----V-G-QNYTD---T---K---N-----S---KNS-N---L---VAYED-AW..K-T-----L---D-VAM-L.HPEQVPSAQG\$	102
SAB1L14	x-X---D-----V-G-QNYTD---T---K---W-----S---KNS-N---L---VAYED-AW..K-T-----L---D-VAM-L.HPEQVPSAQG\$	102
SAB2L16	x-X---D---M-----V-G-QNYTD---T---KQ-W-----NS--HMD---L---IAYED-PW..K-V-----V---D-VAM-L.HPEQVPS-SAS\$	102
SAB2L17	x-X---D---N-----V-G-QNYTD---T---KQ-W-----NS--HMD---L---IAYED-PW..K-V-----V---D-VAM-L.HPEQVPS-SAS\$	102
SAB3L18	x-K---D-----V-G-QNYTD---T---K---W-----S---KNS-N---L---VAYED-PW..K-T-----SI---D-AAM-L.HPEQVPS-SDS\$	102
SAB3L19	x-K---D-----V-G-QNYTD---T---K---W-----S---KNS-N---L---VAYED-PW..K-T-----V---D-VAM-L.HPEQVPS-SDS\$	102
SAB4L10	x-X---D-----V-G-QNYTD---T---KQ-----A-S-A-QND-H---L---IAYED-PW..K-T-----I---D-VAM-L.HPEQVPS-SAS\$	102
SAB4L4	x-K---D-----V-G-QNYTD---T---KQ-----A-S-A-QND-H---L---IAYED-PW..K-T-----I---D-VAM-L.HPEQVPS-SAS\$	102
SYKES.COM	K-DE--H--LQ--H---RI.NYTS---T---LI--NLME---NEIGCL-YE-MTL-L---SITGEPIAW...-N-N-N-N-H--I	209
SYK	K-DE--H--LQ--H---RI.NYTS---T---LI--NLME---NEIGCL-YE-MTL-L---SITGEPIAW...-N-N-N-N-H--I	209

III

Analyses

PART III Analyses

HIV-1 Coreceptor Use: A Molecular Window into Viral Tropism	1
<i>Robert W. Doms and John P. Moore</i>	
The History of HIV-1 Biological Phenotypes Past, Present and Future	13
<i>Eva Maria Fenyö, Hanneke Schuitemaker, Birgitta Åsjö, Jane McKeating, Quentin Sattentau, and the EC Concerted Action HIV Variability</i>	
Updated Proposal of Reference Sequences of HIV-1 Genetic Subtypes	19
<i>Thomas Leitner, Bette Korber, David Robertson, Feng Gao, and Beatrice Hahn</i>	
Intersubtype Recombinant HIV-1 Sequences	25
<i>David L. Robertson, Feng Gao, Beatrice H. Hahn, and Paul M. Sharp</i>	
Fine Structure of HIV and SIV	31
<i>Hans R. Gelderblom</i>	
Reagents for HIV/SIV Vaccine Studies	45
<i>Carla Kulken and Satish Pillai</i>	
gp41, A Multifunctional Protein Involved in HIV Entry and Pathogenesis	55
<i>Eric Hunter</i>	
Global Variation in the HIV-1 V3 Region*	74
<i>Priya K. Dighe, Bette T. Korber and Brian T. Foley</i>	
Mutations in Retroviral Genes Associated with Drug Resistance	206
<i>Jennifer Hammond, Brendan A. Larder, Raymond F. Schinazi, and John W. Mellors</i>	
Drug Resistance Mutations Superimposed on the Structures of HIV-1 Protease and Reverse Transcriptase	250
<i>Holly Tao</i>	

*See pages 91–95 for a listing of two-letter country codes.

HIV-1 Coreceptor Use: A Molecular Window into Viral Tropism

Robert W. Doms¹ and John P. Moore²

¹ *Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104;*

² *The Aaron Diamond AIDS Research Center, The Rockefeller University, 455 First Avenue, New York, NY 10016*

The identification of the chemokine receptors CCR5 and CXCR4 as the major coreceptors for HIV-1 has provided a new framework for understanding viral tropism and pathogenesis at the molecular level. It is now possible to assign molecular designations to virus isolates, based on their coreceptor use, that largely explain viral phenotype (reviewed in [Berger(1997), Broder & Collman(1997), Doms & Peiper(1997), Moore(1997)]). The need to accurately describe viral phenotype arises because HIV-1 strains can exhibit distinct cellular tropisms that have important implications for viral pathogenesis and disease progression. Generally, virus strains that are transmitted between individuals are able to infect both macrophages and primary CD4+ T-cells, but are unable to replicate in transformed T-cell lines [Connor et al.(1993), Roos et al.(1992), Schuitemaker et al.(1992), Zhu et al.(1993)]. As a result, they fail to form syncytia when grown in MT-2 cells or in certain other commonly used cell lines. Viruses with these properties have been referred to both as macrophage tropic (M-tropic) due to their ability to infect macrophages, non-syncytium inducing (NSI) due to their inability to form syncytia on T-cell lines, or slow-low (SL) in reference to their replication kinetics in culture [Fenyo et al.(1988), Schuitemaker et al.(1992)]. With time, typically about 4-5 years after infection, virus strains evolve in some individuals (about 50%) which can infect T-cell lines in addition to primary T-cells [Tersmette et al.(1988), Tersmette et al.(1989)]. While this shift in tropism can sometimes be accompanied by a loss of ability to infect macrophages, more often primary isolates retain this property and so are referred to as dual-tropic [Collman et al.(1992)]. Viruses able to infect T-cell lines have been variously referred to as T-tropic, syncytium-inducing (SI), or rapid-high (RH) according to the nomenclature schemes mentioned above. The emergence of these virus types is correlated with accelerated disease progression [Connor et al.(1993)]. Finally, viruses that are well-adapted to grow on transformed cell lines by continual passage are referred to as T-cell line adapted (TCLA). TCLA viruses exhibit greater sensitivity to neutralization by both antibodies and soluble CD4.

The above three systems for classifying viral strains are problematic for several reasons. The M-tropic designation, for example, can be taken incorrectly to mean that a virus is unable to replicate in primary T-cells. Likewise, T-tropic viruses have been described that retain the ability to infect macrophages. The NSI designation is also misleading, as it suggests that there is something inherently wrong with the env protein from these virus strains. It also implies that SI viruses are more fusogenic and cytopathic, which is not necessarily so when they are grown in primary CD4+ T-cells. It is now known that the SI/NSI designation is an artifact resulting from the fact that T-cell lines express CXCR4 but not CCR5; the env proteins from NSI viruses are perfectly capable of forming syncytia provided that target cells expressing CCR5 are used. Finally, an additional complication arising from the current classification schemes is that the terms M-tropic, NSI, and SL are often used synonymously, as are T-tropic, SI, and RH. While this is sometimes appropriate, there are many instances in which these terms are not synonymous.

Chemokine Receptors and Viral Tropism

The major determinant of viral tropism is at the level of entry, more specifically at the level of membrane fusion with CD4+ cells. This occurs efficiently only if the appropriate coreceptor is present. Thus, M-tropic, NSI viruses use CCR5 in conjunction with CD4 for fusion. Direct interactions between the env glycoprotein and CCR5 have been detected [Trkola et al.(1996), Wu et al.(1996)].

and it is likely that this interaction triggers the final conformational changes in env that promote fusion between the viral and cellular membranes. The importance of CCR5 for HIV-1 infection *in vivo* was shown by the discovery that approximately 1% of Caucasians lack CCR5, and that these individuals are highly resistant but not entirely immune to virus infection [Dean et al.(1996), Liu et al.(1996), Michael et al.(1997), Samson et al.(1996)]. Thus, CCR5 is the major coreceptor for HIV-1 transmission *in vivo*. However, while CD4-positive cells obtained from CCR5-negative individuals are resistant to infection by viruses that require this coreceptor, they are readily infectable by viruses which use CXCR4 [Liu et al.(1996), Rana et al.(1997), Samson et al.(1996), Zhang et al.(1997)]. It is not clear why CCR5-negative individuals are only rarely infected by viruses that can use CXCR4 for cellular entry (for some examples see [Balotta et al.(1997), Theodorou et al.(1997), O'Brien et al.(1997), Biti et al.(1997)]). It appears that viruses which use CXCR4 are inefficient at establishing an infection in a naive host, for unknown reasons.

While M-tropic NSI viruses use CCR5 for cellular entry, CXCR4 is the coreceptor most commonly used by T-tropic, SI virus strains [Feng et al.(1996)]. Viruses that use CCR5 for entry can evolve to use CXCR4 through mutations in the env glycoprotein, usually but not always in the V3- loop. Despite less than 20% amino acid identity between CCR5 and CXCR4 in their extracellular domains, several viruses are known to efficiently use both coreceptors for cellular entry. Use of CXCR4 is largely dependent upon the first and second extracellular loops of this receptor, which are considerably more anionic than the corresponding domains in CCR5 [Brelot et al.(1997), Lu et al.(1997)]. It may be relevant that the V3-loops of T-tropic HIV-1 env proteins tend to be more basic than those found in viruses which use CCR5 as a coreceptor [J. et al.(1992), Fouchier et al.(1992)], suggesting a charge-charge interaction may be involved in the gp120-CXCR4 interaction.

In addition to CCR5 and CXCR4, at least nine other chemokine or orphan receptors have been shown to support the cellular entry of one or more virus strains. These include CCR2b, CCR3, CCR8, GPR1, GPR15, STRL33, US28, V28, and ChemR23 [Choe et al.(1996), Deng et al.(1997), Doranz et al.(1996), Farzan et al.(1997), Liao et al.(1997), Pleskoff et al.(1997b), Reeves et al.(1997), Rucker et al.(1997), Samson et al.(1997)]. Though the *in vivo* significance of these alternative coreceptors is not clear, it is possible that the use of receptors other than CCR5 and CXCR4 may be important for certain aspects of viral pathogenesis. For example, CCR3 is expressed in microglia, and use of this receptor may be correlated with neurotropism [He et al.(1997)]. A major challenge in this rapidly developing field will be to determine if use of these additional receptors has implications for viral pathogenesis.

A Classification System Based on Coreceptor Use

The importance of viral phenotype for pathogenesis and disease progression coupled with the imprecise nature of the current systems for classifying virus strains calls for the development of a more accurate classification scheme. Recently, a new system based on coreceptor use has been proposed [Berger et al.(1997)]. In this scheme, virus strains that use CCR5 as a coreceptor are designated R5, while those that use CXCR4 are designated X4. Viruses that use both coreceptors are designated R5X4. Thus, under this system the HIV-1 strains most commonly transmitted would be called R5 viruses. However, this designation makes no assumptions about the cells in which HIV-1 replication must occur (unlike the M-tropic/T-tropic designation), or the rate of virus replication (unlike the SI/RH system). Viruses that evolve the capacity to use CXCR4 in infected individuals with or without concurrent use of CCR5 would now be called R5X4 or X4 viruses, respectively. The primary advantage of this classification system is that it offers a precise molecular description of the major coreceptors used by any given virus strain without assuming that this necessarily results in the ability to efficiently replicate in a particular target cell. Furthermore, this system can be readily modified, taking into account other coreceptors (such as CCR3 by using an R3 designation) if their use proves to be a major determinant of tropism. We have compiled a table listing the coreceptors used by more than 100 HIV-1 strains, and discuss some of the nuances of co-receptor use below.

Issues Pertaining to Coreceptor Use

As shown in the Table, the major coreceptors used by greater than 100 HIV-1 strains have been determined in little more than a year. All strains listed in the Table have been examined for the ability to use CCR5 or CXCR4, though not necessarily for other coreceptors such as CCR3. From the studies published to date, it is clear that CCR5 and CXCR4 are the major HIV-1 coreceptors since all HIV-1 strains examined thus far use one or both of these receptors. In addition, coreceptor use is phenotype but not genotype dependent. All NSI viruses (irrespective of genetic subtype) studied to date use CCR5 while all SI viruses use CXCR4 (though many also use CCR5). Further, our review of the literature has shown that determining which of these receptors is used by a given virus strain is straightforward and independent of technique or assay. We have found no discrepancies between various virus infection, virus-cell fusion, and cell-cell fusion assays that have been used to determine if virus strains use CCR5 or CXCR4. The only differences concern the relative efficiencies with which R5X4 viruses use CCR5 and CXCR4, but these are relatively insignificant in terms of magnitude. For the purposes of the Table, we have arbitrarily chosen 10% efficiency as the cut-off for relevant coreceptor use. Thus, a virus that uses CCR5 as its most efficient coreceptor is listed as a R5X4 virus if it uses CXCR4 to levels > 10% of that observed for CCR5. The efficiency with which different co-receptors are used by some strains is likely to vary between assay systems, and low-level entry via several co-receptors may not be uncommon. Whether this low level of relative efficiency can support virus infection *in vivo* is not known. As a result, the threshold efficiency for assigning relevance to coreceptor use may have to be raised or lowered in the future.

While many virus strains have been examined for the ability to use CCR5 and CXCR4, only a small number have been tested for use of most of the other viral coreceptors (see Table). Thus, we have not listed coreceptors that are NOT used by a given virus strain. Unlike studies in which use of CCR5 and CXCR4 was examined, we occasionally found discrepancies between different studies with regards to the use of additional coreceptors by some virus strains. The reasons for this are not clear, but may well be due to the fact that use of coreceptors other than CCR5 and CXCR4 tends to be inefficient, thus magnifying assay dependent differences. Perhaps the most significant variable in determining whether a given chemokine or orphan receptor can function as a coreceptor is the level of its expression. Very high levels of CCR3 expression support env-mediated membrane fusion by the majority of HIV-1 strains tested, while at lower levels only a few virus strains could use CCR3 [Rucker et al.(1997)]. Thus, expression levels can strongly influence coreceptor use. Another important variable is the expression level of the coreceptor relative to CD4 [Kozak et al.(1997)]. When CD4 is present on the cell surface at high levels, surprisingly low levels of CCR5 or CXCR4 are needed to support virus entry. However, if CD4 is expressed at low levels, higher expression levels of the coreceptors are needed [Kozak et al.(1997)]. Whether this will hold true for other coreceptors is not known, but clearly both CD4 and coreceptor cell surface expression levels can influence the efficiency of viral entry. As antibodies to additional chemokine and orphan receptors are identified, it will be possible to determine expression levels on relevant target cells. This will make it possible to assess the significance of coreceptor use determined from *in vitro* assays using cell lines or transient expression systems.

Determining the levels at which various receptors are expressed is but one step in determining their significance *in vivo*, since it is also important to determine if they are expressed in relevant target cells. In general, this is likely to be limited to CD4-positive cells. However, HIV-1 has been detected in a variety of CD4-negative cells, and several HIV-2 and many SIV strains have been identified that can utilize either CXCR4 or CCR5 for cellular entry in the absence of CD4 [Edinger et al.(1997), Endres et al.(1996), Reeves et al.(1997)]. In the case of SIV, a number of neurotropic SIV isolates have been shown to infect brain capillary endothelial cells, the principal component of the blood-brain barrier, in a CD4-independent, CCR5-dependent manner [Edinger et al.(1997)]. Thus, expression of an HIV-1 coreceptor in some CD4 negative cells may still have implications for viral pathogenesis. In fact, for many receptors, expression in cells other than T-cells and macrophages may be the only place that they support virus infection, at least during the early stages of disease - it is becoming increasingly clear that the R5 viruses which predominate during the asymptomatic period of the disease fail to infect stimulated PBMCs from individuals who lack CCR5, indicating that receptors such as CCR2b and CCR3 do not play a significant role in supporting virus entry into the primary targets of HIV-1 *in vivo*.

HIV-1 Coreceptor Use

[Zhang et al.(1997)]. However, this may not be the case for viruses which emerge later in the course of disease.

Another important consideration in determining the *in vivo* relevance of an alternative coreceptor is the number and type of virus strains that can use the coreceptor, and if there is any correlation between viral phenotype and its use. Also, studies of sequential virus isolates may reveal correlations between the acquisition of the ability to use a given coreceptor and clinical aspects of HIV-1 disease, such as the development of neurological symptoms. Ultimately, identification of the ligands for orphan receptors and the development of specific antibodies should make it possible to identify the coreceptors that can be used by HIV-1 to enter various cell types.

The expression of a given coreceptor in conjunction with CD4 is not necessarily sufficient for virus infection. For example, CXCR4 is expressed on the surface of macrophages, which are resistant to infection by virus strains which use this receptor. Interestingly, infection of macrophages obtained from CCR5-negative individuals by a dual-tropic virus isolate, 89.6, is inhibited by the CXCR4 ligand SDF-1, suggesting that CXCR4 can be used as a coreceptor on the surface of macrophages, albeit rarely [Yi et al.(1998)]. There are a variety of reasons why CXCR4 may not generally support entry of HIV-1 into macrophages, including differential posttranslational processing, surface expression levels relative to CD4 (as discussed above), and the way in which CXCR4 is presented in relation to CD4. At present, very little is known about the architecture of the fusion site - the number of molecules of env, CD4, and coreceptor that are required for a productive membrane fusion reaction, and their spatial relationships relative to one another.

Summary

The discovery of the HIV-1 coreceptors has important implications for understanding viral tropism and pathogenesis. Identifying which coreceptors are used by particular virus strains, and then determining whether their use correlates with particular aspects of viral pathogenesis, will be of general interest. Thus, a list of coreceptors used by virus strains will be maintained and updated yearly in this database. In addition, this will provide a forum where recent advances in the field can be discussed, particularly with regards as to how coreceptor usage patterns can be determined and our understanding of the significance of receptors other than CCR5 and CXCR4. If other receptors are found to be important for viral pathogenesis, then the classification scheme that we have proposed can be modified to take these new findings into account.

Acknowledgements

The authors thank Satish Pillai and Brian Foley for carefully reviewing the Table, for determining the accession numbers, and for their help and advice.

Strain	Proposed Accession ^a	Primary Clade	Other Tropism ^b	Designation ^c	Receptor ^d	Receptors ^e	References
DJ258	L22939	A	NSI	R5	CCR5		[Trikala et al.(1997)]
92RW026	NA	A	NSI	R5	CCR5		[Trikala et al.(1997)]
93KE101	NA	A	NSI	R5	CCR5		[Zhang et al.(1996)]
93IN103	NA	A	NSI	R5	CCR5		[Zhang et al.(1996)]
92UG037-8	U51190	A	NSI	R5	CCR5	CCR8*	[Bjorndal et al.(1997), Rucker et al.(1997)]
92RW020-5	U08794	A	NSI	R5	CCR5		[Rucker et al.(1997)]
92UG131	L34667	A	NSI	R5	CCR5		[Dittmar et al.(1997)]
92RW20	U08794	A	NSI	R5	CCR5		[Dittmar et al.(1997)]
92UG029	NA	A	SI	X4	CXCR4		[Trikala et al.(1997)]
92RW009	U88823	A	SI	R5X4	CXCR4, CCR5		[Zhang et al.(1996)]
JR-FL	U63632	B	NSI	R5	CCR5	CCR3	[Deng et al.(1997), Farzan et al.(1997), Rucker et al.(1997)]
JR-CSF	M38429	B	NSI	R5	CCR5		[Simmons et al.(1996), Trikala et al.(1997), Zhang et al.(1996)]
SF162	M65024	B	NSI	R5	CCR5	STRL33*	[Liao et al.(1997), Rucker et al.(1997)]
YU2	M93258	B	NSI	R5	CCR5	CCR3, GPR15	[Choe et al.(1996), Farzan et al.(1997)]
ADA	AF004394	B	NSI	R5	CCR5	CCR3, GPR15, STRL33*, CCR8	[Choe et al.(1996), Farzan et al.(1997), Rucker et al.(1997)]
Ba-L	M68893	B	NSI	R5	CCR5	CCR3, STRL33*	[Deng et al.(1997), Dragic et al.(1996), Liao et al.(1997), Rucker et al.(1997)]
92US657	U04908	B	NSI	R5	CCR5		[Trikala et al.(1997)]
92US715.6	U08451	B	NSI	R5	CCR5		[Bjorndal et al.(1997)]
92Br20-4	U08797	B	NSI	R5	CCR5		[Choe et al.(1996), Rucker et al.(1997)]
91US005.11	U27434	B	NSI	R5	CCR5		[Bjorndal et al.(1997), Rucker et al.(1997)]
SL-2	NA	B	NSI	R5	CCR5		[Simmons et al.(1996)]
92TH014.12	U08801	B	NSI	R5	CCR5		[Bjorndal et al.(1997)]
CM243	NA	B	NSI	R5	CCR5	GPR15, STRL33	[Rucker et al.(1997)]
M23	NA	B	NSI	R5	CCR5		[Dittmar et al.(1997)]
B80	NA	B	NSI	R5	CCR5		[Dittmar et al.(1997)]
BR92	NA	B	NSI	R5	CCR5		[Dittmar et al.(1997)]
BR49	NA	B	NSI	R5	CCR5		[Dittmar et al.(1997)]
BR53	NA	B	NSI	R5	CCR5		[Dittmar et al.(1997)]
BR90	NA	B	NSI	R5	CCR5		[Dittmar et al.(1997)]
93HA593 ⁶	U08444	B	SI	R5X4	CXCR4, CCR5		[Zhang et al.(1996)]
92HT593.1 ⁸	U08444	B	NSI	R5X4	CXCR4, CCR5		[Bjorndal et al.(1997)]

HIV-1 Coreceptor Use

Strain	Proposed Accession ^a	Primary Clade	Other Tropism ^b	Designation ^c	Receptor ^d	Receptors ^e	References
2028	NA	B	SI	R5X4	CXCR4, CCR5	CCR3	[Dittmar et al.(1997), Simmons et al.(1996)]
2076	NA	B	SI	R5X4	CXCR4, CCR5		[Dittmar et al.(1997), Simmons et al.(1996), Trkola et al.(1997)]
89.6	U39362	B	SI	R5X4	CXCR4, CCR5	CCR3, CCR2b, CCR8, V28	[Farzan et al.(1997), Rucker et al.(1997)]
DH123	NA	B	SI	R5X4	CXCR4, CCR5		[Trkola et al.(1997)]
Isolate C 7/86	NA	B	SI	R5X4	CXCR4, CCR5		[Trkola et al.(1997)]
92HA594	U08445	B	SI	R5X4	CXCR4, CCR5		[Zhang et al.(1996)]
92HA596	U08446	B	SI	R5X4	CXCR4, CCR5		[Zhang et al.(1996)]
M13	NA	B	SI	R5X4	CXCR4, CCR5		[Simmons et al.(1996)]
2006	NA	B	SI	R5X4	CXCR4, CCR5		[Simmons et al.(1996)]
2044	NA	B	SI	R5X4	CXCR4, CCR5		[Simmons et al.(1996)]
2036	NA	B	SI	R5X4	CXCR4, CCR5		[Simmons et al.(1996)]
2005	NA	B	SI	R5X4	CXCR4, CCR5		[Simmons et al.(1996)]
92HT599.24	U08447	B	SI	X4	CXCR4		[Bjornedal et al.(1997)]
BK132	L03697	B	SI	X4	CXCR4	CCR3*, CCR8*	[Rucker et al.(1997)]
BR65	NA	B	SI	X4	CXCR4		[Dittmar et al.(1997)]
HC4	NA	B	SI	X4	CXCR4		[Trkola et al.(1997)]
SF2	K02007	B	SI/TCLA	R5X4	CXCR4, CCR5		[Trkola et al.(1997)]
RF	M17451	B	SI/TCLA	R5X4	CXCR4, CCR5		[Alkhanb et al.(1996), Deng et al.(1997), Rucker et al.(1997)]
NL 4-3	M19921	B	SI/TCLA	X4	CXCR4		[Trkola et al.(1997), Zhang et al.(1996)]
LA1	X01762	B	SI/TCLA	X4	CXCR4		[Trkola et al.(1997)]
HXBc2	K03455	B	SI/TCLA	X4	CXCR4		[Choe et al.(1996)]
GUN-1	D34590	B	SI/TCLA	R5X4	CXCR4, CCR5		[Simmons et al.(1996)]
BH8	K02011	B	SI/TCLA	X4	CXCR4	CCR3*, STRL33*	[Rucker et al.(1997)]
92ZW101	NA	C	NSI	R5	CCR5		[Zhang et al.(1996)]
92BR025.9	U52953	C	NSI	R5	CCR5		[Bjornedal et al.(1997), Dittmar et al.(1997)]
BR28	U16217	C	NSI	R5	CCR5		[Dittmar et al.(1997)]
93MW965.26	U08455	C	NSI	R5	CCR5		[Bjornedal et al.(1997)]
BR70	NA	C	NSI	R5	CCR5		[Dittmar et al.(1997)]
JW1	NA	C	NSI	R5	CCR5		[Dittmar et al.(1997)]
JW4	NA	C	NSI	R5	CCR5		[Dittmar et al.(1997)]
92ZW102	NA	C	NSI	R5	CCR5		[Zhang et al.(1996)]

HIV-1 Coreceptor Use

Strain	Proposed Accession ^a	Primary Clade	Other Tropism ^b	Designation ^c	Receptor ^d	Receptors ^e	References
DJ259	L22940	C	NSI	R5	CCR5		[Trkola et al.(1997)]
94ZW103	NA	C	NSI	R5	CCR5		[Trkola et al.(1997)]
94ZW109	NA	C	NSI	R5	CCR5		[Trkola et al.(1997)]
92ZW106	NA	C	SI	X4	CXCR4		[Zhang et al.(1996)]
ZAM20	L22956	C	SI	X4	CXCR4		[Trkola et al.(1997)]
94ZW106	NA	C	SI	X4	CXCR4		[Trkola et al.(1997)]
94KE102	NA	D	NSI	R5	CCR5		[Trkola et al.(1997), Zhang et al.(1996)]
94KE103	NA	D	NSI	R5	CCR5		[Trkola et al.(1997), Zhang et al.(1996)]
92UG046	U08737	D	SI	X4	CXCR4		[Trkola et al.(1997)]
UG270	NA	D	SI	X4	CXCR4		[Trkola et al.(1997)]
92UG024.2	U08726	D	SI	X4	CXCR4	CCR8*, V28*, CCR3 ^f	[Bjorndal et al.(1997), Rucker et al.(1997), Trkola et al.(1997)]
92UG021.16	U27399	D	SI	X4	CXCR4		[Bjorndal et al.(1997)]
FW5	NA	D	SI	X4	CXCR4		[Dittmar et al.(1997)]
NDK	M27323	D	SI	X4	CXCR4	GPR15	[Deng et al.(1997), Pleskoff et al.(1997a)]
93ZR001.3	U27419	D	NA	X4	CXCR4	V28*	[Rucker et al.(1997)]
CM235	NA	E	NSI	R5	CCR5		[Trkola et al.(1997)]
92TH001	NA	E	NSI	R5	CCR5		[Trkola et al.(1997)]
M53	NA	E	NSI	R5	CCR5		[Dittmar et al.(1997)]
92TH22	U09131	E	NSI	R5	CCR5		[Dittmar et al.(1997)]
92TH23	NA	E	NSI	R5	CCR5		[Dittmar et al.(1997)]
C2	NA	E	NSI	R5	CCR5		[Dittmar et al.(1997)]
93TH305	NA	E	NSI	R5	CCR5		[Zhang et al.(1996)]
93TH307	NA	E	NSI	R5	CCR5		[Zhang et al.(1996)]
93TH966.8	U08456	E	NSI	R5	CCR5		[Bjorndal et al.(1997)]
93TH976.17	U08458	E	NA	R5	CCR5		[Bjorndal et al.(1997)]
93TH304	NA	E	SI	R5X4	CXCR4, CCR5		[Zhang et al.(1996)]
SL6	NA	E	SI	X4	CXCR4		[Dittmar et al.(1997)]
SL7	NA	E	SI	X4	CXCR4		[Dittmar et al.(1997)]
SL8	NA	E	SI	X4	CXCR4		[Dittmar et al.(1997)]
94TH304	NA	E	SI	X4	CXCR4		[Trkola et al.(1997)]
BR58	NA	F	SI	R5X4	CXCR4, CCR5	CCR3	[Dittmar et al.(1997)]
BZ162	L22084	F	NSI	R5	CCR5		[Trkola et al.(1997)]
R1	NA	F	NSI	R5	CCR5		[Trkola et al.(1997)]
93BR029.2	U27413	F	NA	R5	CCR5		[Rucker et al.(1997)]
92UG975.10	U27426	G	NSI	R5	CCR5		[Bjorndal et al.(1997)]
CA9	NA	O	NSI	R5	CCR5		[Zhang et al.(1996)]
MVP5180	L20571	O	SI	R5X4	CXCR4, CCR5		[Zhang et al.(1996)]

Footnotes:

^a Accession numbers refer to sequences from the described isolate that contain the longest available sequence including Env (full length genome when available, down to a minimum of the V3 region), but may not pertain directly to the sample used to determine the isolate's phenotype. Accession numbers have been deemed "NA" if there are no related locus names in the HIV Database, or if there is insufficient information to conclude that a similarly named locus in the database is referring to the same isolate.

^b The tropism for each virus strain is indicated. SI = syncytium inducing; NSI = non syncytium inducing; SUTCLA = syncytium inducing T-cell line adapted; NA = not available.

^c As described in the text, we have proposed that viruses which use CCR5 be termed R5 viruses, viruses that use CXCR4 be termed X4 viruses, and viruses that use both be termed R5X4 viruses. For a virus to be termed an R5X4 virus, CXCR4 must be used at > 10% of the efficiency of CCR5 for viruses that use CCR5 as their primary receptor, or CCR5 must be used at > 10% of the efficiency of CXCR4 for viruses that use CXCR4 as their primary receptor.

^d All virus strains in the Table have been tested for the ability to use CCR5 and CXCR4. The receptors used by each strain are indicated. For both receptors to be listed, the least efficiently used receptor must support virus entry by > 10% of the levels supported by the most efficiently used receptor. For viruses that use both receptors, CXCR4 is listed first; this does not mean that CXCR4 is used more efficiently than CCR5.

^e Receptors other than CCR5 or CXCR4 that are used by the indicated strains are listed. Since most viruses have not yet been tested for the ability to use receptors other than CCR5 and CXCR4, the absence of other receptors used in the Table should not be taken to mean that a virus uses only CCR5 or CXCR4. For example, if CCR3 is not listed as being used by a given virus strain, it may be because it has not yet been tested. Receptors that have been shown by experiments not to be used by a given virus strain are not indicated. Occasionally there are discrepancies in the literature concerning the use of one or more receptors. In these cases, we opted for 'majority rules'. If two papers report a positive result and one a negative result, the receptor is shown as being used.

^f The virus isolate used CCR3 whereas the cloned env from this isolate did not (Bjornidal et al.(1997)).

*Indicates that coreceptor use was determined only by fusion assays rather than by virus infection.

^g 92HT593.1 (cloned env gene) and 92HA593 (whole virus) come from the same isolate and gave slightly different results in references [Dittmar et al.(1997)] and [Zhang et al.(1996)].

References

- [Alkhatib et al.(1996)] G. Alkhatib, C. Combadiere, C. C. Broder, Y. Feng, P. E. Kennedy, P. M. Murphy, & E. A. Berger. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 272:1955-8, 1996.
- [Balotta et al.(1997)] C. Balotta, P. Bagnarelli, M. Violin, A. L. Ridolfo, D. Zhou, A. Berlusconi, S. Corvasco, M. Corbellino, M. Clementi, M. Clerici, M. Moroni, & M. Galli. Homozygous delta 32 deletion of the CCR-5 chemokine receptor gene in an HIV-1-infected patient. *AIDS* 11:F67-71, 1997.
- [Berger(1997)] E. A. Berger. HIV1 entry and tropism: the chemokine receptor connection. *AIDS* 11 (supplement A):S3-S16, 1997.
- [Berger et al.(1997)] E. A. Berger, R. W. Doms, E.-M. Fenyo, B. T. M. Korber, D. R. Littman, J. P. Moore, Q. J. Sattentau, H. Schuitemaker, J. Sodroski, & R. A. Weiss. HIV-1 phenotypes classified by coreceptor usage. *Nature* In press, 1997.
- [Biti et al.(1997)] R. Biti, R. French, J. Young, B. Bennetts, G. Stewart, & T. Liang. HIV-1 infection in an individual homozygous for the CCR5 deletion allele (letter; comment). *Nat Med* 3:252-3, 1997.
- [Bjornidal et al.(1997)] A. Bjornidal, H. Deng, M. Jansson, J. R. Fiore, C. Colognesi, A. Karlsson, J. Albert, G. Scarlatti, D. R. Littman, & E. M. Fenyo. Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. *J Virol* 71:7478-87, 1997.

- [Brelot et al.(1997)] A. Brelot, N. Heveker, O. Pleskoff, N. Sol, & M. Alizon. Role of the first and third extracellular domains of CXCR-4 in human immunodeficiency virus coreceptor activity. *J Virol* 71:4744-51, 1997.
- [Broder & Collman(1997)] C. C. Broder & R. G. Collman. Chemokine receptors and HIV. *J Leukoc Biol* 62:20-9, 1997.
- [Choe et al.(1996)] H. Choe, M. Farzan, Y. Sun, N. Sullivan, B. Rollins, P. D. Ponath, L. Wu, C. R. Mackay, G. LaRosa, W. Newman, N. Gerard, C. Gerard, & J. Sodroski. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* 85:1135-48, 1996.
- [Collman et al.(1992)] R. Collman, J. W. Balliet, S. A. Gregory, H. Friedman, D. L. Kolson, N. Nathanson, & A. Srinivasan. An infectious molecular clone of an unusual macrophage-tropic and highly cytopathic strain of human immunodeficiency virus type 1. *J Virol* 66:7517-21, 1992.
- [Connor et al.(1993)] R. I. Connor, H. Mohri, Y. Cao, & D. D. Ho. Increased viral burden and cytopathicity correlate temporally with CD4+ T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J Virol* 67:1772-7, 1993.
- [Dean et al.(1996)] M. Dean, M. Carrington, C. Winkler, G. A. Huttley, M. W. Smith, R. Allikmets, J. J. Goedert, S. P. Buchbinder, E. Vittinghoff, E. Gomperts, S. Donfield, D. Vlahov, R. Kaslow, A. Saah, C. Rinaldo, R. Detels, & S. J. O'Brien. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study (see comments) (published erratum appears in Science 1996 Nov 15;274(5290):1069). *Science* 273:1856-62, 1996.
- [Deng et al.(1997)] H. K. Deng, D. Unutmaz, V. N. KewalRamani, & D. R. Littman. Expression cloning of new receptors used by simian and human immunodeficiency viruses (see comments). *Nature* 388:296-300, 1997.
- [Dittmar et al.(1997)] M. T. Dittmar, G. Simmons, S. Hibbitts, M. O'Hare, S. Louisiziritchanakul, S. Beddows, J. Weber, P. R. Clapham, & R. A. Weiss. Langerhans cell tropism of human immunodeficiency virus type 1 subtype A through F isolates derived from different transmission groups. *J Virol* 71:8008-13, 1997.
- [Doms & Peiper(1997)] R. W. Doms & S. C. Peiper. Unwelcomed guests with master keys: how HIV uses chemokine receptors for cellular entry. *Virology* 235:179-90, 1997.
- [Doranz et al.(1996)] B. J. Doranz, J. Rucker, Y. Yi, R. J. Smyth, M. Samson, S. C. Peiper, M. Parmentier, R. G. Collman, & R. W. Doms. A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell* 85:1149-58, 1996.
- [Dragic et al.(1996)] T. Dragic, V. Litwin, G. P. Allaway, S. R. Martin, Y. Huang, K. A. Nagashima, C. Cayanan, P. J. Maddon, R. A. Koup, J. P. Moore, & W. A. Paxton. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC- CKR-5 (see comments). *Nature* 381:667-73, 1996.
- [Edinger et al.(1997)] A. L. Edinger, J. L. Mankowski, B. J. Doranz, B. J. Margulies, B. Lee, J. Rucker, M. Sharon, T. L. Hoffman, J. F. Berson, M. C. Zink, V. M. Hirsch, J. E. Clements, & R. W. Doms. CD4-independent, CCR5-dependent infection of brain capillary endothelial cells by neurovirulent SIV. *Proc. Natl. Acad. Sci. USA* In press, 1997.
- [Endres et al.(1996)] M. J. Endres, P. R. Clapham, M. Marsh, M. Ahuja, J. D. Turner, A. McKnight, J. F. Thomas, B. Stoeckenau-Haggarty, S. Choe, P. J. Vance, T. N. Wells, C. A. Power, S. S. Sutterwala, R. W. Doms, N. R. Landau, & J. A. Hoxie. CD4-independent infection by HIV-2 is mediated by fusin/CXCR4. *Cell* 87:745-56, 1996.
- [Farzan et al.(1997)] M. Farzan, H. Choe, K. Martin, L. Marcon, W. Hofmann, G. Karlsson, Y. Sun, P. Barrett, N. Marchand, N. Sullivan, N. Gerard, C. Gerard, & J. Sodroski. Two orphan seven-transmembrane segment receptors which are expressed in CD4-positive cells support simian immunodeficiency virus infection. *J Exp Med* 186:405-11, 1997.

HIV-1 Coreceptor Use

- [Feng et al.(1996)] Y. Feng, C. C. Broder, P. E. Kennedy, & E. A. Berger. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor (see comments). *Science* 272:872-7, 1996.
- [Fenyo et al.(1988)] E. M. Fenyo, L. Morfeldt-Manson, F. Chiodi, B. Lind, G. von A., J. Albert, E. Olausson, & B. Asjo. Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *J Virol* 62:4414-9, 1988.
- [Fouchier et al.(1992)] R. A. Fouchier, M. Groenink, N. A. Koolstra, M. Tersmette, H. G. Huisman, F. Miedema, & H. Schuitemaker. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol* 66:3183-7, 1992.
- [He et al.(1997)] J. He, Y. Chen, M. Farzan, H. Choe, A. Ohagen, S. Gartner, J. Busciglio, X. Yang, W. Hofmann, W. Newman, C. R. Mackay, J. Sodroski, & D. Gabozda. CCR3 and CCR5 are coreceptors for HIV-1 infection of microglia. *Nature* 385:645-9, 1997.
- [J. et al.(1992)] d. J. J., J. Goudsmit, W. Keulen, B. Klaver, W. Krone, M. Tersmette, & d. A. Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. *J Virol* 66:757-65, 1992.
- [Kozak et al.(1997)] S. L. Kozak, E. J. Platt, N. Madani, F. E. F. Jr, K. Peden, & D. Kabat. CD4, CXCR-4, and CCR-5 dependencies for infections by primary patient and laboratory-adapted isolates of human immunodeficiency virus type 1. *J Virol* 71:873-82, 1997.
- [Liao et al.(1997)] F. Liao, G. Alkhatib, K. W. Peden, G. Sharma, E. A. Berger, & J. M. Farber. STRL33, A novel chemokine receptor-like protein, functions as a fusion cofactor for both macrophage-tropic and T cell line-tropic HIV-1. *J Exp Med* 185:2015-23, 1997.
- [Liu et al.(1996)] R. Liu, W. A. Paxton, S. Choe, D. Ceradini, S. R. Martin, R. Horuk, M. E. MacDonald, H. Stuhlmann, R. A. Koup, & N. R. Landau. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86:367-77, 1996.
- [Lu et al.(1997)] Z. Lu, J. F. Berson, Y. Chen, J. D. Turner, T. Zhang, M. Sharron, M. H. Jenks, Z. Wang, J. Kim, J. Rucker, J. A. Hoxie, S. C. Peiper, & R. W. Doms. Evolution of HIV-1 coreceptor usage through interactions with distinct CCR5 and CXCR4 domains. *Proc Natl Acad Sci USA* 94:6426-31, 1997.
- [Michael et al.(1997)] N. L. Michael, G. Chang, L. G. Louie, J. R. Mascola, D. Dondero, D. L. Birx, & H. W. Sheppard. The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression. *Nat Med* 3:338-40, 1997.
- [Moore(1997)] J. P. Moore. Coreceptors: implications for HIV pathogenesis and therapy. *Science* 276:51-2, 1997.
- [O'Brien et al.(1997)] T. R. O'Brien, C. Winkler, M. Dean, J. A. Nelson, M. Carrington, N. L. Michael, & G. C. W. 2nd. HIV-1 infection in a man homozygous for CCR5 delta 32 (letter). *Lancet* 349:1219, 1997.
- [Pleskoff et al.(1997a)] O. Pleskoff, N. Sol, B. Labrosse, & M. Alizon. Human immunodeficiency virus strains differ in their ability to infect CD4+ cells expressing the rat homolog of CXCR-4 (fusin). *J Virol* 71:3259-62, 1997a.
- [Pleskoff et al.(1997b)] O. Pleskoff, C. Treboute, A. Brelot, N. Heveker, M. Seman, & M. Alizon. Identification of a chemokine receptor encoded by human cytomegalovirus as a cofactor for HIV-1 entry (see comments). *Science* 276:1874-8, 1997b.
- [Rana et al.(1997)] S. Rana, G. Besson, D. G. Cook, J. Rucker, R. J. Smyth, Y. Yi, J. D. Turner, H. H. Guo, J. G. Du, S. C. Peiper, E. Lavi, M. Samson, F. Libert, C. Liesnard, G. Vassart, R. W. Doms, M. Parmentier, & R. G. Collman. Role of CCR5 in infection of primary macrophages and lymphocytes

- by macrophage-tropic strains of human immunodeficiency virus: resistance to patient-derived and prototype isolates resulting from the delta ccr5 mutation. *J Virol* 71:3219-27, 1997.
- [Reeves et al.(1997)] J. D. Reeves, A. McKnight, S. Potempa, G. Simmons, P. W. Gray, C. A. Power, T. Wells, R. A. Weiss, & S. J. Talbot. CD4-independent infection by HIV-2 (ROD/B): use of the 7-transmembrane receptors CXCR-4, CCR-3, and V28 for entry. *Virology* 231:130-4, 1997.
- [Roos et al.(1992)] M. T. Roos, J. M. Lange, d. R. E., R. A. Coutinho, P. T. Schellekens, F. Miedema, & M. Tersmette. Viral phenotype and immune response in primary human immunodeficiency virus type 1 infection. *J Infect Dis* 165:427-32, 1992.
- [Rucker et al.(1997)] J. Rucker, A. L. Edinger, M. Sharron, M. Samson, B. Lee, J. F. Berson, Y. Yi, B. Margulies, R. G. Collman, B. J. Doranz, M. Parmentier, & R. W. Doms. Utilization of chemokine receptors, orphan receptors, and herpesvirus encoded receptors by diverse human and simian immunodeficiency viruses. *J Virol* 71:8999-9007, 1997.
- [Samson et al.(1997)] M. Samson, A. L. Edinger, S. P. J. Rucker, V. Verhasselt, M. Sharron, C. Govaerts, C. Mollereau, G. Vassart, R. W. Doms, & M. Parmentier. ChemR23, a putative chemoattractant receptor, is expressed in dendritic cells and is a coreceptor for SIV and some primary HIV-1 strains. Submitted, 1997.
- [Samson et al.(1996)] M. Samson, F. Libert, B. J. Doranz, J. Rucker, C. Liesnard, C. M. Farber, S. Saragosti, C. Lapoumeronie, J. Cognaux, C. Forceille, G. Muyldermans, C. Verhofstede, G. Burtonboy, M. Georges, T. Inai, S. Rana, Y. Yi, R. J. Smyth, R. G. Collman, R. W. Doms, G. Vassart, & M. Parmentier. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene (see comments). *Nature* 382:722-5, 1996.
- [Schuitemaker et al.(1992)] H. Schuitemaker, M. Koot, N. A. Kootstra, M. W. Dercksen, d. R. E., S. van R. P., J. M. Lange, J. K. Schattenkerk, F. Miedema, & M. Tersmette. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocyctotropic to T-cell-tropic virus population. *J Virol* 66:1354-60, 1992.
- [Simmons et al.(1996)] G. Simmons, D. Wilkinson, J. D. Reeves, M. T. Dittmar, S. Beddows, J. Weber, G. Carnegie, U. Desselberger, P. W. Gray, R. A. Weiss, & P. R. Clapham. Primary, syncytium-inducing human immunodeficiency virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. *J Virol* 70:8355-60, 1996.
- [Tersmette et al.(1988)] M. Tersmette, d. R. E., B. J. Al, I. N. Winkel, R. A. Gruters, H. T. Cuypers, H. G. Huisman, & F. Miedema. Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. *J Virol* 62:2026-32, 1988.
- [Tersmette et al.(1989)] M. Tersmette, J. M. Lange, d. R. E., W. F., J. K. Eeftink-Schattenkerk, P. T. Schellekens, R. A. Coutinho, J. G. Huisman, J. Goudsmit, & F. Miedema. Association between biological properties of human immunodeficiency virus variants and risk for AIDS and AIDS mortality. *Lancet* 1:983-5, 1989.
- [Theodorou et al.(1997)] I. Theodorou, L. Meyer, M. Magierowska, C. Katlama, & C. Rouzioux. HIV-1 infection in an individual homozygous for CCR5 delta 32. Serocon Study Group (letter). *Lancet* 349:1219-20, 1997.
- [Trkola et al.(1996)] A. Trkola, T. Dragic, J. Arthos, J. M. Binley, W. C. Olson, G. P. Allaway, C. Cheng-Mayer, J. Robinson, P. J. Maddon, & J. P. Moore. CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5 (see comments). *Nature* 384:184-7, 1996.
- [Trkola et al.(1997)] A. Trkola, W. A. Paxton, S. P. Monard, J. A. Hoxie, M. A. Siani, D. A. Thompson, L. Wu, C. R. Mackay, R. Horuk, & J. P. Moore. Genetic subtype-independent inhibition of human immunodeficiency virus type 1 replication by CC and CXCR-chemokines. Submitted, 1997.

HIV-1 Coreceptor Use

- [Wu et al.(1996)] L. Wu, N. P. Gerard, R. Wyatt, H. Choe, C. Parolin, N. Ruffing, A. Borsetti, A. A. Cardoso, E. Desjardin, W. Newman, C. Gerard, & J. Sodroski. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5 (see comments). *Nature* 384:179-83, 1996.
- [Yi et al.(1998)] Y. Yi, J. D. Turner, N. Gaddis, & R. G. Collman. Cxcr4 is expressed by primary macrophages and supports ccr5-independent infection by dual-tropic but not t-tropic isolates of hiv-1. *J Virol* Submitted, 1998.
- [Zhang et al.(1997)] L. Zhang, C. D. Carruthers, T. He, Y. Huang, Y. Cao, G. Wang, B. Hahn, & D. D. Ho. Hiv type 1 subtypes, coreceptor usage, and ccr5 polymorphism. *AIDS Res Hum Retroviruses* 13:1357-66, 1997.
- [Zhang et al.(1996)] L. Zhang, Y. Huang, T. He, Y. Cao, & D. D. Ho. HIV-1 subtype and second-receptor use (letter). *Nature* 383:768, 1996.
- [Zhu et al.(1993)] T. Zhu, H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup, & D. D. Ho. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science* 261:1179-81, 1993.

The History of HIV-1 Biological Phenotypes Past, Present and Future

Eva Maria Fenyo,¹ Hanneke Schuitemaker,² Birgitta Åsjo,³ Jane McKeating,⁴
Quentin Sattentau,⁵ and the EC Concerted Action HIV Variability⁶

¹ Microbiology and Tumorbiology Center, Karolinska Institute, 171 77 Stockholm, Sweden

² Department of Clinical Viro Immunology, Central Laboratory of the Netherlands Red Cross Blood Transfusion Service and Laboratory for Experimental and Clinical Immunology, University of Amsterdam, Amsterdam, The Netherlands

³ Center for Research in Virology, University of Bergen, 5020 Bergen, Norway

⁴ Department of Microbiology, University of Reading, RG6 2AH Reading, UK

⁵ Centre d'Immunologie de Marseille-Luminy, Case 906, 13288 Marseille, France

⁶ Active participants of the EC Concerted Action HIV Variability: Karolinska Institute (Stockholm, Sweden): EM. Fenyo/Coordinator; Chester Beatty laboratories (London, United Kingdom): Robin A. Weiss, A. McKnight; University of Liverpool (Liverpool, UK): Thomas Schulz; Centre d'Immunologie de Marseille-Luminy (Marseille, France): Q. Sattentau; University of Reading (Reading, United Kingdom): J. McKeating; Univ. College and Middlesex School of Medicine (London, United Kingdom): P. Balfe; University of Bergen (Bergen, Norway): B. Åsjo; Natl. Public Health Institute (Helsinki, Finland): P. Leinikki, M. Salminen; Inst. of Medical Microbiology and Hygiene (Freiburg, Germany): A. Meyerhans; Central Laboratory of the Netherlands Red Cross Blood Transfusion Service (Amsterdam, The Netherlands): F. Miedema, H. Schuitemaker; Swedish Inst. for Infections Disease Control (Stockholm, Sweden): J. Albert; Instituto Nacional de Saude (Lisboa, Portugal): F. Avelaz, N. Taveira; University of Padova (Padova, Italy): L. Chieco Bianchi, A. De Rossi; University of Edinburgh (Edinburgh, United Kingdom): A. Leigh Brown.

HIV-1 primary isolates have been classified according to their behavior in tissue culture into two distinct groups based on their growth kinetics, cytopathology and cellular tropism. Classification by growth kinetics and cytopathology is indicative of replication rate and syncytium induction in activated peripheral blood mononuclear cells (PBMC). Accordingly, the terms slow/low and rapid/high for replication pattern and non-syncytium inducing (NSI) and syncytium inducing (SI) for cytopathology were introduced [1-3]. In general, the terms SI and NSI correspond broadly to the rapid/high and slow/low phenotypes, respectively [4, 5]. A currently used method to distinguish between these HIV-1 phenotypes is based on the ability of a virus to infect and induce syncytia in the MT-2 T cell line [6]. These two groups of viruses are also distinct at the genetic level, SI viruses generally have a greater positive charge within the third variable region of the envelope glycoprotein than NSI viruses [7-9]. A third phenotype is defined by tropism and includes the terms T cell tropic, macrophage-tropic, dual-tropic [10] and T cell line-adapted (TCLA, reviewed in [11, 12]). Classification of primary HIV-1 isolates by tropism is, however, not as clear a classification as by replicative capacity and cytopathology, since conflicting data exist concerning macrophage tropism of several reported HIV-1 isolates. Macrophage tropism, as defined by the ability of the virus to infect and replicate in primary monocyte/macrophage cultures has been described either as a feature of primary viruses that mostly lack T cell line tropism [13] or as a general characteristic of HIV-1 isolates [10, 14]. Another report describes tropism as a continuous spectrum, ranging from viruses that infect macrophages with high efficiency and T cell lines with very low efficiency, to the converse [15]. Clearly, more experimental data are required to clarify this point. Finally, adaptation to growth in T cell lines changes the biological properties of primary HIV-1 isolates resulting in a variety of phenotypic and genotypic changes leading to altered cellular tropism and increased sensitivity to neutralization by antibody and soluble receptor [12, 16, 17].

The classification of SI and NSI, which corresponds broadly to the rapid/high and slow/low phenotypes, respectively, (see Table 1) is not as simple as it first appears. For example, NSI (slow/low) viruses are able to grow as rapidly as SI (rapid/high) viruses and form syncytia in primary cultures of CD4+ T cells depleted of CD8+ T cells ([18, 19] and unpublished results). Rapid/high viruses are similar

HIV-1 Phenotypes

to SI viruses in that both are able to replicate in CD4+ cell lines of monocytoid or T-cell origin, and are usually isolated from late-stage, immunodeficient patients. Yet in cases where NSI viruses are isolated from AIDS patients, these viruses grow rapidly in PBMC cultures but do not infect monocytoid or T-cell lines. However, it has to be remembered that in spite of these complicating factors, the biological properties of HIV-1 have allowed the definition of virulence markers [20-22]. A more precise definition of such phenotypic traits associated with increased HIV-1 virulence will be of importance.

Table 1. Classification of HIV-1 biological phenotypes

Chemokine receptor usage	New Classification	previous terminology based on			
		cytopathology in MT-2 cells	replication rate in PBMC	tropism	
				concept 1	concept 2
CXCR4	X4	syncytium-inducing	rapid/high	T cell tropic T cell line-adapted	dual-tropic*
CCR5 CCR3/CCR2b	R5 R3/R2b	non-syncytium-inducing	slow/low	macrophage-tropic	dual-tropic*
CXCR4 & CCR5 and/or CCR3	X4R5 X4R5R3 X4R3	syncytium-inducing	rapid/high	T cell tropic	dual-tropic*
Bonzo, Bob and others	yet to be named	unclassified	unclassified	unknown	

* Dual tropism for macrophages and lymphocytes as defined by Valentin et al. [10].

The discovery that HIV-1 uses members of the seven transmembrane domain chemokine receptor family as co-receptors for membrane fusion and entry (reviewed in [23]) has changed our perception about the properties of both T cell line-adapted and primary HIV-1 isolates. The two most well-defined HIV-1 co-receptors are CXCR4 and CCR5, members of the CXCR and CCR chemokine receptor subfamilies, respectively (reviewed in [24]). CXCR4 was the first HIV-1 co-receptor to be characterized, and was shown to be required for the fusion of T cell line-adapted viruses with non-human cells expressing human CD4 [25]. It is expressed on many cell types including transformed T cells, fibroblasts, primary T cells, and macrophages. Subsequently, CCR5 was shown to be the principal co-receptor for primary HIV-1 isolates with the NSI phenotype [26-30], whereas the SI phenotype was associated with the use of CXCR4 alone or in combination with CCR5 [31-33]. Other members of the CC chemokine receptor family, such as CCR2b and CCR3, may also function as co-receptors for NSI virus entry, although generally in a less efficient manner than CCR5 [26, 28, 33]. The almost complete resistance to infection by slow/low, NSI-type viruses of individuals carrying a deletion of 32 base pairs in both alleles of the CCR5 gene confirms the importance of this molecule in HIV-1 transmission *in vivo* [34-36].

The previous classification systems based on replicative and syncytium inducing capacities of primary HIV-1 isolates in PBMC or MT-2 cells can be translated into classification based on co-receptor usage, thereby creating a molecular basis for defining HIV-1 biological phenotype. Thus primary and T cell line adapted HIV-1 isolates can now be grouped on the basis of co-receptor use. The classification systems presently used and their relationship to co-receptor usage are shown in Table 1. Thus viruses previously termed SI (rapid/high) which infect activated PBMC and readily infect CD4+ T cell lines are defined by their use of the CXCR-chemokine receptor CXCR4. NSI (slow/low) viruses which preferentially infect activated PBMC would be defined by their use of members of the CC-chemokine

receptor family, principally CCR-5. According to the terminology based on co-receptor usage, CXCR4 using viruses would be termed X4 and CCR5-using viruses R5 [37]. Viruses that were previously defined as SI (rapid/high), and which use both CXCR4 and CCR5 (and/or CCR3) receptors, would be termed X4R5 (-R3) viruses. Since novel members of the chemokine receptor family which can function as HIV co-receptors continue to be cloned and characterized [38-43] it will be important to allow room for expansion within this classification scheme. A classification system, based on the clustering of a particular virus isolate or clone by the molecules required for virus entry and virus-induced cell-cell fusion should remain a clear and flexible system for the future.

Studies in which HIV-1 envelope genes have been expressed, either in the form of recombinant glycoproteins or as infectious virus, in the presence of CD4 and the appropriate co-receptor have demonstrated that both SI and NSI viruses are equally able to induce syncytium formation. Thus the terms syncytium-inducing and non-syncytium-inducing are not absolute and may only be relevant in terms of co-receptor expression levels on target cells. Similarly, both rapid/high and slow/low viruses appear to have comparable replication kinetics in transformed cell lines expressing both CCR5 and CXCR4 together ([44] and unpublished results) or separately [33, 45]. It is becoming clear that both CD4 and chemokine receptor expression levels are important factors in determining susceptibility of cells to different viruses.

The phenotypic classification system for describing the biological properties of viral isolates continues to be used in parallel with the adaptation of the new nomenclature, and should be interpreted in the context of the underlying molecular interactions. However, many aspects of the molecular relationships relating to viral phenotype are yet to be resolved. We know today that the basis of PBMC suitability for the isolation of a wide variety of HIV isolates is that activated CD4+ T lymphocytes contain populations that express either CCR5 or CXCR4 that can be infected with viruses using either or both molecules [47, 48]. In virus isolation cultures involving cultures with mixed donor PBMC [46], HIV-1 replication patterns still appear as slow or fast, low- or high-titre, respectively, and the level of expression of the relevant co-receptor in the PBMC cultures appears to determine the rate of replication [49]. Phenotyping viruses with the MT-2 cell line remains an efficient way of defining the SI/NSI phenotype. Similarly to other immortalized CD4+ T-cell lines, MT-2 cells express CXCR4 and are susceptible to CXCR4-using viruses. When using MT-2 cells to study viral phenotype, the assay should be indicated (for example, by specifying that the viruses are MT-2 negative or MT-2 positive), along with the more descriptive SI/NSI designation. Furthermore, the T-tropic and M-tropic classification should be used with care, such that a virus should only be called macrophage tropic when tested in primary monocyte/macrophage cultures. It is important to appreciate that classification by co-receptor utilization in cell lines expressing high levels of CD4 and co-receptors may not necessarily indicate that such viruses will use these co-receptors to infect primary resting cells *in vivo*. Thus, until a more complete understanding of the molecular basis for tropism is attained, we suggest that viruses should be classified according to the experimental system employed, in conjunction with the second receptor utilization patterns.

Acknowledgments

This work was completed with help of EC grants from the Concerted Action "HIV Variability" and Programme EVA.

References

- [1] Åsjö, B., L. Morfeldt-Månson, J. Albert, G. Biberfeld, A. Karlsson, K. Lidman, and E.M. Fenyö. Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet*, 1986. ii:660-662.
- [2] Fenyö, E.M., L. Morfeldt-Månson, F. Chiodi, A. Lind, A. von Gegerfelt, J. Albert, E. Olausson, and B. Åsjö. Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *J Virol*, 1988. 62:4414-4419.

- [3] Tersmette, M., R.E.Y. De Goede, B.J. Al, I. Winkel, B. Gruters, H.T. Cuypers, H.G. Huisman, and F. Miedema, Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. *J Virol*, 1988, 62:2026-2032.
- [4] Miedema, F., L. Meygaard, M. Koot, M.R. Klein, M.T.L. Roos, M. Groenink, R.A.M. Fouchier, A.B. Van't Wout, M. Tersmette, P.T.A. Schellekens, and H. Schuitemaker, Changing virus-host interactions in the course of HIV-1 infection. *Immunological Reviews*, 1994, 140:35-72.
- [5] Rübsamen-Waigmann, H., H. von Briesen, H. Holmes, Å. Björndal, B. Korber, R. Esser, S. Ranjbar, P. Tomlinson, B. Galvao-Castro, E. Karita, S. Sempala, C. Wasi, S. Osmanov, E.M. Fenylö, and the WHO Network for HIV Isolation and Characterization, Standard conditions of virus isolation reveal biological variability of HIV-1 in different regions of the world. *AIDS Research and Human Retroviruses*, 1994, 10:1401-1408.
- [6] Koot, M., A.H.V. Vos, R.P.M. Keet, R.E.Y. de Goede, W. Dercksen, F.G. Terpstra, R.A. Coutinho, F. Miedema, and M. Tersmette, HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay. *AIDS*, 1992, 6:49-54.
- [7] Fouchier, R.A.M., M. Groenink, N.A. Kootstra, M. Tersmette, H.G. Huisman, F. Miedema, and H. Schuitemaker, Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol*, 1992, 66:3183-3187.
- [8] de Jong, J.-J., A. de Ronde, W. Keulen, M. Tersmette, and J. Goudsmit, Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. *J Virol*, 1992, 66:6777-6780.
- [9] de Wolf, F., E. Hogerworst, J. Goudsmit, E.M. Fenylö, H. Rübsamen-Waigmann, H. Holmes, B. Galvao-Castro, E. Karita, S.D.K. Sempala, C. Wasi, E. Baan, F. Zorgdrager, V. Lukashov, S. Osmanov, C. Kuiken, M. Cornelissen, and the WHO Network for HIV Isolation and Characterization, Syncytium-inducing (SI) and non-syncytium inducing (NSI) capacity of human immunodeficiency virus type 1 (HIV-1) subtypes other than B: Phenotypic and genotypic characteristics. *AIDS Research and Human Retroviruses*, 1994, 10:1387-1400.
- [10] Valentin, A., J. Albert, R. Fredriksson, E.M. Fenylö, and B. Åsjö, Dual tropism for lymphocytes and mononuclear phagocytes is a common feature of primary HIV-isolates. *J Virol*, 1994, 68:6684-6689.
- [11] Levy, J.A., HIV and the pathogenesis of AIDS. 1994, Washington DC: ASM Press.
- [12] Moore, J.P. and D.D. Ho, HIV-1 neutralization: the consequences of viral adaptation to growth on transformed T cells. *AIDS*, 1995, 9(suppl A):S117-S136.
- [13] Schuitemaker, H., M. Koot, N.A. Kootstra, M.W. Dercksen, R.E.Y. De Goede, R.P. Van Steenwijk, J.M. Lange, J.K. Eeftink Schattenkerk, F. Miedema, and M. Tersmette, Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus populations. *J Virol*, 1992, 66:1354-1360.
- [14] Stent, G., G.-B. Joo, P. Kierulf, and B. Åsjö, Macrophage tropism: fact or fiction? *Journal of Leukocyte Biology*, 1997, 62:4-11.
- [15] Collin, M., P. Illei, W. James, and S. Gordon, Definition of the range and distribution of human immunodeficiency virus macrophage tropism using PCR-based infectivity measurements. *J Gen Virol*, 1994, 75:1597-1603.
- [16] Sullivan, N., Y. Sun, J. Li, W. Hofmann, and J. Sodroski, Replicative function and neutralization sensitivity of envelope glycoproteins from primary and T-cell line-passaged human immunodeficiency virus type 1 isolates. *J Virol*, 1995, 69:4413-4422.
- [17] Zhang, Y.-j., R. Fredriksson, J. McKeating, and E.M. Fenylö, Passage of HIV-1 molecular clones into different cell lines confers differential sensitivity to neutralization. *Virology*, 1997, 238:254-264.

- [18] Todd, B.J., P. Kedar, and J.H. Pope, Syncytium induction in primary CD4+ T-cell lines from normal donors by human immunodeficiency virus type 1 isolates with non-syncytium-inducing genotype and phenotype in MT-2 cells. *J Virol*, 1995. 69:7099-7105.
- [19] Tanaka, Y., Y. Koyanagi, R. Tanaka, Y. Kumazawa, T. Nishimura, and N. Yamamoto, Productive and lytic infection of human CD4+ type 1 helper T cells with macrophage-tropic human immunodeficiency virus type 1. *J Virol*, 1997 71:465-470.
- [20] Cheng-Mayer, C., D. Seto, M. Tateno, and J.A. Levy, Biological features of HIV-1 that correlate with virulence in the host. *Science*, 1988. 240:80-82.
- [21] Tersmette, M., B.A. Gruters, F. de Wolf, R.E.Y. de Goede, J.M.A. Lange, P.T.A. Schellekens, J. Goudsmit, H.G. Huismans, and F. Miedema, Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential isolates. *J Virol*, 1989. 63:2118-2125.
- [22] Fiore, J.R., Å. Björndal, K. Aperia-Peipke, M. Di Stefano, G. Angarano, G. Pastore, H. Gaines, E.M. Fenyo, and J. Albert, The biological phenotype of HIV-1 is usually retained during and after sexual transmission. *Virology*, 1994. 204:297-303.
- [23] D'Souza, M.P. and V.A. Harden, Chemokines and HIV-1 second receptors. *Nature Medicine*, 1996. 2:1293-1300.
- [24] Moser, B., Chemokines and HIV: a remarkable synergism. *Trends Microbiol.* 1997. 5(3):88-90.
- [25] Feng, Y., C.C. Broder, P.E. Kennedy, and E.A. Berger, HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*, 1996. 272:872-877.
- [26] Choe, H., M. Farzan, Y. Sun, N. Sullivan, B. Rollins, P.D. Ponath, L.J. Wu, C.R. Mackay, G. Larosa, W. Newman, N. Gerard, C. Gerard, and J. Sodroski, The beta-chemokine receptor CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*, 1996. 85:1135-1148.
- [27] Deng, H.K., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Dimarzio, S. Mamon, R.E. Sutton, C.M. Hill, C.B. Davis, S.C. Peiper, T.J. Schall, D.R. Littman, and N.R. Landau, Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, 1996. 381:661-666.
- [28] Doranz, B.J., J. Rucker, Y.J. Yi, R.J. Smyth, M. Samson, S.C. Peiper, M. Parmentier, R.G. Collman, and R.W. Doms, A dual-tropic primary hiv-1 isolate that uses fusin and the beta-chemokine receptors ckr-5, ckr-3, and ckr-2b as fusion cofactors. *Cell*, 1996. 85:1149-1158.
- [29] Dragic, T., V. Litwin, G.P. Allaway, S.R. Martin, Y.X. Huang, K.A. Nagashima, C. Cayanan, P.J. Maddon, R.A. Koup, J.P. Moore, and W.A. Paxton, HIV-1 entry into CD4(+) cells is mediated by the chemokine receptor CC-CKR-5. *Nature*, 1996. 381:667-673.
- [30] Alkhatib, G., C. Combadiere, C.C. Broder, Y. Feng, P.E. Kennedy, P.M. Murphy, and E.A. Berger, CCCKRs - A rantes, MIP-1-alpha, MIP-1-beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science*, 1996. 272:1955-1958.
- [31] Simmons, G., D. Wilkinson, J.D. Reeves, M.T. Dittmar, S. Beddows, J. Weber, G. Carnegie, U. Desselberger, P.W. Gray, R.A. Weiss, and P.R. Clapham, Primary, syncytium-inducing human immunodeficiency virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as co-receptors for virus entry. *J Virol*, 1996. 70:8355-8360.
- [32] Zhang, L., Y. Huang, T. He, Y. Cao, and D.D. Ho, HIV-1 subtype and second receptor use. *Nature*, 1996. 383:768.
- [33] Björndal, Å., H. Deng, M. Jansson, J.R. Fiore, C. Colognesi, A. Karlsson, J. Albert, G. Scarlatti, D.R. Littman, and E.M. Fenyo, Co-receptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. *J Virol*, 1997. 71:7478-7487.
- [34] Paxton, W.A., S.R. Martin, D. Tse, T.R. O'Brien, J. Skurnick, N.L. VanDevanter, N. Padian, J.P. Braun, D.P. Kotler, S.M. Wolinsky, and R.A. Koup, Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposures. *Nature Medicine*, 1996. 2:412-417.

- [35] Liu, R., W.A. Paxton, S. Choe, D. Ceradini, S.R. Martin, R. Horuk, M.E. MacDonald, H. Stuhlman, R.A. Koup, and N.R. Landau, Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*, 1996. 86:367-377.
- [36] Samson, M., F. Libert, B.J. Doranz, J. Rucker, C. Liesnard, C.-M. Farber, S. Saragosti, C. Lapoumeroulie, J. Cognaux, C. Forceille, G. Muyldermans, C. Verhofstede, G. Burtonboy, M. Georges, T. Inai, S. Rana, Y. Yi, R.J. Smyth, R.G. Collman, R.W. Doms, G. Vassart, and M. Parmentier, Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*, 1996. 382:722-725.
- [37] Berger, E.A., R.W. Doms, E.M. Fenyö, B.T.M. Korber, D.R. Littman, J.P. Moore, Q.J. Sattentau, H. Schuitemaker, J. Sodroski, and R.A. Weiss, HIV-1 phenotypes classified by Co-receptor usage. *Nature*, 1998. 391:240.
- [38] Pleskoff, O., C. Trèbute, A. Brélot, N. Heveker, M. Seman, and M. Alizon, Identification of a chemokine receptor encoded by human Cytomegalovirus as a cofactor for HIV-1 entry. *Science*, 1997. 276:1874-1878.
- [39] Reeves, J.D., A. McKnight, S. Potempa, G. Simmons, P.W. Gray, C.A. Power, T. Wells, R.A. Weiss, and S.J. Talbot, CD4-independent infection by HIV-2 (ROD/B): Use of the 7-transmembrane receptors CXCR-4, CCR3, and V28 for entry. *Virology*, 1997. 231:130-134.
- [40] Alkhatib, G., F. Liao, E.A. Berger, and J.M. Farber, A new SIV co-receptor, STRL33. *Nature*, 1997. 388:238.
- [41] Deng, H., D. Unutmaz, V.N. KewalRamani, and D.R. Littman, Expression cloning of new receptors used by simian and human immunodeficiency viruses. *Nature*, 1997. 388:296-300.
- [42] Clapham, P.R. and R.A. Weiss, Immunodeficiency viruses. Spoilt for choice of co-receptors. *Nature*, 1997. 388:230-231.
- [43] Loetscher, D.F., D.F. Amara, D.F. Oberlin, D.F. Brass, and D.F. Legler, TYMSTR, a putative chemokine receptor selectively expressed in activated T cells, exhibits HIV-1 co-receptor function. *Curr Biol*, 1997. 7:652-660.
- [44] Lusso, P., F. Cocchi, C. Balotta, P.D. Markham, A. Louie, P. Farci, R. Pal, R.C. Gallo, and M.S. Reitz, Growth of macrophage-tropic and primary human immunodeficiency virus type 1 (HIV-1) isolates in a unique CD4+ T-cell clone (PM1): Failure to downregulate CD4 and to interfere with cell-line-tropic HIV-1. *J Virol*, 1995. 69:3712-3720.
- [45] Scarlatti, G., E. Tresoldi, Å. Björndal, R. Fredriksson, C. Colognesi, H. Deng, M.S. Malnati, A. Pfebani, A.G. Siccardi, D.R. Littman, and P. Lusso, *in vivo* evolution of HIV-1 coreceptor usage and sensitivity to chemokine-mediated suppression. *Nature Medicine*, 1997. 3:1259-1265.
- [46] World, Health, and Organization, WHO Guidelines for standard HIV isolation procedures. 1994, WHO/GPA/VAD.
- [47] Bleul, C.C., L. Wu, J.A. Hoxie, T.A. Springer, and C.R. Mackay, The HIV co-receptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proc Natl Acad Sci USA*, 1997. 94:1925-1930.
- [48] Wu, L., W.A. Paxton, N. Kassam, N. Ruffing, J.B. Rottman, N. Sullivan, H. Choe, J. Sodroski, W. Newman, R.A. Koup, and C.R. Mackay, CCR5 levels and expression pattern correlate with infectability by macrophage-tropic HIV-1, *in vitro*. *J Exp Med*, 1997. 185:1681-1691.
- [49] Bleul, C.C., M. Farzan, H. Choe, C. Parolin, I. Clark-Lewis, J. Sodroski, and T.A. Springer, The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature*, 1996. 382:829-833.

Updated Proposal of Reference Sequences of HIV-1 Genetic Subtypes

Thomas Leitner,¹ Bette Korber,¹ David Robertson², Feng Gao,³ Beatrice Hahn³

¹ Theoretical Biology and Biophysics, Group T-10, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545;

² Laboratory of Structural & Genetic Information, CNRS-EP 91, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France;

³ Department of Medicine and Microbiology, University of Alabama at Birmingham, 701 S. 19th Street, LHRB 613, Birmingham, AL 35294.

The known universe of HIV-1 viruses is divided into two groups, a major "M" group and an outlier "O" group, (reviewed in [9]). HIV-1 subtypes are clusters of sequences within the M group that are defined by phylogenetic analysis [6, 12]. In the *Human Retroviruses and AIDS 1996* compendium, a reference set of HIV-1 sequences representing the different subtypes was proposed. Since that issue of the compendium, many more full length sequences from the various HIV-1 M group subtypes have been produced. Partly as a consequence of the availability of these new sequences, and partly due to further analysis of pre-existing data, new information has been accumulating concerning subtyping and HIV-1 alignments and hybrid genomes [3, 6, 8, 14]. In response we are updating the proposed list of reference sequences that appeared in Table 1 page III-30 of *Human Retroviruses and AIDS 1996*. We anticipate updating this table and the accompanying alignments on a regular basis in the future as a service to the scientific community, particularly since new submissions are likely to include a considerable number of complex mosaic genomes.

Table 1 lists reference sequences for 9 group M subtypes in major coding regions, i.e., *gag*, *pol*, *env* and *nef*. The criteria for inclusion of a reference sequence have changed since 1996, with a shift in emphasis to using full length HIV genomic sequences to serve as representatives for the subtypes, when such sequences are available. These are supplemented with sequences spanning intact coding regions, or else the longest gene fragments currently available. Table 1A summarizes a basic list of reference sequences for international variation and subtyping efforts; Table 1B provides additional information, including GenBank accession numbers, citations describing the respective sequence, sampling year, and the country where the virus was collected. Protein and nucleotide sequence alignments are available from the HIV database for each of the four coding regions on the web site at <http://hiv-web.lanl.gov>. These alignments also contain the HIV-1 O group sequences ANT70 and MVP5180 and chimpanzee viral sequences CPZ-GAB and CPZ-ANT. Full length nucleotide genome alignments are also available in Part I of this volume. The alignments were generated using first a hidden Markov model [4] at the nucleotide level, and then manually corrected at the amino acid level to keep open reading frames in frame, and finally back-translated into nucleotides. We include here maximum likelihood trees [5, 15] showing the phylogenetic relationships of the representative sequences selected for the *gag*, *pol*, *env* and *nef*. Trees for V3 and p17 sequences are also included, because: 1) these genomic regions are commonly sequenced; 2) sequences are available for all of the selected reference strains, including the shorter fragments; and 3) the organization of the V3 region tree is somewhat altered relative to the full length *env* tree.

Some of the HIV-1 subtypes are more clearly defined than others. Each subtype A, B, C, D, F, and H, contains at least one full length, apparently non-recombinant genome available as a reference sequence, as well as multiple additional full length *env* and *gag* sequences. Non-recombinant means there are no identified conflicting subtype associations in different regions of the sequence. Our current database and methods, however, have limitations, and additional sequences and analyses in the future may change our current understanding of subtype relationships. Short fragments of genes are unreliable for subtyping a sequence. For example, while the full-length representative subtype H sequence is essentially subtype H throughout the genome (see trees), phylogenetic analyses of V3 sequences occasionally produce trees that show a close association of subtype H and A sequences. Since this

discordance is not always observed but depends on the particular alignments and programs used, it is not sufficient evidence for recombination. Nevertheless, the fact that it can occur, underscores the possibility that ambiguities can arise with subtyping efforts, particularly those focused on short regions of the genome.

All full length representatives of subtypes E and G that have been sequenced to date represent mosaic genomes, with parts of the viral genome clustering with the A subtype in phylogenetic analysis, and parts of the genome forming the two clearly distinguishable clades designated either E or G. The E subtype is clearly "E" in envelope, appears to be an A/E mosaic in the regulatory regions, and is essentially A-associated in *gag* and *pol* [3, 8]. All of the longer subtype G sequences available to date have stretches of subtype A-associated sequence interspersed. In contrast to the A/E mosaics, however, these A/G mosaic sequences have many different patterns of A-like sequence, suggesting that they resulted from independent recombination events [6, 13, 14]. Subtype I sequences [10] have also been reanalyzed [7, 16] and there is now evidence from analysis of a complete genome that they are multiply mosaic and composed of 3 (or of possibly 4) subtypes [7]. Further analysis is necessary to determine the potential they have to represent a distinct subtype. Only fragments of subtype I *env* and *gag* genes are available at this time [11], so further work is needed in this case as well to fully characterize these strains. Nevertheless, these sequences are included as reference sequences to assist in identification of possible I subtype related strains in the future. In addition to the two subtype I sequences listed in the table, three other sequences (GM4, GM5, GM7) have been found to cluster close to this subtype over the *env* V3 region [1]. However, the sequence of GM4 is suggested to be a G/I/C recombinant [2], where the question mark represents a ~600 bp section that includes the I-like V3 region.

REFERENCES

- [1] Blouin, J. C., E. A. Guzman, and B. T. Foley. 1996. Global variation in the HIV-1 V3 region, p. III-77-III-201. In G. Myers, B. Korber, B. Foley, K.-T. Jeang, J. W. Mellors, and S. Wain-Hobson (ed.), *Human Retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences*. Los Alamos National Laboratory, Los Alamos, NM.
- [2] Bobkov, A., R. Cheingsong-Popov, M. Salminen, F. McCutchan, J. Louwagie, K. Ariyoshi, H. Whittle, and J. Weber. 1996. Complex mosaic structure of the partial envelope sequence from a Gambian HIV type 1 isolate. *AIDS Res. Hum. Retrovirus*. 12:169-171.
- [3] Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Artenstein, P. A. Hegerich, D. St. Louis, D. S. Burke, and F. E. McCutchan. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* 70:5935-5943.
- [4] Eddy, S. 1995. HMMER Hidden Markov Models of Protein and DNA Sequence, 1.8 ed. Washington University School of Medicine, St. Louis, MO.
- [5] Felsenstein, J. 1993. PHYLIP: Phylogeny Inference Package, 3.52c ed. University of Washington, Seattle, WA.
- [6] Gao, F., D. L. Robertson, C. D. Carruthers, S. G. Morrison, B. Jian, Y. Chen, F. Barre-Sinoussi, M. Girard, A. Srinivasan, A. G. Abimiku, G. M. Shaw, P. M. Sharp, and B. H. Hahn. submitted 1997. Non-recombinant reference clones and sequences for human immunodeficiency virus type 1 subtypes A, C, D, E, and H.
- [7] Gao, F., D. L. Robertson, and B. H. Hahn. 1997, unpublished data.
- [8] Gao, F., D. L. Robertson, S. G. Morrison, H. Hui, S. Craig, J. Decker, P. N. Fultz, M. Girard, G. M. Shaw, B. H. Hahn, and P. M. Sharp. 1996. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol.* 70:7013-7029.
- [9] Korber, B., I. Lousert-Ajakai, J. Blouin, and S. Saragosti. 1997. A comparison HIV-1 group M and group O functional and immunogenic domains in the *gag* p24 protein and the C2V3 region of the envelope protein, p. III-41-III-56. In G. Myers, B. Korber, B. Foley, K.-T. Jeang, J. W. Mellors, and S. Wain-Hobson (ed.), *Human Retroviruses and AIDS 1996: a compilation and analysis of nucleic acid and amino acid sequences*. Los Alamos National Laboratory, Los Alamos, NM.

- [10] Kostrikis, L. G., E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, and D. D. Ho. 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J. Virol.* 69:6122-6130.
- [11] Leitner, T., A. Alaeus, S. Marquina, E. Lilja, K. Lidman, and J. Albert. 1995. Yet another subtype of HIV type 1? *AIDS Res. Hum. Retrovirus.* 11:995-997
- [12] Louwagie, J., F. E. McCutchan, M. Peeters, T. P. Brennan, B. E. Sanders, G. A. Eddy, G. van der Groen, K. Fransen, G.-M. Gershy-Damet, R. Deleys, and D. S. Burke. 1993. Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS.* 7:769-780.
- [13] McCutchan, F. 1997, personal communication.
- [14] McCutchan, F. E., M. O. Salminen, J. K. Carr, and D. S. Burke. 1996. HIV- 1 genetic diversity. *AIDS.* 10 (suppl 3):S13-S20.
- [15] Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek. 1994. fastDNAm1: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.* 10:41-48.
- [16] Salminen, M. 1996, personal communication.

HIV-1 Genetic Subtypes

Table 1A Updated Proposal of Reference Sequences of HIV-1 Genetic Subtypes

Subtype	<i>gag</i>	<i>pol</i>	<i>env</i>	<i>nef</i>
A	U455 92UG037.1 K89 VI32	U455 92UG037.1	U455 92UG037.1 K89(KENYA) SF170	U455 92UG037.1
B	HXB2 JRFL OYI RF	HXB2 JRFL OYI RF	HXB2 JRFL OYI RF	HXB2 JRFL OYI RF
C	ETH2220 92BR025.8 UG268 ZAM18	ETH2220 92BR025.8	ETH2220 92BR025.8 UG268 ZAM18	ETH2220 92BR025.8
D	NDK Z2Z6 ELI 94UG114.1 ¹	NDK Z2Z6 ELI 94UG114.1 ¹	NDK Z2Z6 ELI 94UG114.1 ¹	NDK Z2Z6 ELI 94UG114.1 ¹
E ²			CM240 TN235 90CR402.1 93TH253.3	
F	93BR020.1 BZ162 VI69 VI174	93BR020.1	93BR020.1 BZ163 BZ126 RJ103	93BR020.1
G ³	92NG003.1 ⁴ 92NG083.2 SE6165 ⁵	92NG083.2 SE6165 ⁵	92NG003.1 ⁴ 92NG083.2 92UG975.10 92RU131.9	92NG003.1 ⁴ 92NG083.2
H	90CF056.1 VI557 ⁵	90CF056.1	90CF056.1 VI557 ⁵ CA13 ⁵	90CF056.1
J	SE7022 ⁵ SE7887 ⁵		SE7022 ⁵ SE7887 ⁵	

¹ The sequence 94UG114.1 is the most distant complete genome D subtype sequence (see trees), tending to branch off closest to the B/D root in most analyses. In some subgenomic regions, it may even move outside the B/D cluster.

² E in most of *env*, A in *gag* and *pol*, mixture of A & E in regulatory genes [3, 8, 14].

³ The G reference sequences may show resemblance to subtype A in regions of *pol* and *vif*, see also footnote 4.

⁴ 92NG003.1 is a full length sequence, but is not included in the *pol* alignment because of A-like regions in this gene [6].

⁵ Full length gene sequences of *gag*, *pol*, or *env* are not yet available, see Table 1B.

Table 1B Sequence descriptions

Subtype	Sequence	Acc. No.	Source	Region	Sampling year	Sampling country (origin)
HIV-1 M Group sequences in alignments						
A	U455	M62320	Oram, J.D. et al, <i>ARHR</i> 6:1073-1078 (1990)	complete	NA	Uganda
A	92UG037.1	U51190	Gao, F. et al, <i>J. Virol.</i> 70:7013-7029 (1996)	complete	1992	Uganda
A	K89	L11774	Louwagie, J. et al, <i>AIDS</i> 7:769-780 (1993)	<i>gag</i>	NA	Kenya
A	VI32	L11788	Louwagie, J. et al, <i>AIDS</i> 7:769-780 (1993)	<i>gag</i>	NA	Burundi
A	SF170	M66533	Evans, L. et al, <i>PNAS</i> 85:2815 (1988)	<i>env</i>	NA	Rwanda
B	HXB2	K03455, M38432	Wong-Staal, F. et al, <i>Nature</i> 313:277-284 (1985)	complete	NA	France
B	JRFL	U63632	O'Brien, W.A. et al, <i>Nature</i> 348:69 (1990)	complete	NA	US
B	OYI	M26727	Wain-Hobson, S. et al, <i>AIDS</i> 3:707 (1989)	complete	NA	Gabon
B	RF	M17451, M12508	Starck, B.R. et al, <i>Cell</i> 45:637-648 (1986)	complete	1983	US (Haiti)
C	ETH2220	U46016	Salminen, M.O. et al, <i>ARHR</i> 12:1329-1339 (1996)	complete	1986	Ethiopia
C	92BR025.8	U52953	Gao, F. et al, in preparation (1997)	complete	1992	Brazil
C	UG268	L11799	Louwagie, J. et al, <i>AIDS</i> 7:769-780 (1993)	<i>gag</i>	1993	Uganda
C	UG268	L22948	Louwagie, J. et al, <i>J. Virol.</i> 69:263-271 (1995)	<i>env-nef</i>	1993	Uganda
C	ZAM18	L03705	McCutchan, F. et al, <i>AIDS</i> 5:441-449 (1991)	<i>gag</i>	1989	Zambia
C	ZAM18	L22954	Louwagie, J. et al, <i>J. Virol.</i> 69:263-271 (1995)	<i>env</i>	1989	Zambia
D	NDK	M27323	Spire, B. et al, <i>Gene</i> 81:275-284 (1989)	complete	NA	Zaire
D	Z26	M22639	Theodore, T. et al, unpublished (1988)	complete	NA	Zaire
D	ELI	K03454, X04414	Alizon, M. et al, <i>Cell</i> 46:63-74 (1986)	complete	NA	Zaire
D	94UG114.1	U88824	Gao, F. et al, in preparation (1997)	complete	1994	Uganda
E	CM240	U54771	Carr, J.K. et al, <i>J. Virol.</i> 70:5935-5943 (1996)	complete	1990	Thailand
E	TN235	L03698	McCutchan, F.E. et al, <i>ARHR</i> 8:1887-1895 (1992)	<i>env</i>	NA	Thailand
E	90CR402.1	U51188	Gao, F. et al, <i>J. Virol.</i> 70: 7013-7029 (1996)	complete	1990	Central African Republic
E	93TH253.3	U51189	Gao, F. et al, <i>J. Virol.</i> 70: 7013-7029 (1996)	complete	1993	Thailand
F	BZ162	L11751	Louwagie, J. et al, <i>AIDS</i> 7:769-780 (1993)	<i>gag</i>	NA	Brazil
F	VI69	L11796	Louwagie, J. et al, <i>AIDS</i> 7:769-780 (1993)	<i>gag</i>	NA	Belgium (Rwanda)
F	VI174	L11782	Louwagie, J. et al, <i>AIDS</i> 7:769-780 (1993)	<i>gag</i>	NA	Zaire
F	BZ163	L22085	Louwagie, J. et al, <i>ARHR</i> 10:561-567 (1994)	<i>env-nef</i>	NA	Brazil
F	BZ126	L22082	Louwagie, J. et al, <i>ARHR</i> 10:561-567 (1994)	<i>env-nef</i>	NA	Brazil
F	93BR020.1	AF005494	Gao, F. et al, in preparation (1997)	complete	1993	Brazil
F	RJ103	U08974	Sabino, E.C. et al, <i>J. Virol.</i> 68:6340-6346 (1994)	partial <i>env</i>	NA	Brazil
G	92NG003.1	U88825	Gao, F. et al, in preparation (1997)	complete	1992	Nigeria
G	SE6165	L40752, L40761	Leitner, T. et al, <i>Virology</i> 209:136-146 (1995)	p17, RT	1993	Sweden (Central Africa)
G	92NG083.2	U88826	Gao, F. et al, in preparation (1997)	complete	1992	Nigeria
G	92UG975.10	U27426	Gao, F. et al, <i>J. Virol.</i> 70:1651-1657 (1996)	<i>env</i>	1992	Uganda
G	93RU131.9	U30312	Gao, F. et al, <i>ARHR</i> 10:1651-1657 (1996)	<i>env-nef</i>	1992	Russia
H	90CF056.1	AF005496	Gao, F. et al, in preparation (1997)	complete	1990	Central African Republic
H	VI557	U09666	Janssens, W. et al, <i>ARHR</i> 10:877-879 (1994)	V3-V5	NA	Zaire
H	VI557	L11793	Louwagie, J. et al, <i>AIDS</i> 7:769-780 (1993)	<i>gag</i>	NA	Zaire
H	CA13	U09667	Janssens, W. et al, <i>ARHR</i> 10:877-879 (1994)	V3-V5	NA	Cameroon
J	SE7022	L41177, L41179	Leitner, T. et al, <i>ARHR</i> 11:995-997 (1995)	V3, p17	1993	Sweden (Zaire)
J	SE7887	L41176, L41178	Leitner, T. et al, <i>ARHR</i> 11:995-997 (1995)	V3, p17	1994	Sweden
Additional sequences available in alignments						
O Group	ANT70	L20587	Vanden Haesevelde, M. et al, <i>J. Virol.</i> 68:1586-1596 (1994)	complete	NA	Cameroon
O Group	MVP5180	L20571	Gurtler, L. et al, <i>J. Virol.</i> 68:1581-1585 (1994)	complete	1991	Cameroon
Chimpanzee	SIV-CPZANT	U42720	Vanden Haesevelde, M. et al, <i>Virology</i> 221:346-350 (1996)	complete	1986	Zaire
Chimpanzee	SIV-CPZGAB	X52154	Huet, T. et al, <i>Nature</i> 345:356-359 (1990)	complete	NA	Gabon

HIV-1 Genetic Subtypes

Unrooted phylogenetic trees calculated with maximum likelihood methods [5, 15]. The alignments for the various subgenomic regions are derived from the complete genome alignment, as presented in the HIV data base 1997. Each of the alignments are available from the HIV database at <http://hiv-web.lanl.gov>. All alignments were globally gap/aligned for the generation of the trees to 1398 sites for the gag gene; 2894 sites for the pol gene; 2250 sites for the env gene; 282 sites for the nef gene; 426 sites for the p17/gag fragment; and 213 sites for the env V3 fragment. All trees, except the nef tree, were constructed using the program DNAML under the F84 substitution model [5] where nucleotide frequencies were derived from the datasets and the transition/transversion parameter was set to 3.0 for the gag gene, 2.0 for the pol gene, 1.5 for the env gene, 3.0 for the p17 fragment, and 1.42 for the V3 fragment. The nef gene tree was calculated with the programs fastDNAML and DNAsites [15], to allow for different substitution rates across sites. This proved to be important for the topology of this gene tree in resolving subtypes B and D. Although G subtype sequences in the trees shown cluster separate from subtype A, they cluster in subtype A in some subgenomic regions [6]. Subtypes B and D are generally close to each other in all analyses, but with the nef gene it may be difficult to completely resolve them from each other. All trees are drawn to the same scale, thereby indicating the relative information density in the different regions.

Intersubtype Recombinant HIV-1 Sequences

David L. Robertson,¹ Feng Gao,² Beatrice H. Hahn,² Paul M. Sharp³

¹ *Laboratory of Structural and Genetic Information, CNRS-EP 91, 31 Chemin Joseph Aiguier, 13402 Marseilles Cedex 20, France;*

² *Departments of Medicine and Microbiology, University of Alabama at Birmingham, 702 S. 19th Street, LHRB 613, Birmingham, AL 35294;*

³ *Division of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, U.K.*

Retroviruses are highly recombinogenic, but recombination can only occur between genomes packaged within the same virion. Until about 1994, it was generally thought that individuals do not become infected with multiple distinct HIV-1 strains, and so the possibility that recombination between divergent viruses could contribute to the evolution of HIV-1 was not widely considered. However, not long after the recognition that the considerable global diversity of the major (M) group of HIV-1 could be described in terms of discrete sequence subtypes, it was also realized that a significant fraction of HIV-1 strains are in fact intersubtype recombinants [1,2]. Analyses of viruses characterized over the last few years have confirmed that substantial numbers of HIV-1 isolates have mosaic genomes [3-16]. The finding of such hybrid viruses has a number of implications both for attempts at viral characterization (phylogenetic analyses, and tracking of the epidemic), and for concerns about the future genetic diversity of HIV-1 and its impact on vaccine development.

Hybrid viruses can be detected because their phylogenetic affinities vary depending on the region of genome analysed. A useful initial approach is to examine the extent of sequence divergence/similarity between a new sequence and a bank of reference sequences of different subtypes, for example as a diversity plot [9], or using the RIP program [17]; if the extent of relative similarity to different subtypes varies along the sequence, this may indicate that the sequence is a recombinant. However, fuller investigation must involve a phylogenetic approach, comparing trees derived by analyses of different regions of the genome, and assessing the confidence of phylogenetic clustering by a statistical approach such as the bootstrap. A thorough analysis would involve taking a window of sequence of a certain size, and moving this window along the genome in steps of a defined size, generating perhaps hundreds of trees for visual examination in the process! There are at least two short cuts. One is to analyse only a few windows, defining selected regions according to the output of the diversity analysis. Another is to not examine the entire phylogenetic tree of all subtypes, but to focus on one particular phylogenetic question. Thus, if the initial analyses suggest that a sequence may be a recombinant between two particular subtypes, it is possible to ask simply what is the bootstrap value for the clustering of the new sequence with one or another particular subtype, and plot these values as a function of position along the genome; this is the basis of the "bootscanning" approach [18]. Once the subtypes putatively involved in the recombination event have been identified, and the crossover points have been approximately localized, more precisely defined breakpoints can be determined, and their statistical significance assessed, using informative site analysis [2,6,19]. (For a fuller discussion of recombination analysis methods, see Ref.16.)

Among the full-length HIV-1 sequences available currently (i.e., in December 1997), there are 12 that appear to be intersubtype recombinants (Table 1, Figure 1). Clearly, the delineation of recombinant genomes, and localization of the crossover points, relies on the availability of nonrecombinant reference sequences. Since such sequences are, as yet, available only for subtypes A, B, C, D, E, and H (see Leitner et al., pp. III.19-III.24, this volume), the breakpoints between regions of sequence of different subtype origin can be mapped only tentatively in those viruses derived from recombination events involving subtypes E, G, I or J; this includes the majority of the full length genome hybrids—in fact, all but the first three in the accompanying Figure and Table. In several viruses there are regions (shown in white) which have not yet been assigned to any subtype; the origin of these segments may be clarified as more reference sequences become available, but it is also possible that some of these regions come from subtypes that are no longer extant.

Table 1 Full-length Intersubtype Recombinant HIV-1 Sequences

Sequence	Acc. No.	Subtypes	Origin	Year	References
ZAM184	U86780	A,C	Zambia	1990	[12]
92RW009.6	U88823	A,C	Rwanda	1992	[16]
93BR029.4	AF005495	B,F	Brazil	1993	[16]
90CF402.1	U51188	A,E	C.A.R.	1990	[9]
93TH253.3	U51189	A,E	Thailand	1993	[9]
CM240	U54771	A,E	Thailand	1990	[8]
92NG083.2	U88826	A,G	Nigeria	1992	[16]
92NG003.2	U88825	A,G	Nigeria	1992	[16]
IbNg	L39106	A,G	Nigeria	1991	[33]
MAL	K03456	A,D,I,?	Zaire	1985	[20]
Z321B	U76035	A,G,I,?	Zaire	1976	[14]
94CY032.3	AF049337	A,G,I,?	Cyprus	1994	[25]

The references indicate the sequence publication. IbNg [9] and MAL [19,22] were only subsequently recognized as recombinants. Utilizing the latest set of reference sequences for the various subtypes (see Leitner et al., pp. III.19-III.24), we have subjected all of these sequences to further analysis. The locations of the segments of different subtype origins, revised according to these latest analyses (Robertson, Gao, Sharp and Hahn; in preparation), are shown in Figure 1.

Most of the recombinant viruses originate from Africa (Table 1), and a substantial majority (11 of the 12 full length hybrids) include segments of subtype A (Figure 1). This is not surprising, since it appears that the various subtypes of HIV-1 first arose in central Africa and have co-circulated there for many years. Furthermore, in that region, subtype A is the most common. Among the hybrids of African origin, MAL is interesting because it was one of the first African (and, in retrospect, nonsubtype B) viruses to be characterized [20]. Li et al. first suggested that MAL may be a recombinant virus in 1987 [21,22], but the realization of the widespread occurrence of recombination among HIV-1 subtypes did not occur for about another 7 years. While MAL includes regions that are clearly of subtype A and D origin, there is a long region encompassing the 5' half of the *pol* gene that has defied characterization [19]. Now it appears that part of that region may be subtype I (see below), but the origin of the remainder has still not been determined (Figure 1). Thus, the ancestry of MAL may have involved successive recombination events of sequences derived from three or four different subtypes. Z321B, which also appears to include sequence blocks from three (or four?) different subtypes, is the oldest of the known hybrid viruses, having been isolated in 1976. Clearly, by that date, the group M subtypes had diverged sufficiently that the various segments recombined within Z321B are detectable as having distinct subtype origins. This implies that the subtypes must have begun to diverge substantially earlier than the 1970s, and is consistent with the recent finding that a virus from 1959 appears, from phylogenetic analyses, to have postdated the group M radiation [23].

Among the hybrid viruses isolated outside of Africa are two subtype "E" viruses (CM240 and 93TH253.3) from Thailand. As shown in Figure 1, a large fraction of the genome of these subtype "E" viruses, including the *gag* and *pol* genes, appears to be of subtype A origin [8,9]. Since all Thai "E" viruses appear to be closely related, and since the mosaic structure of their genomes appears to be identical to that of the other full length subtype "E" virus (90CF402.1) from the Central African Republic, it is most likely that a single recombination event occurred in Africa, and that the subtype "E" viruses in Thailand are due to the introduction of a descendant of that recombination event [9]. While the subtype "E" viruses clearly indicate that a subtype E arose in the past, it is possible that full length subtype E viruses no longer exist. In contrast, the recombination event between subtypes B and F to generate 93BR029.4 most likely occurred in Brazil, where both of these subtypes are common.

Figure 1. Full-length intersubtype recombinant HIV-1 sequences. The color code indicates the subtype origin of different segments of the genomes; regions in white are sequences whose origin has not been determined. The LTRs are hatched because they have not been analysed in detail. See Table 1 for more details.

HIV-1 Recombinants

At least three other viruses from Brazil, for each of which only partial sequences are available, appear to represent independently generated B/F hybrids [3,6,18]. The other non-African example, 94CY032.3, was isolated in Cyprus. Initially, partial sequences from this virus formed the basis of the proposition of subtype I [24]. Analysis of a recently determined full length sequence [25] suggests that it is a recombinant of at least three subtypes, A, G and I. Since 94CY032.3 is the prototypic subtype I virus, it is most parsimonious to assume that all segments which cannot be assigned to any other subtype are of a common (i.e., subtype I) origin. Two of the recombinant viruses from Africa, MAL and Z321B, contain short segments of sequence closely related to 94CY032.3, but not similar to other subtypes; thus, these regions have also been designated as being of subtype I origin (Figure 1). It is not clear whether the multiple recombination events in the ancestry of 94CY032.3 occurred in Africa, or in Cyprus, where a number of different subtypes are also found [24].

It is notable that most of the mosaic genomes have a "patchwork" appearance due to multiple cross-over events (Figure 1). This is as expected given the known properties of reverse transcriptase. As more recombinant viruses are characterized, it may be possible to discern whether there are some consistent features to the patterns of genetic exchange. For example, certain sites may be hotspots for recombination; a hotspot may exist because the genome adopts a secondary structure that induces stalling of the reverse transcriptase, or because a region is conserved and so has a higher level of sequence identity. Alternatively, some exchanges may be "forbidden;" numerous components of the HIV-1 genome interact, either at the level of their RNA or the encoded translation product, and certain combinations of divergent interacting components may be inviable.

From the current data it is not possible to estimate the frequency of multiple infection or of recombination events. The 12 intersubtype recombinants come from a total of 53 available full length sequences. Of the 53 full length sequences, 30 come from subtype B, which is the focus of attention in the Western world but constitutes only about 16% of the global pandemic. Among the 12 hybrids, three are similar subtype "E" viruses, selected for investigation because of their impact on the epidemic in Thailand, but which likely all arose from a single ancestral recombination event (see above). Thus, counting only one of the three subtype "E" genomes, 10/21 full length non-subtype B sequences are independently generated recombinants of different subtypes. Among the much larger numbers of viruses that have been partially characterized, there are many additional examples of intersubtype recombinants [2-6,10,11,13,15]. We have previously estimated that roughly 10% of viruses for which at least a near full-length *gag* or *env* gene sequence is available are recombinants [2]. This figure is possibly an underestimate of the frequency of recombinant viruses, both because of the excess representation of subtype B, and because a virus for which only one gene has been sequenced may be a recombinant but not have a crossover within that gene; for example, if judged only on the basis of their *gag* or *env* gene sequences (see Figure 1), 92NG003.1 and 92NG083.2 appear to represent nonrecombinant subtype G viruses, while 93BR029 was originally classified as subtype F on the basis of an *env* gene phylogeny [6]. Finally, the frequency of new recombinants in any given geographic area will obviously depend on the number and prevalence of co-circulating subtypes: multiple subtypes well represented in the same population can be expected to yield a higher number of new hybrid viruses.

Recombinant viruses have also been detected among HIV-2 strains [19,26] and among SIVs [27,28]. As yet, no viruses have been described that are recombinants of the two highly divergent major groups of HIV-1, M and O, nor have recombinants of HIV-1 and HIV-2 been found. Individuals multiply infected with HIV-1 groups M and O [29], and with both HIV-1 and HIV-2 [30], have been reported. Recombination between these more divergent viruses may not occur because of the extent of sequence dissimilarity or may yield replication attenuated or defective viruses. Rather few HIV-1 intrasubtype recombinants have been described [7,31,32], but in this case this is probably merely a problem of detection.

References

- [1] Hahn, B.H., Robertson, D.L., McCutchan, F.E. and Sharp, P.M. Recombination and diversity of HIV: implications for vaccine development. Pp. 87-94 in: *Retroviruses of Human AIDS and Related Diseases*. Eds. Girard, M. and Dodet, B. Neuvieme Colloque des Cent Gardes, 1994.

- [2] Robertson, D.L., Sharp, P.M., McCutchan, F.E. and Hahn, B.H. Recombination in HIV-1. *Nature* 374:124-126, 1995.
- [3] Sabino, E.C., Shpaer, E.G., Morgado, M.G., Korber, B.T., Diaz, R.S., Bongertz, V., Cavalcante, S., Galvao-Castro, B., Mullins, J.I. and Mayer, A. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J. Virol.* 68:6340-6346, 1994.
- [4] Hahn, B.H., Robertson, D.L. and Sharp, P.M. Intersubtype recombination in HIV-1 and HIV-2. Pp. III.22-III.29 in: *Human Retroviruses and AIDS: a Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Eds. Myers, G., Korber, B., Hahn, B.H., Jeang, K.-T., Mellors, J.W., McCutchan, F.E., Henderson, L.E. and Pavlakis, G.N. Los Alamos National Laboratory, Los Alamos, New Mexico, 1995.
- [5] Leitner, T., Escanilla, D., Marquina, S., Wahlberg, J., Brostrom, C., Hansson, H.B., Uhlen, M. and Albert, J. Biological and molecular characterization of subtype D, subtype G, and subtype A/D recombinant HIV-1 transmissions in Sweden. *Virology* 209:136-146, 1995.
- [6] Gao, F., Morrison, S.G., Robertson, D.L., Thornton, C.L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B., von Briesen, H., Beddows, S., Weber, J., Sharp, P.M., Shaw, G.M., Hahn, B.H., and the WHO and NIAID Networks for HIV Isolation and Characterization. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. *J. Virol.* 70:1651-1667, 1996.
- [7] Diaz, R.S., Sabino, E.C., Mayer, A., Mosley, J.W., Busch, M.P. and The Transfusion Study Group. Dual human immunodeficiency virus type-1 infection and recombination in a dually exposed transfusion recipient. *J. Virol.* 69: 3273-3281, 1995.
- [8] Carr, J.K., Salminen, M.O., Koch, C., Gotte, D., Arntstein, A.W., Hegerich, P.A., St. Louis, D., Burke, D.S. and McCutchan, F.E. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* 70:5935-5943, 1996.
- [9] Gao, F., Robertson, D.L., Morrison, S.G., Hui, H., Craig, S., Decker, J., Fultz, P.N., Girard, M., Shaw, G.M., Hahn, B.H. and Sharp, P.M. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol.* 70: 7013-7029, 1996.
- [10] Cornelissen, M., Kampinga, G., Zorgdrager, F., Goudsmit, J. and UNAIDS Network for HIV Isolation and Characterization. Human immunodeficiency virus type 1 subtypes defined by *env* show high frequency of recombinant *gag* genes. *J. Virol.* 70:8209-8212, 1996.
- [11] Bobkov, A., Cheingsong-Popov, R., Salminen, M., McCutchan, F., Louwagie, J., Ariyoshi, K., Whittle, H. and Weber, J. Complex mosaic structure of the partial envelope sequence from a Gambian HIV Type 1 isolate. *AIDS Res. Hum. Retrovirus.* 12:169-171, 1996.
- [12] Salminen, M.O., Carr, J.K., Robertson, D.L., Hegerich, P., Gotte, D., Koch, C., Sanders-Buell, E., Gao, F., Sharp, P.M., Hahn, B.H., Burke, D.S. and McCutchan, F.E. Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. *J. Virol.* 71: 2647-2655, 1997.
- [13] Wieland, U., Seelhoff, A., Hofmann, A., Kuhn, J.E., Eggers, H.J., Mugenyi, P. and Schwander, S. Diversity of the *vif* gene of human immunodeficiency virus type 1 in Uganda. *J. Gen. Virol.* 78:393-400, 1997.
- [14] Choi, D.J., Dube, S., Spicer, T.P., Slade, H.B., Jensen, F.C. and Poiesz, B.J. HIV type 1 isolate Z321, the strain used to make a therapeutic HIV type 1 immunogen, is intersubtype recombinant. *AIDS Res. Hum. Retrovirus.* 13: 357-361, 1997.
- [15] Takehisa, J., Oseikwasi, M., Ayisi, N.K., Hishida, O., Miura, T., Igarashi, T., Brandful, J., Ampofo, W., Netty, V.B.A., Mensah, M., Yamashita, M., Ido, E. and Hayami, M. Phylogenetic analysis of human immunodeficiency virus 1 in Ghana. *Acta. Virology* 41:51-54, 1997.
- [16] Gao, F., Robertson, D.L., Carruthers, C.D., Morrison, S.G., Jian, B., Chen, Y., Barre-Sinoussi, F., Girard, M., Srinivasan, A., Abimiku, A.G., Shaw, G.M., Sharp, P.M., and Hahn, B.H. A comprehensive panel of near full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol.* (Submitted).

- [17] Siepel, A.C., Halpern, A.L., Macken, C. and Korber, B.T.M. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res. Hum. Retrovirus*. 11:1413-1416, 1995.
- [18] Salminen, M., Carr, J.K., Burke, D.S. and McCutchan, F.E. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retrovirus*. 11:1423-1425, 1995.
- [19] Robertson, D.L., Hahn, B.H. and Sharp, P.M. Recombination in AIDS viruses. *J. Mol. Evol.* 40:249-259, 1995.
- [20] Alizon, M., Wain-Hobson, S., Gluckman, J.-C. and Sonigo, P. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* 46:63-74, 1986.
- [21] Li, W.-H., Tanimura, M. and Sharp, P.M. Rates and dates of divergence between AIDS virus nucleotide sequences. *Genetics* 116:s44, 1987.
- [22] Li, W.-H., Tanimura, M. and Sharp, P.M. Rates and dates of divergence between AIDS virus nucleotide sequences. *Molec. Biol. Evol.* 5:313-330, 1988.
- [23] Zhu, T., Korber, B.T., Nahmias, A.J., Hooper, E., Sharp, P.M. and Ho, D.D. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 391:594-597, 1998.
- [24] Kostrikis, L.G., Bagdades, E., Cao, Y., Zhang, L., Dimitriou, D. and Ho, D.D. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J. Virol.* 69:6122-6130, 1995.
- [25] Gao, F., Robertson, D.L., Carruthers, C.D., Salminen, M., Kostrikis, L., Ho, D.D., Sharp P.M., and Hahn B.H. Unpublished.
- [26] Gao, F., Yue, L., Robertson, D.L., Hill, S.C., Hui, H., Biggar, R.J., Neequaye, A.E., Whelan, T.M., Ho, D.D., Shaw, G.M., Sharp, P.M. and Hahn, B.H. Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J. Virol.* 68: 7433-7447, 1994.
- [27] Jin, M.J., Hui, H., Robertson, D.L., Muller, M.C., Barre-Sinoussi, F., Hirsch, V.M., Allan, J.S., Shaw, G.M., Sharp, P.M. and Hahn, B.H. Mosaic genome structure of simian immunodeficiency virus from West African green monkeys. *EMBO J.* 13:2935-2947, 1994.
- [28] Chen, Z., Telfer, P., Gettie, A., Reed, P., Zheng, L., Ho, D.D. and Marz, P.A. Genetic characterization of new West African simian immunodeficiency virus STVsm: geographic clustering of household-derived STV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J. Virol.* 70:3617-3627, 1996.
- [29] Takehisa, J., Zekeng, L., Miura, T., Ido, E., Yamashita, M., Mboudjeka, I., Gurtler, L.G., Hayami, M. and Kaptue, L. Triple HIV-1 infection with group O and Group M of different clades in a single Cameroonian AIDS patient. *J. AIDS Human Retrovirol.* 14:81-2, 1997.
- [30] Grez, M., Dietrich, U., Balfe, P., von Briesen, H., Maniar, J.K., Mahambre, G., Delwart, E.L., Mullins, J.I., Rubsamen-Waigmann, H. Genetic analysis of Human Immunodeficiency Virus Type 1 and 2 (HIV-1 and HIV-2) mixed infections in India reveals a recent spread of HIV-1 and HIV-2 from a single ancestor for each of these viruses. *J. Virol.* 68:2161-2168, 1994.
- [31] Liu, S.-L., Mulvania, T., He, X., Kosloff, B., Rodrigo, A., Corey, L. and Mullins, J.I. Human immunodeficiency virus type 1 recombination in vivo associated with selection advantage. (In preparation.)
- [32] Zhu, T.F., Wang, N., Carr, A., Wolinsky, S., and Ho, D.D. Evidence for co-infection by multiple strains of Human Immunodeficiency Virus Type 1 subtype B in an acute seroconverter. *J. Virol.* 69:1324-1327, 1995
- [33] Howard, T.M. and Rasheed, S. Genomic structure and nucleotide sequence analysis of a new HIV type 1 subtype A strain from Nigeria. *AIDS Res. Hum. Retrovirus*. 12:1413-1425, 1996.

Fine Structure of HIV and SIV

Hans R. Gelderblom

Robert Koch-Institut, Nordufer 20, D-13353 Berlin, Germany

Aim and Contents of this Note

This note, which is especially directed to scientists working in the fields of molecular genetics and immunology, summarizes essential aspects of morphogenesis and fine structure of HIV and SIV. All retroviruses, including HIV and SIV, show three functional states during their extracellular life cycle: assembly, maturation and aging. These states have been elucidated by transmission electron microscopy (TEM) and closely correlated to viral functions using wild-type virus as well as a variety of functional mutants (for review see Gelderblom, 1991; Hunter, 1994; Nermut and Hockley, 1996). In addition, two early steps in virus-cell interaction, virus adsorption and entry, have been studied by EM.

The possibility of clearly distinguishing between three morphologically and functionally distinct states is unique for retroviruses. Because of the close relationship between viral structure and function, EM studies have contributed key information to the understanding of HIV and, ultimately, to the development of antiretroviral strategies.

This note covers three topics. In Section I, techniques applied in the EM analysis of retroviruses are presented together with an overview of comparative retrovirus classification using both morphological and genetic information. In Section II morphogenesis and maturation of HIV/SIV are discussed in detail. Section III outlines the structure and composition of both mature and aging virus particles and thereby also covers the presence and possible role of host cell components with the virion.

I. Electron microscopy and morpho-classification of retroviruses

Electron microscopical techniques The majority of structural findings on HIV and SIV have been collected by TEM analysis of positive-contrasted thin sections of specimens, usually virus-producing cells (see Figs. 4–6). The specimens are fixed using aldehyde and osmium-tetroxide, dehydrated, and embedded in resins. Sections of 40 to 60 nm in thickness are post-stained using heavy metal salt solutions (for details see *Electron Microscopy* by Bozzola and Russel, 1991). The “positive” contrast observed in thin section EM is mainly due to the heavy metals bound to the specimen and their scattering power, i.e., the distribution of heavy elements is shown rather than the details of the object itself (Figs. 1, 5).

Fig. 1. Demonstration of positive and negative contrast phenomena in electron microscopy of biological specimens. Positive contrast is achieved by incubation with electron-dense heavy metal ions and covalent binding of this stain. In negative contrast the same or similar stains are used to wrap the biological structures without interacting chemically with the object.

Thin section EM involves chemical fixation and complete dehydration of the specimen which definitely abolish any "natural" state of biological structures. Nevertheless, thin section EM has been very rewarding in the structural analysis of retroviruses, due to the reproducible differential positive contrast and the appreciable fine structure attained in spite of all the artifacts.

Cryo-EM, in contrast, developed during the last two decades, has already given "molecular" information in many instances (for review see Dubochet et al., 1988; Müller, 1992). Cryo-EM involves physical fixation, i.e. rapid freezing instead of chemical fixation, and the analysis of frozen-hydrated specimens at temperatures below -140°C . The image obtained in cryo-EM depends on the scattering of the biological matter itself, as there is no heavy metal staining involved, and the first results obtained have been rewarding (Nermut et al., 1993). Specimen preparation, operation of the cryo-EM and image processing are, however, technically demanding. The biohazard of working with cryo-fixed, i.e. "living" virus poses additional obstacles. Therefore, in a recent cryo-EM study on the organization of HIV, non-infectious, immature particles generated by recombinant techniques and by "therapeutic" antiviral treatment (Fuller et al., 1997) were investigated.

In negative staining, biological structures usually remain unstained but are surrounded by an electron-dense stain and therefore appear lighter, i.e., in negative contrast, compared to the dark background. As with positive staining, salts of heavy metals—tungsten, uranium, and molybdenum—are used to create contrast. The stain, applied as a one to two percent solution, surrounds the virus particle pre-adsorbed to the grid and dries out (Figs. 1, 5). This rapid technique can result in high resolution information and is used to study selected aspects of retroviruses, (e.g. the presence and shape of SIV glycoprotein knobs) and also in conjunction with immuno-EM (IEM) (Gelderblom et al., 1988; Grief et al., 1989; Hockley et al., 1988; Chrystie and Almeida, 1988, 1989). But, because of osmotic damage and other artifacts, negative staining is not recommended as the preferred EM procedure for the elucidation of retrovirus fine structure.

Finally, scanning EM (SEM) has been used to study retroviruses and their interaction with the host cell. Conventional SEM is well suited to the study of virus-infected cell cultures (Gelderblom et al., 1988; Hockley et al., 1988; Nermut et al., 1993). State-of-the-art high resolution SEM can resolve the shape of virus envelope knobs (Renz et al., 1995), however, TEM is mostly preferred in virus fine structural studies, because SEM information is limited to the viral surface.

Table 1 Morphological criteria for the differentiation of retroviruses

Criteria	Mammalian type B	Mammalian type C	Avian type C	Type D group	HTLV-BLV group	Lentivirus group	Spumavirus group
Morphogenesis							
Cores preformed in the cytoplasm	+	-	-	+	-	-	+
Core assembly concomitant with budding	-	+	+	-	+	+	-
Distance of core shell to prospective envelope	wide	wide	wide	wide	narrow	narrow	wide
Morphology of the mature virion							
Shape of the core capsid	isometric	isometric	isometric	cone	isometric	cone-to-tube	isometric
Site of core	eccentric	concentric	concentric	diagonal	concentric	diagonal	concentric
Length of the envelope projections	8-9 nm	5 nm	5 nm	5 nm	5 nm	9-10 nm	12 nm
Anchorage stability of envelope knobs	high	relatively high	relatively high	relatively high	low	low	low

Fig. 2. Using the morphological criteria listed in Table 1, retroviruses form seven clearly distinguishable groups of agents. The members of each particular group share specific biological properties (for detail see Gelderblom et al., 1985a; Gelderblom, 1991).

Fig. 3. Phylogenetic tree of 19 "prototypic" members of the retrovirus family (modified from Griffith et al., 1997). The genetic classification fits remarkably well the morphological grouping shown in Fig. 2.

Criteria used for the morphological differentiation of retroviruses Starting with the pioneering work of Wilhelm Bernhard in the late 1950s, TEM has been used to differentiate and classify newly detected retroviruses (Bernhard, 1960). Comparative studies on members of the different retrovirus groups have revealed well-defined structural differences that can be assessed either during budding and in the immature state or with the mature and aging virion. The distinctive criteria listed in Table I have often been used to assign "new" retrovirus isolates according to morphology into the proper genus (Fig. 2; for review see Gelderblom, 1991).

II. Morphogenesis and maturation of HIV and SIV

The extracellular "life" of a retrovirus consists of several states that can be studied directly by EM (for review see Gelderblom, 1991; Nakai and Goto, 1996; Nermut and Hockley, 1996).

1. The virus is formed at the plasma membrane by an ordered self-assembly of virus structural Gag and Gag-Pol precursor polyproteins (Pr55 gag, Pr155 Gag-Pol) together with genomic RNA.
2. After closure of the spherical ribonucleoprotein shell (RNP) the immature virion is released from the cell.
3. This particle undergoes a functional and morphological maturation as a consequence of proteolytic cleavage of the precursor proteins by the viral protease (PR). The cleavage occurs during assembly or after release of the immature particle.
4. The mature virion ages, rapidly losing infectivity.

Retrovirus-coded proteins are named according to function and localization within the virion by 2-letter acronyms (Leis et al., 1988). This nomenclature is widely applicable in comparative studies on enveloped viruses and is also used in this article (for details see Fig. 7). During maturation, Pr55

Gag is cleaved by the viral protease (PR) into the four major structural proteins of the mature virion (Henderson et al., 1988, 1992): membrane associated or matrix protein (p17 MA), capsid, (p24 CA), nucleocapsid (p10 NC), and the C-terminal core-envelope link (CEL, p6 LD) (Höglund et al., 1992).

The budding process HIV and SIV, like most retroviruses, self-assemble at plasma membranes (Figs. 4, 5, 6). As a first indication, an electron-dense patch appears directly underneath the lipid bilayer, while on the bud, on the outside of the bilayer, individual virus envelope knobs can be seen. The dense patch grows continuously by side-to-side assembly of the Gag- and Gag-Pol-polyproteins ultimately forming a spherical, electron-dense RNP shell. This immature core is surrounded by a tightly fitting lipid bilayer. High resolution EM of immature virions clearly shows lateral interaction of Gag-rods in the RNP (Hockley et al., 1988; Nermut et al., 1994; Fuller et al., 1997). The immature virion is densely studded with glycoprotein knobs. Only after complete sealing of the RNP core shell, does the concentrically organized immature virion separate from the plasma membrane and pinch off into the extracellular space (Gelderblom, 1991; Göttinger et al., 1991). The immature virion is slightly bigger than the mature particle (120–140 nm versus 110–130 nm; Nermut et al., 1993; Gentile et al., 1994) and appears to be more rounded than the latter which often shows an angular outline. The angular shapes have led to the conclusion that SIV and HIV represent isometric, probably icosahedral, structures (Marx et al., 1988; Özel et al., 1988; Goto et al., 1990; Nermut et al., 1993, 1994; Nermut and Hockley, 1996).

Virus formation and release depend on a number of structural requirements. Very early, intact Pr55 gag was shown using recombinant techniques to contain all the information necessary for viral assembly (Gheysen et al., 1989; Wills and Craven, 1991). The authentic viral RNP core shell of HIV is formed from two proteins, Pr55 Gag and Pr155 Gag-Pol, co-assembled at a ratio of approximately 20 to 1. Virus formation also involves oligomerization of the polyprotein and essentially represents a very constrained process (Jowett et al., 1992; Nermut et al., 1994; Zhang et al., 1996). A few factors shall be mentioned here.

Requirements for a functional MA Two factors, co-translational myristoylation and a highly basic region of the N-terminal MA domain, which binds to phospholipids, are essential for the stable association of MA with the plasma membrane and the formation of infectious HIV (Göttinger et al., 1989; Bryant and Ratner, 1990; Zhou et al., 1994). Other regions of MA determine the targeting of Pr55 Gag to the plasma membrane, the regular site of viral assembly (Fäcke et al., 1993; Yuan et al., 1993). MA functions are required for both the self-assembly of the Gag- and the Gag-Pol-polyproteins and the insertion of the viral Env protein into the prospective envelope (Yu et al., 1992; Dorfman et al., 1994 b). Significantly, MA is part of the viral pre-integration complex and acts in targeting this complex to the nucleus in quiescent cells, a process that permits the infection of non-dividing cells (Bukrinsky et al., 1993).

Properties of a functional CA The side-to-side assembly of the polyprotein rods depends on multiple protein-protein interactions. This process can be inhibited by the intercalation of short peptides derived from the respective domains (Niedrig et al., 1993). The core capsid is made up of p24 CA (Gelderblom et al., 1987a; Chrystie and Almeida, 1989). Domains mainly at the C-terminus of p24 CA are essential for proper assembly of the RNP shell (Zhang et al., 1996), while the N-terminal region is required for the formation of the mature, cone-shaped core. The middle region of CA containing the "major homology motif" is well conserved among different retroviruses and essential for the formation of infectious virions (Manumano et al., 1994). Finally, Pr55 Gag functions are required for the co-incorporation of Pr155 Gag-Pol into the growing RNP shell (Smith et al., 1993).

The role of p2 Between CA and NC a spacer of 14 amino acids (p2) is situated; another spacer, p1 is located between NC and p6 CEL (Henderson et al., 1992). P2 is readily detectable in lysates of infected cells, but virtually absent from cell-released virions. The C-terminal portion of CA including part of p2 is crucial for HIV morphogenesis. The spacer is required for the ordered assembly of the spherical RNP shell. The removal of p2 from the C-terminus of p25 CA occurs late during the controlled Pr55 and Pr155 protein cleavage kinetics and is essential for the transition of the spherical RNP shell into the cone-shaped core, i.e., for viral maturation (Göttinger et al., 1989 and personal communication; Pettit et al., 1994; Kräusslich et al., 1995).

Fig. 4. Thin section EM of HIV-1 strain IIIB. Between cellular processes and villi, HIV particles reveal different stages in development. Among many mature virions—characterized by the cone-shaped cores (arrows)—are shown budding viruses (arrow heads), cell released immature virions, and one particle representing an intermediate state in maturation (double arrow). Magnification: $\times 45,000$

Fig. 5. EM of HIV comparing negative staining (a), and thin section positive staining, (b). (a) The phage-like particles result from the deleterious effects of the negative staining procedure on the labile envelope of isolated HIV. At the bottom a single free cone-shaped core is seen, still partly wrapped by remnants of the viral envelope. (b) Part of Fig. 4 shown at higher magnification. The two mature HIV reveal core-envelope-links (CEL), lateral bodies (arrow heads) and loss of envelope knobs. Two immature virions are still densely covered with surface knobs, during or just after release from the cell. Magnification: $\times 120,000$

The role of NC For effective encapsidation of the viral genome the NC domain interacts during assembly with the viral RNA via two zinc-finger-like motifs (Clavel and Orenstein, 1990; Gorelick et al., 1990; Bess et al., 1992). In addition, a functional NC is required for the effective assembly (Jowett et al., 1992).

The function of p6 The C-terminal portion of the Pr55 polyprotein forms p6 essential for the last step in viral budding (Götlfinger et al., 1991). Deletions in p6 prohibit both the closure of the spherical RNP shell and particle release. From fine structural studies, p6 was inferred to also form the core-envelope-link (CEL) (Figs. 5b, 7; Nilsson et al., 1992; Höglund et al., 1992).

III. The assembly of the viral envelope

The two viral glycoproteins SU (surface protein, Gp120) and TM (trans-membrane protein, Gp41) are proteolytically processed from the viral Pr160 Env precursor in the Golgi region (Hallenberger et al., 1992). The incorporation of the glycoprotein knobs depends on the N-terminal part of MA (Yu et al., 1992; Dorfman et al., 1994 b; Freed and Martin, 1996). In HIV and a number of other retroviruses (Pepinski and Vogt, 1984) evidence exists for a close proximity of the MA and TM (Bugelski et al., 1995; for review see Hunter, 1994). The membrane spanning region of TM can be replaced by that of

Fig. 6. (a-c) Three different stages in the formation and maturation of HIV illustrating schematically the side-to-side assembly of the Pr55 and Pr155 polypeptide precursor, the immature virion, and the structural rearrangement of the processed proteins in the mature virion. (d) Mature SIV sm particles with cone-shaped cores, lateral bodies and a relatively dense fringe of surface projections. When the glycoprotein knobs are cut tangentially, a trimeric outline of the knobs becomes evident. Magnification: $\times 120,000$. (e, f) Immuno EM localization of HLA class I (beta 2 microglobulin) and of class II (HLA DR), respectively, on budding and mature HIV-1 (from Gelderblom et al., 1987b). Magnification: $\times 80,000$.

Fig. 7. Diagram of murine type C oncogenic retrovirus and HIV-1 combining morphological observations and IEM findings (Frank et al., 1978; Gelderblom, 1991). In contrast to other retroviruses, lentiviruses contain an additional C-terminal p6 domain. The definition of the CEL at the small base of the core (Höglund et al., 1992) is based on both morphological evidence (see Fig. 5b) and molecular genetic findings (Göttlinger et al., 1991), which suggest an essential role for LI p6 in finishing the budding process. Modified from Gelderblom et al., 1987a; Nermut and Hockley, 1996.

an unrelated TM protein without changing viral functions, and the C-terminus of TM appears to be dispensable (Manmano et al., 1994; Wilk et al., 1996). The 72 glycoprotein knobs inserted into the viral envelope form a regular T = 7 laevo pattern (Özel et al., 1988).

The knobs are oligomers of loosely connected TM-SU complexes. The degree of oligomerization, however, is not clear. From cross-linking and molecular-genetic experiments both trimer- and dimer-formation have been reported (Schawaller et al., 1989; Weiss et al., 1990). From morphologic studies a trimeric shape of the knobs became evident (Gelderblom et al., 1987a; Gelderblom et al., 1988; Grief et al., 1989).

IV. Structure and composition of the mature virion

Most probably starting during the assembly of the precursor proteins, Pr55 and Pr155 are cleaved sequentially (Pettit et al., 1994) into the proteins observed in mature HIV/SIV (Henderson et al., 1988, 1992). This processing is required for the rearrangement of the inner proteins which takes place only after release of the immature virion from the cell.

Constituents and organization of the viral core While MA in the mature virion remains tightly bound to the inner leaflet of the lipid-bilayer, the other Gag-proteins condense to form the cone-shaped core typical of members of the lentivirus family (Figs. 4–7; Gelderblom et al., 1987a). In addition, cleavage of Pr155 Gag-Pol liberates multiple copies of the three essential viral enzymes: reverse transcriptase (RT), integrase (IN), and protease (PR). A single virion contains about 1200 molecules of CA and 80 of RT (Layne et al., 1992). In addition, two accessory virus-coded proteins, Vpr and Vpx are present in amounts equivalent to Gag. Vpr is present in all lentiviruses, while Vpx is restricted to HIV-2 and SIV. The incorporation of Vpr depends on C-terminal domains of p6 (Paxton et al., 1993). Vpr and Vpx have been localized by IEM between the core shell and the viral envelope inside the virion (Liska et al., 1994). High resolution localization of these proteins, however, is still required as well as an IEM analysis to clarify the composition of the electron-dense masses ("lateral bodies") along the axis of the cone-shaped cores (Figs. 6b, 7). The fine organization of the viral genome inside the core appears complex (Höglund et al., 1992; Takasaki et al., 1997).

In addition to virus structural and accessory proteins, a number of host cell constituents e.g., actin, ubiquitin, and cyclophilin A are incorporated into the virion. Interestingly, the latter binds specifically

to CA of HIV-1 and was shown to be necessary for the infectivity of HIV-1 strains of the M-group viruses (Franke et al., 1994; Thali et al., 1994).

Constituents of the viral envelope The lipids of the viral envelope are derived from the plasma membrane of the virus-producing cell during budding (Aloia et al., 1993). Besides SU and TM, a number of cellular proteins have been demonstrated to be present on HIV and SIV by immuno EM and other immunological techniques (Gelderblom et al., 1987b; Henderson et al., 1987; Hoxie et al., 1987; Arthur et al., 1992; Meerloo et al., 1993). For example, HLA class I or HLA DR (class II) proteins are tightly inserted during budding into the viral envelope (Fig. 6) and exceed in number the viral SU and TM. Besides masking the virion as a "wolf in sheep's clothing" and having an impact on and interaction with the immune system, these host proteins might also exert other biological functions when present on the virion. Effective purification systems for viruses free of host components are lacking (Bess et al., 1997; Gluschkof et al., 1997).

Virus aging and comparative aspects Because of weak SU-TM interactions, the viral knobs are shed spontaneously from the viral surface (Gelderblom et al., 1985b). For HIV-1 strain IIIB a shedding half-life of 30 hours was determined. This rate easily explains the rapid decay of viral infectivity of IIIB (Layne et al., 1992). The SU knobs on HIV-2 and SIV, however, and in particular on fresh HIV isolates, have been reported to be more tightly associated with the viral envelope.

Except for possible differences in anchorage stability of the knobs, no consistent morphological differences are apparent between HIV-1, HIV-2, and SIV (Palmer and Goldsmith, 1988; Gelderblom, 1991; Nermut and Hockley, 1996). Finally, it should be mentioned that the fine structural data have been collected by EM studies on viruses grown in cell culture. The validity of the structural model still has to be proven by the EM evaluation of infectious plasma HIV.

Acknowledgments

The author thanks Stefan Biel, Thomas Leitner, and Georg Pauli for critical reading of the typescript and Matthias Niedrig for supplying the diagrams A-C of Fig. 6.

References

- [1] Aloia, R.C., H. Tian and F.C. Jensen. 1993. Lipid composition and fluidity of the HIV envelope and host cell plasma membranes. *Proc. Natl. Acad. Sci. USA* 90:5181-5185.
- [2] Arthur, L.O., J.W. Bess, R.C. Sowder, R. Benveniste, D. Mann, J. Chermann and L. Henderson. 1992. Cellular proteins bound to immunodeficiency viruses: implications for pathogenesis and vaccines. *Science* 258:1935-1938.
- [3] Bernhard, W. 1960. The detection and study of tumor viruses with the electron microscope. *Cancer Res.* 20:712-727.
- [4] Bess Jr., J.W., R.J. Gorelick, W.J. Bosche, L.E. Henderson and L.O. Arthur. 1997. Microvesicles are a source of contaminating cellular proteins found in purified HIV-1 preparations. *Virology* 230:134-144.
- [5] Bess, J.W., P.J. Powell, H.J. Issaq, L.J. Schumack, M.K. Grimes, L.E. Henderson and L.O. Arthur. 1992. Tightly bound zinc in human immunodeficiency virus type 1, human T-cell leukemia virus type 1, and other retroviruses. *J. Virol.* 66:840-847.
- [6] Bozzola, J.J. and L.D. Russell. 1991. *Electron Microscopy. Principles and Techniques for Biologists*. Jones and Bartlett Publishers, Boston.
- [7] Bryant, M. and L. Ratner. 1990. Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc. Natl. Acad. Sci. USA* 87:523-527.
- [8] Bugelski, P.J., B.E. Maleeff, A.M. Klinkner, J. Ventre and T.K. Hart. 1995. Ultrastructural evidence of an interaction between env and gag proteins during assembly of HIV type 1. *AIDS Res. Human Retroviruses* 11:55-64.

- [9] Bukrinsky, M.L., S. Haggerty, M.P. Dempsey, N. Sharova, A. Adzhubei, L. Spitz, P. Lewis, D. Goldfarb, M. Emerman and M. Stevenson. 1993. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* 363:666-669
- [10] Chrystie, I.L. and J.D. Almeida. 1988. Further studies of HIV morphology by negative staining. *AIDS* 2:459-464
- [11] Chrystie, I.L. and J.D. Almeida. 1989. Recovery of antigenically reactive HIV-2 cores. *J. Med. Virol.* 27:188-195
- [12] Clavel, F. and J.M. Orenstein. 1990. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.* 64:5230-5234
- [13] Dorfman, T., A. Bukovsky, A. Ohagen, S. Höglund and H.G. Göttinger. 1994 a. Functional domains of the capsid protein of human immunodeficiency virus type 1. *J. Virol.* 68:8180-8187
- [14] Dorfman, T., F. Mammano, W.A. Haseltine and H.G. Göttinger. 1994 b. Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. *J. Virol.* 68:1689-1696
- [15] Dubochet, J., M. Adrian, J.J. Chang, J.C. Homo, J. Lepault, A.W. McDowell and P. Schultz. 1988. Cryo-electron microscopy of vitrified specimens. *Quarterly Review of Biophysics* 21:129-228
- [16] Fäcke, M., A. Janetzko, R.L. Shoeman and H.G. Kräusslich. 1993. A large deletion in the matrix domain of the human immunodeficiency virus gag gene redirects virus particle assembly from the plasma membrane to the endoplasmic reticulum. *J. Virol.* 67:4972-4980
- [17] Frank, H., H. Schwarz, T. Graf, and W. Schäfer. 1978. Properties of mouse leukemia viruses. XV. Electron microscopic studies on the organization of Friend leukemia and other mammalian C-type viruses. *Z. Naturforsch.* 33c:124-138
- [18] Franke, E.K., H.E. Yuan and J. Luban. 1994. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* 372:359-362
- [19] Freed, E.O. and M.A. Martin. 1996. Domains of the human immunodeficiency virus type 1 matrix protein and gp41 cytoplasmic tail required for envelope incorporation into virions. *J. Virol.* 70:341-351
- [20] Fuller, S.D., T. Wilk, B.E. Gowen, H.-G. Kräusslich, and V.M. Vogt. 1997. Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. *Current Biology* 7:729-738.
- [21] Gelderblom, H.R. 1991. Assembly and morphology of HIV: potential effect of structure on viral function. *AIDS* 5:617-637
- [22] Gelderblom, H.R., E.H.S. Hausmann, M. Özel, G. Pauli and M.A. Koch. 1987a. Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology* 156:171-176
- [23] Gelderblom, H.R., M. Özel, E.H.S. Hausmann, T. Winkel, G. Pauli and M.A. Koch. 1988. Fine structure of human immunodeficiency virus (HIV), immunolocalization of structural proteins and virus-cell relation. *Micron Microscopica* 19:41-60
- [24] Gelderblom, H.R., M. Özel and G. Pauli. 1985a. T-Zell-spezifische Retroviren des Menschen: Vergleichende morphologische Klassifizierung und mögliche funktionelle Aspekte. *Bundesgesundheitsblatt* 28:161-171
- [25] Gelderblom, H.R., H. Reupke and G. Pauli. 1985b. Loss of envelope antigens of HTLV-III/LAV, a factor in AIDS pathogenesis? *Lancet* II:1016-1017
- [26] Gelderblom, H.R., H. Reupke, T. Winkel, R. Kunze and G. Pauli. 1987b. MHC-Antigens: Constituents of the envelopes of human and simian immunodeficiency viruses. *Z. Naturforsch.* 42c:1328-1334
- [27] Gentile, M., T. Adrian, A. Scheidler, M. Ewald, F. Dianzani, G. Pauli and H.R. Gelderblom. 1994. Determination of the size of HIV using adenovirus type 2 as an internal length marker. *J. Virol. Meth.* 48: 43-52

- [28] Gheysen, D., E. Jacobs, F. de Foresta, C. Thiriart, M. Francotte and D. Thirnes. 1989. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell* 59:103-112
- [29] Gluschkof, P., Mondor, I., H.R. Gelderblom and Q.J. Sattentau. 1997. Cell membrane vesicles are a major contaminant of gradient-enriched human immunodeficiency virus type-1 preparations. *Virology* 230:125-133
- [30] Gorelick, R.J., S.M. Nigida, J.R. Bess, L.O. Arthur, L.E. Henderson and A. Rein. 1990. Non-infectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* 64:3207-3211
- [31] Goto, T., K. Ikuta, J.J. Zhang, C. Morita, K. Sano, M. Komatsu, S. Kato and M. Nakai. 1990. The budding of defective human immunodeficiency virus (HIV-1) particles from cell clones persistently infected with HIV-1. *Arch. Virol.* 111:87-101.
- [32] Göttinger, H.G., T. Dorfman, J.G. Sodroski and W.A. Haseltine. 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc. Natl. Acad. Sci. USA* 88:3195-3199
- [33] Göttinger, H.G., J.G. Sodroski and W.A. Haseltine. 1989. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* 86:5781-5785
- [34] Grief, C., D.J. Hockley, C.E. Fromholz and P.A. Kitchin. 1989. The morphology of simian immunodeficiency virus as shown by negative staining electron microscopy. *J. Gen. Virol.* 70:2215-2219
- [35] Griffiths, D.J., J.W. Venables, R.A. Weiss and M.T. Boyd. 1997. A novel exogenous retrovirus sequence identified in humans. *J. Virol.* 71:2866-2872
- [36] Hallenberger, S., Bosch, V., H. Angliker, E. Shaw, H. Klenk and W. Garten. 1992. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature* 360:358-361
- [37] Henderson, L.E., R.E. Benveniste, R. Sowder, T.D. Copeland, A.M. Schultz and S. Oroszlan. 1988. Molecular characterization of gag proteins from simian immunodeficiency virus (SIVMne). *J. Virol.* 62:2587-2595
- [38] Henderson, L.E., M.A. Bowers, R.C. Sowder, S.A. Serabyn, D.G. Johnson, J.W. Bess, L.O. Arthur, D.K. Bryant and C. Fenselau. 1992. Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processing, and complete amino acid sequences. *J. Virol.* 66:1856-1865
- [39] Henderson, L.E., R. Sowder, T.D. Copeland, S. Oroszlan, L.O. Arthur, W.G. Robey and P.J. Fischinger. 1987. Direct identification of class II histocompatibility DR proteins in preparations of human T-cell lymphotropic virus type III. *J. Virol.* 61:629-632
- [40] Hockley, D.J., R.D. Wood, J.P. Jacobs and A.J. Garrett. 1988. Electron microscopy of human immunodeficiency virus. *J. Gen. Virol.* 69:2455-2469
- [41] Hosaka, Y. and J. Hosoi. 1983. Study of negatively stained images of Sendai virus nucleocapsids using minimum-dose system. *J. Ultrastruct. Res.* 84:140-150
- [42] Hoxie, J.A., T.P. Fitzharris, P.R. Youngbar, D.M. Matthews, J.L. Rackowski and S.F. Radka. 1987. Nonrandom association of cellular antigens with HTLV-III virions. *Hum. Immunol.* 18:39-52
- [43] Höglund, S., A. Ohagen, K. Lawrence and D. Gabuzda. 1994. Role of vif during packing of the core of HIV-1. *Virology* 201:349-355
- [44] Höglund, S., L.G. Öfverstedt, A. Nilsson, P. Lundqvist, H.R. Gelderblom, M. Özel and U. Skoglund. 1992. Spatial visualization of the maturing HIV-1 core and its linkage to the envelope. *AIDS Res. Human Retroviruses* 8:1-7
- [45] Hunter, E. 1994. Macromolecular interactions in the assembly of HIV and other retroviruses. *Sem. Virol.* 5:71-83

- [46] Jowett, J., D. Hockley, M.V. Nermut and I.M. Jones. 1992. Distinct signals in human immunodeficiency virus type 1 Pr55 necessary for RNA binding and particle formation. *J. Gen. Virol.* 73:3079-3086
- [47] Kräusslich, H.G., M. Fäcke, A.M. Heuser, J. Konvalinka and H. Zentgraf. 1995. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J. Virol.* 69:3407-3419
- [48] Layne, S.P., M.J. Merges, M. Dembo, J.L. Spouge, S.R. Conley, J.P. Moore, J.L. Raina, H. Renz, H.R. Gelderblom and P.L. Nara. 1992. Factors underlying spontaneous inactivation and susceptibility to neutralization of human immunodeficiency virus. *Virology* 189:695-714
- [49] Leis, J., D. Baltimore, J.M. Bishop, J. Coffin, E. Fleissner and S. Golf. 1988. Standardized and simplified nomenclature for proteins common to all retroviruses. *J. Virol.* 62:1808-1809
- [50] Liska, V.D. Spehner, M. Mehtali, D. Schmitt, A. Kim and A.M. Aubertin. 1994. Localization of viral protein X in simian immunodeficiency virus macaque strain and analysis of its packaging requirements. *J. Gen. Virol.* 75:2953-2962
- [51] Mammano, F., A. Ohagen, S. Höglund and H.G. Göttinger. 1994. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J. Virol.* 68:4927-4936
- [52] Marx, P.A., R.J. Munn, K.I. Joy. 1988. Computer emulation of thin section electron microscopy predicts an envelope-associated icosahedral capsid for human immunodeficiency virus. *Lab. Invest.* 58:112-118
- [53] Massiah, M.A., M.R. Starich, C. Paschal, M.F. Summers, A.M. Christensen and W.I. Sundquist. 1994. Three-dimensional structure of the human immunodeficiency virus type 1 matrix protein. *J. Mol. Biol.* 244:198-223
- [54] Meerloo, T., M.A. Sheikh, A.C. Bloem, A. de Ronde, M. Schutten, C.A. van Eijs, P.J. Roholl, P. Joling, J. Goudsmit and H.J. Schuurman. 1993. Host cell membrane proteins on human immunodeficiency virus type 1 after in vitro infection of H9 cells and blood mononuclear cells. An immuno-electron microscopy study. *J. Gen. Virol.* 74:129-135
- [55] Müller, M. 1992. The integrating power of cryofixation-based electron microscopy in biology. *Acta Microscópica* 1:37-44
- [56] Nakai, M., T. Goto. 1996. Ultrastructure and morphogenesis of human immunodeficiency virus. *J. Electron Microsc.* 45:247-257
- [57] Nermut, M.V., C. Grief, S. Hashmi and D. Hockley. 1993. Further evidence of icosahedral symmetry in human and simian immunodeficiency virus. *AIDS Res. Hum. Retroviruses* 9:929-938
- [58] Nermut, M.V. and D.J. Hockley. 1996. Comparative morphology and structural classification of retroviruses. *Curr. Top. Microbiol. Immunol.* 214:1-24
- [59] Nermut, M.V., D.J. Hockley, J.B. Jowett, I.M. Jones, M. Garreau and D. Thomas. 1994. Fullerene-like organization of the HIV gag-protein shell in virus-like particles produced by recombinant baculovirus. *Virology* 198:288-296
- [60] Niedrig, M., H. Bickhard, H.R. Gelderblom, G. Pauli, S. Modrow, A. V. Pöblitzki, R. Wagner and H. Wolf. 1993. Assembly of HIV gag proteins: a target for antiviral therapy? *Vaccines 93*, Cold Spring Harbor Laboratory Press: 209-214
- [61] Nilsson, A., P. Lundqvist, A. Love, B. Torfason, H.R. Gelderblom and S. Höglund. 1992. Spatial visualization of progressive states of maturing lentivirus. *Vet. Microbiol.* 33:333-340
- [62] Özel, M., G. Pauli and H.R. Gelderblom. 1988. The organization of the envelope projections on the surface of HIV. *Arch. Virol.* 100:255-266
- [63] Palmer, E. and C.S. Goldsmith. 1988. Ultrastructure of human retroviruses. *J. Electron. Microsc. Tech.* 8:3-5

- [64] Paxton, W., R.L. Connor and N. Landau. 1993. Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis. *J. Virol.* 67:7229-7237
- [65] Papinsky, R.B. and V.M. Vogt. 1984. Fine structure analysis of lipid-protein and protein-protein interactions of gag protein p19 of the avian sarcoma and leukemia viruses by cyanogen bromide mapping. *J. Virol.* 52:145-153
- [66] Pettit, S.C., M.D. Moody, R.S. Webbe, A.H. Kaplan, P.V. Nantermet, C.A. Klein and R.A. Swanstrom. 1994. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.* 68:8017-8027
- [67] Reicin, A.S., S. Paik, R.D. Berkowitz, J. Luban, Lowy, I. and S.P. Goff. 1995. Linker insertion mutations in the human immunodeficiency virus type 1 gag gene: effects on virion particle assembly, release, and infectivity. *J. Virol.* 65:4350-4358
- [68] Renz, H., G. Weinberg, M. Özel, H.R. Gelderblom, G. Pauli, and M. Müller. 1993. High resolution scanning electron microscopy of HIV after conventional- and cryo-preparation. Comparison with transmission electron microscopical results. *Beitr. Elektronenmikroskop. Direktabb. Oberfl.* 26: 209-219
- [69] Schawaller, M., G.E. Smith, J.J. Skehel and D.C. Wiley. 1989. Studies with crosslinking reagents on the oligomeric structure of the env glycoprotein of HIV. *Virology* 172:367-369
- [70] Smith, A.J., N. Srinivasakumar, M.L. Hammariskjold and D. Rekosh. 1993. Requirements for incorporation of Pr160 gag-pol from human immunodeficiency virus type 1 into virus-like particles. *J. Virol.* 67:2266-2275
- [71] Takasaki, T., Kurane, I., H. Aihara, N. Ohkawa and J. Yamaguchi. 1997. Electron microscopy study of human immunodeficiency virus type 1 (HIV-1) core structure: two RNA strands in the core of mature and budding particles. *Arch. Virol.* 142:375-382
- [72] Thali, M., A. Bukovsky, E. Kondo, B. Rosenwirth, C. Walsh, J. Sodroski and H.G. Göttinger. 1994. Functional association of cyclophilin A with HIV-1 virions. *Nature* 372:363-365
- [73] Weiss, C.D., J.A. Levy and J.M. White. 1990. Oligomeric organization of gp120 on infectious human immunodeficiency virus type 1 particles. *J. Virol.* 64:5674-5677
- [74] Wilk, T., T. Pfeiffer, A. Bukovsky, G. Moldenhauer and V. Bosch. 1996. Glycoprotein incorporation and HIV-1 infectivity despite exchange of the gp 160 membrane-spanning domain. *Virology* 218:269-274
- [75] Wills, J.W. and R.C. Craven. 1991. Form, function and use of retroviral Gag proteins. *AIDS* 5:639-654
- [76] Yu, X., X. Zuan, Z. Matsuda, T.H. Lee and M. Essex. 1992. The matrix protein of human immunodeficiency virus type 1 is required for incorporation of viral envelope protein into mature virions. *J. Virol.* 66:4966-4971
- [77] Yuan, X., X. Yu, T.H. Lee and M. Essex. 1993. Mutations in the N-terminal region of human immunodeficiency virus type 1 matrix protein block intracellular transport of the Gag precursor. *J. Virol.* 67:6387-6394
- [78] Zhang, W.H., D. Hockley, M.V. Nermut, Y. Morikawa and I.A. Jones. 1996. Gag-Gag interactions in the C-terminal domain of human immunodeficiency virus type 1 p24 capsid antigen are essential for Gag particle assembly. *J. Gen. Virol.* 77:743-751
- [79] Zhou, K., L.J. Parent, J.W. Wills and M.D. Resh. 1994. Identification of a membrane binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J. Virol.* 68:2556-2569.

Reagents for HIV/SIV Vaccine Studies

Carla Kuiken and Satish Pillai

Theoretical Biology and Biophysics, Group T-10, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545

Progress in developing a vaccine against HIV has been much slower than expected. There are many reasons why development of a vaccine against HIV is extremely difficult.

- There is no clear protective immunity even after natural infection
- The virus is highly variable
- The virus is able to escape aspects of the immune pressure
- It can infect different cell types over the course of infection
- Not all infected cells are recognizable by the immune system
- The virus affects cells of the immune system itself
- There is a risk of inducing infection-enhancing antibodies
- There is no perfect animal model for the infection

In spite of years of research into HIV, much is still unknown about the virus. What exactly makes it so pathogenic? What are the mechanisms that appear to protect some people against infection? What determines the dramatically different rates of progression to AIDS after infection? An important part of the research into these questions is done using animal, mostly primate, models. Unfortunately, no single primate model is ideally suited for the job.

Chimpanzees are the primates genetically most closely related to humans. They can be infected with HIV, but progress to AIDS very slowly after HIV infection. Thus far, only one chimpanzee has died from an AIDS-like illness, ten years after infection with HIV-1 (Villinger et al., 1997). This is a drawback in the light of recent discussions about the lack of necessity for sterilizing immunity (the total prevention of infection, as opposed to partial immunity, where infection after vaccination can occur but is less severe). Even a vaccine that does not prevent infection may still boost the immune system enough to prevent or delay disease in the event of infection. In fact, vaccines that protect against other viral diseases generally do not induce sterilizing immunity (polio, measles, varicella, smallpox). Since chimpanzees develop disease only very slowly, this effect cannot easily be studied using chimpanzee models. In addition, chimpanzees are a threatened species, they are very expensive, they cannot be put down unless they are extremely ill, and since HIV usually does not make them sick they must be kept in isolation for their lifetime, up to 40 years. Being HIV-infected, they pose a risk for their caretakers. Finally, the use of chimpanzees in potentially lethal experiments gives rise to complicated ethical considerations. After the initial, disappointing results involving HIV-1 vaccines (usually subunit vaccines) based on lab strains, the chimpanzee trials have largely been abandoned in favor of the cheaper and less demanding macaque monkey.

Macaques in the wild are not natural hosts for SIV, and most SIV strains are highly pathogenic to them, so they provide a very good model system to study pathogenesis. HIV-2, which is much more similar to SIV than HIV-1, quickly produces an AIDS-like disease in the pig-tailed macaque, *M. nemestrina*. This is also the only macaque species that can be productively infected with HIV-1, although thus far without developing AIDS (Heeney, 1996). Challenge strains used in macaques are most frequently derived from SIVsm, usually after several passages in macaques. A recent development is the use of chimeric SIV/HIV (SHIV) viruses that can infect macaques, cause an AIDS-like disease just like other SIVs, but share a varying number of genes with HIV-1. Only in the last few years have SHIV strains become available that are sufficiently pathogenic to be used as challenge strains in vaccine trials (see Fig. 1c).

In many cases the results of vaccine trials are different in different animal models. It is not always known when this is a virus effect (SIV vs. HIV), and when the host immune system responds differently to different immunizations and challenges. This problem obviously makes the interpretation

Figure 1. An overview of the derivation of SIVmac (a), SIVsm (b) and SHIV (c) strains. Rectangles indicate passages in a rhesus macaque, ovals indicate isolates or clones derived from these. Figures 1a and 1b were updated from (Schultz & Hu, 1993); Figure 1c was produced by Dr. Alan Schultz (NIAID) based on information from the original references. The figures were reprinted with permission. References: SIVmac (Figure 1a): 251/BK28 (Kornfeld et al., 1987); IA11 (Marthas et al., 1990); 32H-C8 (Rud et al., 1992); 32H-J5 (Rud et al., 1992); 139 (Kestler et al., 1990); 239 derivatives (Desrosiers et al., 1998); 17E (Anderson et al., 1993); 316 (Mori et al., 1992). SIVsm (Figure 1b): B670 (Conway et al., 1991); H4 (Novembre et al., 1993); PB114 (Dewhurst et al., 1990); PB114-Δnef (Novembre et al., 1996); E543 (Hirsch et al., 1997). SHIV (Figure 1c): KU1 (Joag et al., 1996); KU2 (Raghavan et al., 1997); 3.2 clone (Sodroski, unpublished); 89.6 (Reimann et al., 1996b); 89.6P (Reimann et al., 1996a); KB9 (Karlsson et al., 1997); SP33 (Luciw et al., 1995); SP33A Luciw, unpublished; DH12 (Shibata et al., 1997a).

of vaccine studies in animal models difficult. Furthermore, subtle differences in the design of the vaccine (adjuvants, quantity and variability of antigens) and the trials (frequency and timing of inoculations, timing and severity of challenge, virus type, cell-free or cell associated) all can influence the outcome.

By far the best protection against both cell-free and cell-bound infection with HIV and SIV so far has been obtained using live attenuated virus vaccines. The protection is not type-specific, and SIV-vaccinated monkeys can even be protected against superinfection with virulent SHIV strains (Heeney, 1996). Live attenuated virus vaccines basically establish an infection, which appears to be the best way to prime the immune system to react to the presence of an alien invader; both humoral and cellular immune responses are optimized in this situation. However, there are many risks associated with the use of attenuated virus vaccines. First of all, it is uncertain that the attenuated virus is really non-pathogenic. It has been reported some time ago that virus that was apathogenic in adult monkeys could cause disease in neonates (Ruprecht et al., 1996). According to a recent report (Cohen, 1997), several adult monkeys that have been vaccinated with an attenuated virus several years ago have now also fallen ill. Apparently, this is not a matter of repair of the crippling mutations, but rather an inability of the immune system to control even very weak immunodeficiency viruses in the long run. However, even if this approach is ultimately deemed to be too dangerous, vaccine trials using attenuated SIV variants provide very important insights into the nature of protective immunity against immunodeficiency viruses.

In recent years, many successes have been reported using attenuated and SHIV based vaccines. Without striving for completeness, we cite some examples of successful vaccination attempts here. Successful protection from vaginal challenge in rhesus macaques has been reported by Miller et al. (1997). Dunn et al. (1997) reported that macaques infected with the SIVmac251 clone BK28 were protected against subsequent challenge with a chimeric strain, SHIVsbg. No virus could be recovered from three of the five challenged monkeys, although SHIV-infected cells were found. Two monkeys did develop viremia, but the viral load was 100-fold reduced compared to unvaccinated control animals. Comparable results were reported by Stevens et al. (1997) using attenuated SIVmac or SHIV as a vaccine and a pathogenic SHIV-KU1 for challenge. Interestingly, animals in this study that were productively infected with a pathogenic virus were later unable to fend off superinfection with a second virulent strain. Quesada-Rolander reported protection of monkeys vaccinated with SHIV-4 against rectal challenge with a virulent SIVsm (Quesada-Rolander et al., 1996). From two out of four monkeys, no virus could be recovered a year after challenge; two other showed initial viremia, which was later suppressed. Linhart et al. reported on an attempt to use attenuated viruses in a post-exposure vaccination, using simultaneous inoculation. The study showed no effect of co-injection of a non-virulent variant along with a virulent one (Linhart et al., 1997). An attempt was made to prevent infection of newborn macaques by vaccinating pregnant monkeys with the attenuated strain SIVmac1A11. Two out of three newborns were protected against mucosal challenge after birth (Van Rompay et al., 1996). Stahl-Hennig et al. reported sterilizing immunity in 4 out of 8 macaques vaccinated with the attenuated strain 32H-C8 (Stahl-Hennig et al., 1996). Putkonen et al. report protection of cynomolgous monkeys against SIV challenge using an attenuated HIV-2 strain as a vaccine (Putkonen et al., 1995). Other successful vaccinations have been reported using SIVmac1A11 (Otsynla et al., 1996), SIVmac316-Δnef and 239-Δ3 (Wyand et al., 1996) and several other attenuated SIV strains.

The number of different viral strains used in vaccine research is rapidly growing. In Figures 1-3 we present an overview of the virus strains that are frequently used in these studies, and their derivation. A distinction must be made between strains used for vaccination, which obviously must be non-pathogenic, and strains used for challenge, which tend to be pathogenic. However, it should be mentioned that some strains have been used both as vaccine and as challenge strains. Pathogenicity is a relative concept, and depends on the virus, the host species, and the individual host (most probably on characteristics of the MHC). For example, SIVsm is not pathogenic in its natural host, the sooty mangabey, but can be highly pathogenic to macaques.

SIVmac viruses are used most extensively in these studies. The SIVmac isolates 251 and 32H both have a reduced pathogenic counterpart, clones 1A11 and C8, respectively. These clones are genetically very similar to the quasispecies they were derived from, but they have one or more attenuating genetic

Figure 2 Schematic representation of common attenuated SIV strains used in vaccine research. Deletions are indicated in white, mutations are shown as black lines. Mutations may include insertion or deletion of a stop codon.

Chimeric SIV/HIV Strains

Chimeric SIV/HIV Strains (cont'd)

Name	HIV strain	Construct	Reference
NM-8	NL43		Shibata, <i>ARHR</i> 8: 403 (1992)
NM-9	NL43		Shibata, <i>ARHR</i> 8: 403 (1992)
NM-10	NL43		Shibata, <i>ARHR</i> 8: 403 (1992)
NS-1	NL43		Shibata, <i>ARHR</i> 8: 403 (1992)
NS-3	NL43		Shibata, <i>ARHR</i> 8: 403 (1992)
NS-6	NL43		Shibata, <i>ARHR</i> 8: 403 (1992)
NS-7	NL43		Shibata, <i>ARHR</i> 8: 403 (1992)
SHIV _{MD1}	DH12/NL43		Shibata, <i>J Virol</i> 71: 8141 (1997)
SHIV _{Han-2}	Han-2/NL43		Kuwata, <i>AIDS</i> 10: 1331 (1996)
SHIV _{SF2}	SF2		Kuwata, <i>AIDS</i> 10: 1331 (1996)
SHIV _{SF13}	SF13		Kuwata, <i>AIDS</i> 10: 1331 (1996)
NM-3m	NL43		Kuwata, <i>J Gen Virol</i> 76: 2151 (1995)

Figure 3. Schematic representation of common chimeric HIV/SIV strains used in vaccine research. The information was obtained from the references shown in the Figure. Although every attempt was made to contact the authors to clarify ambiguities, in some cases the exact location of the borders between HIV and SIV could not be established. The uncertain regions are shown in gray. The HIV genes used in these constructs are from the clones HXB2, 89.6, SF33, and DH12.

deletions. Another important SIVmac isolate, 239, is a clonal isolate from rhesus monkey #239. SIVmac239 is pathogenic, but a long series of reduced- or non-pathogenic strains with varying number of deletions in the genome has been derived from it. These strains cover a spectrum of pathogenicity, ranging from highly pathogenic to apparently non-infectious (unable to replicate in the host) (Desrosiers et al., 1998). It has recently become apparent that monkeys infected with SIVmac239-Δ3, in the lower mid range of the pathogenicity spectrum, do develop AIDS after several years (Cohen, 1997). Two macrophage-tropic variants have also been derived from SIVmac-239: SIVmac316 and 17E (see Figure 1a).

SIVsm isolates can be highly pathogenic to macaques. Pathogenic challenge stocks in this group are usually bulk (rather than clonal) isolates, passaged in one of several macaques. Genetic clones of this group (such as SIVsmH4) tend to be much less pathogenic. An important and highly pathogenic strain is SIVsmPBj14, which kills a majority of infected macaques at primary infection, within a few weeks; monkeys that survive primary infection usually die of an AIDS-like illness within two years. Other isolates that have been used as challenge stocks are B670 and E660. Derivation of commonly used isolates from this group is shown in Figure 1b.

SHIV strains have recently been developed that are highly pathogenic. SHIV strains have traditionally been used for both protection and challenge. Especially SHIVs that have been composed from a CXCR4-using HIV strain appear to be pathogenic. The viruses usually are pathogenic only in their derivative form, after passage in several monkeys. Derivation of some frequently used SHIV strains is shown in Figure 1c.

In recent years the repertoire of non-pathogenic vaccine strains has been extended by the creation of artificially attenuated virus variants. This is usually done either by creating stop codons in non-vital sections, or deleting sections from the genome of a virulent strain. Since these attenuated virus strains often are not included in Genbank, we have produced an alignment for them. The alignment itself is not reproduced here, but it will be made available on the web page (<http://hiv-web.lanl.gov>). Instead, Figure 2 shows a schematic representation of attenuated SIV strains. The diagram shows where changes have been made or documented with respect to the wild type. The number of SHIV chimeras is also growing rapidly; a fairly large number of pathogenic SHIV strains is now available. The diagram in Figure 3 gives an overview of SHIV strains that are presently in use in the vaccine field, and indicates which section of the genomes are derived from HIV-1 (and which strain of HIV-1), and which from a SIV strain.

Notes

SIVmac239, a frequently used pathogenic strain, contains a stop codon (TAA) in *nef*.

BK28 is a clone from the SIVmac251 bulk isolate; it is the only sequence available from SIVmac251.

SHIV KU1 and KU2 are passages of SHIV-4.

SHIV 89.6P and KB9 are passages of SHIV89.6.

SP33A is a passage of SP33.

Acknowledgments

We are grateful to Drs. Jim Bradac and Alan Schultz from NIAID for providing invaluable background information and helpful suggestions, and Dr. Marta Marthas from the California Regional Primate Research Center of the University of California at Davis for critical reading of the manuscript.

References

- Anderson, M. G., Hauer, D., Sharma, D. P., Joag, S. V., Narayan, O., Zink, M. C. and Clements, J. E. (1993). Analysis of envelope changes acquired by SIVmac239 during neuroadaptation in rhesus macaques. *Virology* 195:616-26.
- Beer, B., Baier, M., zur Megede, J., Norley, S. and Kurth, R. (1997). Vaccine effect using a live attenuated nef-deficient simian immunodeficiency virus of African green monkeys in the absence of detectable vaccine virus replication in vivo. *Proc Natl Acad Sci USA* 94:4062-7.
- Chakrabarti, L., Baptiste, V., Khatissian, E., Cumont, M. C., Aubertin, A. M., Montagnier, L. and Hurtrel, B. (1995). Limited viral spread and rapid immune response in lymph nodes of macaques inoculated with attenuated simian immunodeficiency virus. *Virology* 213:535-48.
- Cohen, J. (1997). Weakened SIV vaccine still kills [news]. *Science* 278:24-5.
- Conway, M. D., Davison-Fairburn, B., Martin, L. N., Insler, M. S. and Murphey-Corb, M. (1991). Infection of rhesus monkeys with topical instillation of simian immunodeficiency virus (SIV)B670 into the conjunctival sac. *J Med Primatol* 20:152-5.
- Desrosiers, R. C. (1992). HIV with multiple gene deletions as a live attenuated vaccine for AIDS. *AIDS Res Hum Retroviruses* 8:411-21.
- Desrosiers, R. C., Lifson, J. D., Gibbs, J. S., Czajak, S. C., Howe, A. Y., Arthur, L. O. and Johnson, R. P. (1998). Identification of highly attenuated mutants of simian immunodeficiency virus. *J Virol* 72:1431-7.
- Dewhurst, S., Embretson, J. E., Anderson, D. C., Mullins, J. I. and Fultz, P. N. (1990). Sequence analysis and acute pathogenicity of molecularly cloned SIVSMM-PBj14 [see comments]. *Nature* 345:636-40.
- Dunn, C. S., Beyer, C., Kieny, M. P., Gloeckler, L., Schmitt, D., Gut, J. P., Kim, A. and Aubertin, A. M. (1996). High viral load and CD4 lymphopenia in rhesus and cynomolgus macaques infected by a chimeric primate lentivirus constructed using the env, rev, tat, and vpu genes from HIV-1 Lai. *Virology* 223:351-61.
- Dunn, C. S., Hurtrel, B., Beyer, C., Gloeckler, L., Ledger, T. N., Moog, C., Kieny, M. P., Mehtali, M., Schmitt, D., Gut, J. P., Kim, A. and Aubertin, A. M. (1997). Protection of SIVmac-infected macaque monkeys against superinfection by a simian immunodeficiency virus expressing envelope glycoproteins of HIV type 1. *AIDS Res Hum Retroviruses* 13:913-22.
- Heeney, J. L. (1996). Primate models for AIDS vaccine development. *AIDS* 10 Suppl A, S115-22.
- Hirsch, V., Adger-Johnson, D., Campbell, B., Goldstein, S., Brown, C., Elkins, W. R. and Montefiori, D. C. (1997). A molecularly cloned, pathogenic, neutralization-resistant simian immunodeficiency virus, SIVsmE543-3. *J Virol* 71:1608-20.
- Hoch, J., Lang, S. M., Weeger, M., Stahl-Hennig, C., Coulibaly, C., Dittmer, U., Hunsmann, G., Fuchs, D., Muller, J., Sopper, S. and et al. (1995). vpr deletion mutant of simian immunodeficiency virus induces AIDS in rhesus monkeys. *J Virol* 69:4807-13.
- Joag, S. V., Li, Z., Foresman, L., Stephens, E. B., Zhao, L. J., Adany, I., Pinson, D. M., McClure, H. M. and Narayan, O. (1996). Chimeric simian/human immunodeficiency virus that causes progressive loss of CD4+ T cells and AIDS in pig-tailed macaques. *J Virol* 70:3189-97.
- Karlsson, G. B., Halloran, M., Li, J., Park, I. W., Gomila, R., Reimann, K. A., Axthelm, M. K., Liff, S. A., Letvin, N. L. and Sodroski, J. (1997). Characterization of molecularly cloned simian-human immunodeficiency viruses causing rapid CD4+ lymphocyte depletion in rhesus monkeys. *J Virol* 71:4218-25.
- Kestler, H., Kodama, T., Ringler, D., Marthas, M., Pedersen, N., Lackner, A., Regier, D., Sehgal, P., Daniel, M., King, N. and et al. (1990). Induction of AIDS in rhesus monkeys by molecularly cloned simian immunodeficiency virus. *Science* 248:1109-12.
- Kornfeld, H., Riedel, N., Viglianti, G. A., Hirsch, V. and Mullins, J. I. (1987). Cloning of HTLV-4 and its relation to simian and human immunodeficiency viruses. *Nature* 326:610-3.

- Kuwata, T., Igarashi, T., Ido, E., Jin, M., Mizuno, A., Chen, J. and Hayami, M. (1995). Construction of human immunodeficiency virus 1/simian immunodeficiency virus strain mac chimeric viruses having vpr and/or nef of different parental origins and their in vitro and in vivo replication. *J Gen Virol* 76:2181-91.
- Kuwata, T., Shioda, T., Igarashi, T., Ido, E., Ibuki, K., Enose, Y., Stahl-Hennig, C., Hunsmann, G., Miura, T. and Hayami, M. (1996). Chimeric viruses between SIVmac and various HIV-1 isolates have biological properties that are similar to those of the parental HIV-1. *AIDS* 10:1331-7.
- Li, J., Lord, C. I., Haseltine, W., Letvin, N. L. and Sodroski, J. (1992). Infection of cynomolgus monkeys with a chimeric HIV-1/SIVmac virus that expresses the HIV-1 envelope glycoproteins. *J Acquir Immune Defic Syndr* 5:639-46.
- Linhart, H., Gundlach, B. R., Sopper, S., Dittner, U., Matz-Rensing, K., Kuhn, E. M., Muller, J., Hunsmann, G., Stahl-Hennig, C. and Uberla, K. (1997). Live attenuated SIV vaccines are not effective in a postexposure vaccination model. *AIDS Res Hum Retroviruses* 13:593-9.
- Luciw, P. A., Pratt-Lowe, E., Shaw, K. E., Levy, J. A. and Cheng-Mayer, C. (1995). Persistent infection of rhesus macaques with T-cell-line-tropic and macrophage-tropic clones of simian/human immunodeficiency viruses (SHIV). *Proc Natl Acad Sci USA* 92:7490-4.
- Luciw, P. A., Shaw, K. E., Unger, R. E., Planellas, V., Stout, M. W., Lackner, J. E., Pratt-Lowe, E., Leung, N. J., Banapour, B. and Marthas, M. L. (1992). Genetic and biological comparisons of pathogenic and nonpathogenic molecular clones of simian immunodeficiency virus (SIVmac). *AIDS Res Hum Retroviruses* 8:395-402.
- Marthas, M. L., Sutjipto, S., Higgins, J., Lohman, B., Torten, J., Luciw, P. A., Marx, P. A. and Pedersen, N. C. (1990). Immunization with a live, attenuated simian immunodeficiency virus (SIV) prevents early disease but not infection in rhesus macaques challenged with pathogenic SIV. *J Virol* 64:3694-700.
- Miller, C. J., McChesney, M. B., Lu, X., Dailey, P. J., Chutkowski, C., Lu, D., Brosio, P., Roberts, B. and Lu, Y. (1997). Rhesus macaques previously infected with simian/human immunodeficiency virus are protected from vaginal challenge with pathogenic SIVmac239. *J Virol* 71:1911-21.
- Mori, K., Ringler, D. J., Kodama, T. and Desrosiers, R. C. (1992). Complex determinants of macrophage tropism in env of simian immunodeficiency virus. *J Virol* 66:2067-75.
- Novembre, F. J., Johnson, P. R., Lewis, M. G., Anderson, D. C., Klumpp, S., McClure, H. M. and Hirsch, V. M. (1993). Multiple viral determinants contribute to pathogenicity of the acutely lethal simian immunodeficiency virus SIVsmmPBj variant. *J Virol* 67:2466-74.
- Novembre, F. J., Lewis, M. G., Saucier, M. M., Yalley-Ogunro, J., Brennan, T., McKinnon, K., Bellah, S. and McClure, H. M. (1996). Deletion of the nef gene abrogates the ability of SIV smmPBj to induce acutely lethal disease in pigtail macaques. *AIDS Res Hum Retroviruses* 12:727-36.
- Otsyula, M. G., Miller, C. J., Tarantal, A. F., Marthas, M. L., Greene, T. P., Collins, J. R., van Rompay, K. K. and McChesney, M. B. (1996). Fetal or neonatal infection with attenuated simian immunodeficiency virus results in protective immunity against oral challenge with pathogenic SIVmac251. *Virology* 222:275-8.
- Putkonen, P., Walther, L., Zhang, Y. J., Li, S. L., Nilsson, C., Albert, J., Biberfeld, P., Thorstensson, R. and Biberfeld, G. (1995). Long-term protection against SIV-induced disease in macaques vaccinated with a live attenuated HIV-2 vaccine. *Nat Med* 1:914-8.
- Quesada-Rolander, M., Makitalo, B., Thorstensson, R., Zhang, Y. J., Castanos-Velez, E., Biberfeld, G. and Putkonen, P. (1996). Protection against mucosal SIVsm challenge in macaques infected with a chimeric SIV that expresses HIV type 1 envelope. *AIDS Res Hum Retroviruses* 12:993-9.
- Raghavan, R., Stephens, E. B., Joag, S. V., Adany, I., Pinson, D. M., Li, Z., Jia, F., Sahni, M., Wang, C., Leung, K., Foresman, L. and Narayan, O. (1997). Neuropathogenesis of chimeric simian/human immunodeficiency virus infection in pig-tailed and rhesus macaques. *Brain Pathol* 7:551-61.

- Ranjbar, S., Jones, S., Stott, E. J. and Almond, N. (1997). The construction and evaluation of SIV/HIV chimeras that express the envelope of European HIV type 1 isolates. *AIDS Res Hum Retroviruses* 13:797-800.
- Reimann, K. A., Li, J. T., Veazey, R., Halloran, M., Park, I. W., Karlsson, G. B., Sodroski, J. and Letvin, N. L. (1996a). A chimeric simian/human immunodeficiency virus expressing a primary patient human immunodeficiency virus type 1 isolate env causes an AIDS-like disease after in vivo passage in rhesus monkeys. *J Virol* 70:6922-8.
- Reimann, K. A., Li, J. T., Voss, G., Lekutis, C., Tenner-Racz, K., Racz, P., Lin, W., Montefiori, D. C., Lee-Parritz, D. E., Lu, Y., Collman, R. G., Sodroski, J. and Letvin, N. L. (1996b). An env gene derived from a primary human immunodeficiency virus type 1 isolate confers high in vivo replicative capacity to a chimeric simian/human immunodeficiency virus in rhesus monkeys. *J Virol* 70:3198-206.
- Rud, E. W., Yon, J. R., Larder, B. A., Clarke, B. E., Cook, N. and Cranage, M. (1992). Infectious molecular clones of SIVmac32: Nef deletion controls ability to reisolate virus from rhesus macaques. In *Vaccines*, pp. 229-235. Edited by F. Brown, R. M. Chanock, H. S. Ginsberg and R. A. Lerner. New York: Cold Spring Harbor Laboratory Press.
- Ruprecht, R. M., Baba, T. W., Liska, V., Bronson, R., Penninck, D. and Greene, M. F. (1996). "Attenuated" simian immunodeficiency virus in macaque neonates. *AIDS Res Hum Retroviruses* 12:459-60.
- Schultz, A. M. and Hu, S. L. (1993). Primate models for HIV vaccines. *AIDS* 7 Suppl 1, S161-70.
- Shibata, R. and Adachi, A. (1992). SIV/HIV recombinants and their use in studying biological properties. *AIDS Res Hum Retroviruses* 8:403-9.
- Shibata, R., Kawamura, M., Sakai, H., Hayami, M., Ishimoto, A. and Adachi, A. (1991). Generation of a chimeric human and simian immunodeficiency virus infectious to monkey peripheral blood mononuclear cells. *J Virol* 65:3514-20.
- Shibata, R., Maldarelli, F., Siemon, C., Matano, T., Parta, M., Miller, G., Fredrickson, T. and Martin, M. A. (1997a). Infection and pathogenicity of chimeric simian-human immunodeficiency viruses in macaques: determinants of high virus loads and CD4 cell killing. *J Infect Dis* 176:362-73.
- Shibata, R., Siemon, C., Czajak, S. C., Desrosiers, R. C. and Martin, M. A. (1997b). Live, attenuated simian immunodeficiency virus vaccines elicit potent resistance against a challenge with a human immunodeficiency virus type 1 chimeric virus. *J Virol* 71:8141-8.
- Stahl-Hennig, C., Dittmer, U., Nisslein, T., Petry, H., Jurkiewicz, E., Fuchs, D., Wachter, H., Matz-Rensing, K., Kuhn, E. M., Kaup, F. J., Rud, E. W. and Hunsmann, G. (1996). Rapid development of vaccine protection in macaques by live-attenuated simian immunodeficiency virus. *J Gen Virol* 77:2969-81.
- Stephens, E. B., Joag, S. V., Atkinson, B., Sahni, M., Li, Z., Foresman, L., Adany, I. and Narayan, O. (1997). Infected macaques that controlled replication of SIVmac or nonpathogenic SHIV developed sterilizing resistance against pathogenic SHIVKU-1. *Virology* 234:328-39.
- Van Rompay, K. K., Otsyula, M. G., Tarara, R. P., Canfield, D. R., Berardi, C. J., McChesney, M. B. and Marthas, M. L. (1996). Vaccination of pregnant macaques protects newborns against mucosal simian immunodeficiency virus infection. *J Infect Dis* 173:1327-35.
- Villinger, F., Brar, S. S., Brice, G. T., Chikkala, N. R., Novembre, F. J., Mayne, A. E., Bucur, S., Hillyer, C. D. and Ansari, A. A. (1997). Immune and hematopoietic parameters in HIV-1-infected chimpanzees during clinical progression toward AIDS. *J Med Primatol* 26:11-8.
- Wyand, M. S., Manson, K. H., Garcia-Moll, M., Montefiori, D. and Desrosiers, R. C. (1996). Vaccine protection by a triple deletion mutant of simian immunodeficiency virus. *J Virol* 70:3724-33.

gp41, A Multifunctional Protein Involved in HIV Entry and Pathogenesis

Eric Hunter

Department of Microbiology and Center for AIDS Research, University of Alabama at Birmingham, Birmingham, AL 35294-2170

Introduction

The entry of HIV into a target cell represents the key initial step in the replication cycle of this virus. The entry process involves a coordinated series of molecular interactions between the two components of the virus glycoprotein complex (gp120(SU)/gp41(TM)) and the two components of the receptor complex (CD4 and a chemokine receptor). It results in the fusion of the viral membrane with that of the host cell and the introduction of the viral replicative machinery into the cytoplasm. It has been well established over the past several years that conserved domains of gp120 come together to form the binding site for amino-terminal residues of the CD4 molecule (Moore et al., 1993). For most HIV and SIV isolates, this SU-CD4 interaction defines the cell type to which the virus binds and ultimately enters (Sattentau and Moore, 1993). Recent data have shown, however, that for the process of membrane fusion to proceed, a second interaction, with a member of the chemokine receptor family of multiple membrane-spanning proteins, must occur. Natural isolates of HIV-1 which primarily infect macrophages and monocytes but lack the ability to infect T-cell lines utilize CCR5, while T cell-line adapted and syncytium-inducing (SI) viruses primarily utilize CXCR4 (reviewed in Berger, 1997; Doms and Peiper, 1997; Moore, 1997). Binding of the receptor complex results in conformational changes in both gp120 and in the TM protein gp41. By analogy with the hemagglutinin of influenza virus, the TM protein is thought to be the engine that drives the membrane fusion process, and it is this multifaceted molecule that is the subject of this review.

Biosynthesis of Env

The SU and TM components of the envelope glycoprotein complex are synthesized, as a single, co-translationally glycosylated, polyprotein precursor (gp160), on ribosomes associated with the endoplasmic reticulum (ER). Individual precursor molecules, shown schematically in Figure 1, initiate folding, presumably in the presence of ER-associated chaperonins, and assemble to form an oligomeric structure (Hunter and Swanstrom, 1990). Whether this is a trimeric or tetrameric molecule remains unresolved, although recent data suggest that a trimer is the biologically relevant form of the oligomer and this would be consistent with the findings for other retroviruses (see below). Mutagenic studies in both HIV and Rous sarcoma virus suggest that it is regions within the TM domain that direct and stabilize the oligomerization process (Earl and Moss, 1993; Einfeld and Hunter, 1994; Einfeld and Hunter, 1997). It is likely that this region is structurally conserved between HIV and SIV, since mixed Env oligomers have been described (Doms et al., 1990). Surprisingly, the membrane-spanning domain is not required for oligomers to form, although it and regions within the cytoplasmic domain may provide stability to the structure (Earl and Moss, 1993; Einfeld and Hunter, 1988). Formation of an Env oligomer appears to be required for efficient transport out of the ER.

Shortly after the stable oligomer is formed, it is transported from the ER to the Golgi complex where the bulk of the mannose-rich oligosaccharide side chains are processed to hybrid/terminally glycosylated forms. In a late Golgi compartment, most likely the trans-Golgi network, the precursor is cleaved to gp120 and gp41 by a cellular protease. Furin and PC7 appear to be the most likely candidates for this action (Decroly et al., 1997; Hallenberger et al., 1997). These enzymes, members of the subtilisin family of proteases, recognize and cleave at a highly conserved basic sequence (RXKR) that links the C-terminus of gp120 to the N-terminus of gp41. Cleavage is important since it frees the fusion peptide at the amino-terminus of gp41 from the C-terminus of gp120, which is critical for its

Figure 1. Organization of the HIV-1 env gene product and functional domains of gp41. The env open reading frame is depicted with vertical arrows showing sites for cleavage by the host-encoded signal peptidase and furin-like Golgi proteinase. Oligosaccharide addition sites that are utilized *in vivo* are shown by vertical lines and CHO. The position and size of critical functional domains as well as T- and B-cell epitopes are shown to scale. The position of the cysteine-bounded loop that is part of epitope Cluster I is shown as two upper case Cs.

future role in fusion. It is also probably accompanied by conformational changes in the oligomeric complex since uncleaved gp160 is inefficiently incorporated into virions (Bosch and Pawlita, 1990; Dubay et al., 1995). A post-cleavage conformational change is consistent with studies on influenza virus, where, in the crystal structure of the cleaved hemagglutinin, the C-terminus of HA1(SU) and the N-terminus of HA2(TM) are separated by 21 Å (Wilson et al., 1981). In the case of HIV Env, gp120 and gp41 are non-covalently associated following cleavage of the precursor. It can be concluded from an analysis of mutant glycoproteins that residues within conserved regions (C1 and C5) at the N-terminus and at the C-terminus of gp120 play a key role in this association (Helseth et al., 1991). In gp41, a region (residues 528–562) immediately carboxy-terminal to the fusion peptide is important for this association, although mutations outside this region can also be disruptive (Cao et al., 1993).

It is important to emphasize that the stable oligomeric structure, which forms in the ER, is folded from uncleaved polyprotein precursors. This clearly allows the folding of regions of both SU and TM into domains that are structurally constrained in a manner that would be impossible if they were assembled from the mature proteins, gp120 and gp41 (see below). If one considers the folding of this constrained structure as the setting of a spring-loaded trap, cleavage of the precursor might then be analogous to the removal of the safety catch.

Immune Responses to gp41

Although the bulk of gp41 might be expected to be sequestered from the humoral immune system within the Env oligomeric complex, several regions of this molecule are recognized by sera from individuals infected with HIV. The first of these regions, often referred to as Cluster I extends from around residue 588 to residue 615 of gp41. The immunodominant epitope is centered around a short disulfide-bonded loop (residues 607–613) that is located within the central region of the gp41 ectodomain (1987b; Gnann et al., 1987c). Almost all sera from infected individuals react with this epitope even though some sequence diversity has been found in viruses of the O-subtype (Eberle et al., 1997). Nevertheless, antibodies to this epitope are not neutralizing and do not block infection by HIV. On the contrary, monoclonal antibodies reactive with this region have been reported to enhance infection through a complement-dependent mechanism (Robinson et al., 1990a; Robinson et al., 1990b). The immunologically active form of the seven-residue peptide, CSGKLIC, contains an intramolecular disulfide bond and NMR studies showed that this domain maintains a preference for a folded conformation, including a type I reverse turn about the residues SGKL. The presence of the disulfide bond is integral to the formation of the structure of the loop in solution. Mutations (Cys to Ser) which eliminate the possibility of loop formation result in the failure of antibody binding to this peptide (Oldstone et al., 1991).

In addition to this primary immunodominant epitope, a second ectodomain region, located proximal to the membrane-spanning region of gp41, is also recognized by sera from infected individuals (Cluster II; Horal et al., 1991; Xu et al., 1991). While a majority of antibodies reactive with this region do not neutralize HIV infection, one human monoclonal antibody (2F5) has been identified that neutralizes a variety of laboratory strains and clinical isolates of HIV-1. The amino acid sequence ELDKWA (residues 671–676) has been defined as the epitope recognized by this antibody (Muster et al., 1993). Sequence analysis of a variety of primary isolates suggests that the major determinant of MAbs 2F5 binding corresponds to the amino-acid sequence LDKW, since naturally occurring and *in vitro* selected neutralization-resistant viruses contain changes in the D and K positions of the ELDKWA motif (Purtscher et al., 1996). While 2F5 recognizes a putatively linear determinant present on short peptides, it seems likely that the epitope reflects a specific conformation of the contributing amino acids, since attempts to elicit neutralizing antibodies to a synthetic peptide encompassing this region or to the epitope presented on the hepatitis B surface antigen have been unsuccessful (Eckhart et al., 1996). However, insertion of this epitope into the loop of antigenic site B of the influenza virus hemagglutinin did elicit ELDKWA-specific, neutralizing immunoglobulins in antisera of mice (Muster et al., 1994). The mechanism of neutralization is not understood, although Neurath and colleagues have reported multiple effects of binding the 2F5 monoclonal antibody to virions that are suggestive of post-binding conformational changes in Env (Neurath et al., 1995). It is also relevant that the 2F5 epitope overlaps the amino acid sequence of a peptide derived from gp41 that is a potent inhibitor of fusion (see below)

and a tryptophan-rich region that we have recently shown plays a critical role in fusion and infectivity (Salzwedel and Hunter, unpublished).

An enigma, as far as anti-gp41 neutralizing antibodies are concerned, has been the fact that monoclonal antibodies raised against a hydrophilic peptide corresponding to residues 743-760 (DR-PEGIEEGGERDRDRS) of the cytoplasmic domain could neutralize a variety of HIV-1 isolates. Both monoclonal antibodies and polyclonal antisera raised against this epitope have neutralizing potential (Dalglish et al., 1988; Evans et al., 1989). There is currently no evidence that this region of the cytoplasmic domain can be translocated to the outer surface of the cell and so how such antibodies can neutralize has remained a puzzle. Recent experiments, however, indicate that a monoclonal antibody to this epitope can bind with equal efficiency to infected and uninfected cells (Sattentau et al., 1995), suggesting that the epitope is also present on a cell surface protein. Since it is clear that the lipid envelope of HIV incorporates a large number of cell surface proteins (Arthur et al., 1992), it is possible that neutralization by antibodies to this epitope is mediated through one of these cell-derived molecules.

The TM protein also elicits activity from the cellular arm of the immune response. One of the CTL epitopes, identified through the use of synthetic peptides and fusion proteins carrying portions of the HIV-1 env gene, overlaps with the immunodominant B-cell epitope (amino acids 591-602 of the most-likely HIV env sequence - AVERYLKDQQLL). Only two positions within this epitope showed variation among North American HIV-1 isolates, and the substitutions were conservative in nature (Johnson et al., 1992). Interestingly, a Lys to Arg substitution at position 597 abolished recognition, probably by interfering with the peptide-MHC interactions (Dai et al., 1992; Hammond et al., 1991). In additional studies, CTL from four subjects recognized at least three distinct epitopes in this region of gp41 in the context of the MHC class I alleles A24, B8, B14, and B27 (Shankar et al., 1996). The other region of gp41 recognized by CTL is a stretch of 20 amino acids at the C-terminus of the cytoplasmic domain. CTL from seven subjects recognized epitopes within this region that were restricted by A30, B7, B8 and B35 class I alleles (Shankar et al., 1996).

Role of gp41 in membrane fusion

a) Conformational changes on receptor binding

Several enveloped viruses including influenza virus, vesicular stomatitis virus, and the alpha-viruses require the acidic environment of the endosome to activate the fusogenic potential of their envelope glycoprotein complexes. Thus after binding a receptor molecule on the cell surface, they are taken up into coated pits, undergo endocytosis and then release their replicative machinery into the cell following fusion with the endosomal membrane. In contrast the entry of HIV is not dependent on exposure to an acidic environment (McClure et al., 1988; Sinangil et al., 1988; Stein et al., 1987).

It is known, for the hemagglutinin of influenza virus and the glycoprotein complex of Semliki Forest virus, that the acid environment of the endosomal compartment induces a conformational change in the proteins that is necessary for activation of the fusion process (reviewed in White et al., 1983; White, 1992). In the case of HIV, and SIV analogous conformational changes have been observed following binding of receptor molecules (Sattentau and Moore, 1991; Sattentau et al., 1993), and recent experiments suggest that these receptor-induced conformation changes may parallel in molecular terms those induced by acid pH in the influenza virus HA.

Initially, experiments aimed at defining the mechanism by which a soluble form of the HIV receptor (sCD4) neutralized the infectivity of HIV demonstrated that treatment of laboratory strains of this virus with sCD4 could induce rapid shedding of gp120 and loss of the characteristic surface spikes seen in EMs of virions (Kirsh et al., 1990; Moore et al., 1990). Dissociation was temperature dependent and no significant dissociation of gp120 was observed between 4° C and 16° C (Hart et al., 1991; Moore and Klasse, 1992; Moore et al., 1991; Sattentau and Moore, 1991).

In the case of influenza virus HA, brief treatment of the virus with low pH, which activates fusion, results in the exposure of antigenic epitopes and protease cleavage sites that are normally sequestered within the HA molecule (Daniels et al., 1983; Skehel et al., 1982; White et al., 1983). Similarly,

dissociation of gp120 by CD4 is accompanied by exposure of epitopes on gp41 that are not reactive in the native molecule (Hart et al., 1991; Sattentau and Moore, 1991). Epitope exposure is not dependent upon loss of gp120 since similar results were obtained following CD4 binding at 4° C (Sattentau and Moore, 1991; Sattentau et al., 1995). A majority of studies have employed sCD4 treatment of HIV virions or infected cells, but Sattentau and Moore (1993) have demonstrated a similar exposure of gp41 epitopes at the interface of fusing cells. Thus membrane anchored CD4, like sCD4, can induce gp41 epitope exposure following gp120 binding.

The exposure of epitopes in both TM and SU, coupled with the dissociation of gp120 that has been observed with laboratory adapted isolates of HIV-1, appears to represent the most dramatic effect of sCD4 on Env conformation and may reflect the relative instability of the glycoprotein complexes of these isolates. The susceptibility of primary isolates of HIV-1 to sCD4 neutralization is significantly lower than that of cell-line adapted isolates (Daar et al., 1990) and sCD4 is very inefficient at inducing gp120 dissociation in the non-adapted strains (Moore et al., 1992), perhaps reflecting the lower affinity for CD4 exhibited by the oligomeric form of Env from natural isolates (Moore et al., 1992).

The glycoproteins of HIV-2 and SIV also have a lower affinity for CD4 and these viruses are similarly resistant to neutralization by sCD4. In contrast to the laboratory adapted strains of HIV-1, both SIV and HIV-2 infection and syncytium formation can be enhanced by sub-inhibitory concentrations of the soluble receptor. Incubation of sCD4 with SIVagm viruses not only accelerates the kinetics of infection but also leads to a 100-fold increase in the titer of the virus (Allan et al., 1990; Clapham et al., 1991; Sekigawa et al., 1990; Werner et al., 1990). A similar enhancement of infectivity has recently been observed with HIV-1 when poorly neutralizable primary isolates are treated with sub-inhibitory concentrations of sCD4 (Sullivan et al., 1995). Treatment of SIV or HIV-2 infected cells with the soluble receptor does not result in dissociation of gp120 or increased exposure of TM epitopes - apparently as a consequence of stronger interactions between the subunits in these viruses (Allan et al., 1992; Sattentau et al., 1993). Nevertheless, it does result in some conformational changes, since an increase was observed in the exposure of the V2 and V3 loops, as detected by antibody binding (Sattentau et al., 1993), and antibodies from SIVagm infected monkeys, which failed to neutralize virus in vitro, could neutralize the sCD4-induced enhancement of SIVagm infection (Allan et al., 1990). This is consistent with the suggestion that sCD4 enhancement of SIVagm (and presumably of HIV-1 primary isolates and HIV-2) is a form of receptor mediated activation (Allan et al., 1990), analogous to the low-pH activation observed with influenza virus or Semliki Forest virus (Marsh and Helenius, 1989).

b) Molecular rearrangements following receptor binding - formation of coiled-coils

Both crystallographic and peptide studies with influenza virus have shown that the low pH-mediated conformational changes that occur in HA involve the structural transition of a region within the HA2 (TM) protein. Thus residues C-terminal to the fusion peptide, that are constrained as an extended polypeptide chain within the inactive HA, form an elongated helical structure following low pH activation (Bullough et al., 1994; Carr and Kim, 1993). The newly formed helices of the HA trimer form a triple-helix, coiled-coil structure that is postulated to position the fusion peptide, located at the N-terminus of HA2, for insertion into the target cell membrane. A heptad repeat (leucine-zipper-like) region capable of forming a similar coiled-coil structure is found within the N-terminal 100 amino acids of the TM protein of a majority of retroviruses. Peptides corresponding to this domain form stable coiled-coil structures (Wild et al., 1992) and fusion of the heptad repeat to the C-terminus of a monomeric form of bacterial protein-A or maltose-binding protein results in assembly of oligomeric chimeric proteins (Bernstein et al., 1995; Shugars et al., 1996). Thus this region has the capability to both oligomerize and mediate oligomerization of protein molecules. The heptad repeat domain is highly conserved among HIV-1 isolates and was originally postulated to play a role in the assembly oligomerization of the glycoprotein complex that occurs in the ER (Deltwart and Mosialos, 1990; Gallaher et al., 1989). Mutational analyses of the region have argued against a direct role in oligomer assembly in the ER but showed that even minor amino acid substitutions had dramatic effects on the biological activity of the HIV-1 glycoprotein, by blocking cell-cell fusion and virus infectivity (Chen, 1994; Chen et al., 1993; Dubay et al., 1992). Peptides corresponding to this region are potent inhibitors of HIV-1 induced cell fusion as well as virus infection, and this property correlates directly with the ability of the peptides to

form stable coiled-coil structures; the same mutations which block the biological function of Env also reduce the melting temperature of the peptide's coiled-coil structure as well as abrogate their inhibitory potential (Wild et al., 1994b). We have postulated therefore that binding of CD4 to gp120 might induce equivalent conformational changes to that seen during low pH treatment of the influenza virus HA, with the resulting formation of an extended coiled-coil that can present the N-terminal fusion peptide to the target cell membrane (Figure 2A and 2B) (Bernstein et al., 1995; Wild et al., 1994b). Such a mechanism would explain why mutations which prevent cleavage of the Env precursor effectively block fusion, since in the absence of a free amino-terminus, the fusion peptide would be unable to undergo such a dramatic displacement.

c) Insertion of the fusion peptide

While the formation of a coiled-coil structure might present the fusion peptide to the target cell membrane, the mechanism of action of the fusion peptide itself remains obscure. How, for example, does the amino-terminal fusion peptide of gp41 interact with the membrane and what molecular interactions must occur following insertion? Mutational analyses of the fusion peptides of HIV-1, HIV-2 and STV, which show sequence homology to the fusion peptides of the paramyxoviruses (Gallaheer, 1987), have demonstrated the importance of hydrophobic residues within this region; substitution of hydrophobic with hydrophilic residues reduces syncytium formation and virus infectivity, whereas enhancement of hydrophobicity increases both (Bosch et al., 1989; Freed and Myers, 1992; Freed et al., 1990; Steffy et al., 1992). Recent NMR analyses of the gp41 fusion domain indicated that a peptide corresponding to this region inserts into lipid micelles primarily as a helix (59%), with substantial beta-structure (26.7%). Fluorescence experiments with a tryptophan analogue of the peptide pointed to deep penetration of the peptide into the apolar hydrocarbon core. The region C-terminal to an alanine-15/glycine-16 flexible hinge interacts in a substantial way with the micelle, suggesting that it lies on the surface of the bilayer (Chang et al., 1997). Despite this data, it is still not clear if, in order to function during virus entry, the fusion peptide interacts only with lipid or whether it also interacts with membrane-associated proteins in a manner analogous to the signal peptides of secreted and membrane-spanning proteins.

d) Structural analyses of gp41: implications for the fusion process

Although the three-dimensional crystal structure of the influenza virus HA was established over 15 years ago (Wilson et al., 1981), similar attempts to determine the structure of the intact Env complex of HIV have been unsuccessful. Using the known structure of HA2 and structure-prediction algorithms, Gallaheer (1989) proposed a secondary structure for the HIV TM that included a number of structural motifs common to the TMs of diverse retroviral isolates. While this model has been conceptually valuable, protein-structure prediction remains unreliable, particularly for membrane proteins, and, more importantly, it fails to take into account the potential for formation of intermediate, quasi-stable structures such as those observed in HA. Within the past two years, however, significant progress has been made in obtaining high-resolution structure information on the fusion-active form of the HIV TM.

Following expression of the ectodomain of gp41 in bacteria, the insoluble product could be converted to a soluble form through limited proteolysis. This generated two peptide fragments N51 and C43, derived from the N- and C-regions of the gp41 ectodomain respectively (Lu et al., 1995). The N51 peptide corresponds to the 4-3 heptad repeat region (residues 549-599) discussed previously, while the C43 peptide is derived from a C-terminal potentially helical region (residues 633-675). Biophysical studies showed that these two peptides associate to form a highly thermostable, helical, trimeric complex of heterotrimers in which the N51 and C34 helices are arranged in an anti-parallel fashion (Figure 3A; Lu et al., 1995). An almost identical, antiparallel, coiled-coil bundle could be isolated using the same approach with the STV TM ectodomain (Blacklow et al., 1995). This provided the initial indication that the N- and C-terminal regions of TM can fold back onto themselves to form a stable, predominantly helical, complex. While these helical bundles lacked the interconnecting 33 amino acid region containing the immunodominant epitope (residues 600-632), expression of the intact gp41 ectodomain (minus the fusion peptide) in insect cells yielded soluble oligomeric (trimer) proteins that appear in the electron microscope to form similar rod-like structures and which by circular dichroism have high alpha-helical content (~80%) (Weissenhorn et al., 1996). Thus the presence of the

Figure 2. Ribbon diagram of the neutral and low pH forms of the influenza virus HA. A. Neutral pH form of the HA monomer crystal structure - HA1 is shown in grey, HA2 in colors that correspond to the amino terminal helix, the central extended chain and the carboxy-terminal helix (Wilson et al., 1981). B. Crystal structure of the low pH form of HA2 showing the formation of an extended helix following a major conformational change (Bullough et al., 1994). Note that the region of extended polypeptide chain between the N-terminal (yellow) helix and the C-terminal (red) helix is incorporated into the extended helix following acidification. The fusion peptide can thus be moved approximately 100Å.

Figure 3. Coiled-coil bundle formed by the N-terminal and C-terminal helical domains of gp41. Panel A shows proposed packing of helices in protease-resistant core from bacterially expressed gp41. The 33 amino-acid connecting loop with cysteine-bonded loop that is missing from the structure determined by Lu et al., (1996) is included schematically in this figure. Panel B shows a ribbon diagram of the coiled coil bundle formed by N36(555-590) and C34(637-670) peptides from gp41. Panel C shows a side-view of the space-filling model of the same structure (from the coordinates of Chan et al., 1997). White spheres are water molecules.

Figure 4. Hypothetical two-stage model for the conformational changes resulting in membrane proximity during fusion. A. Model for trimeric HIV Env complex. One SU molecule is cut away to show N-terminal heptad repeat constrained in an unstructured form. B. Following binding of the receptor complex, the extended-chain, heptad repeat region (shown in cyan) can assemble into a trimeric coiled-coil, thereby allowing the fusion peptide to insert into the target cell membrane. C. A second (rate limiting?) step in the conformational change is the formation of the coiled-coil bundle which results in apposition of the viral and target cell membranes. Peptide inhibitors might act by competing for the C-terminal heptad repeat domain during this latter step.

cysteine-loop region does not appear to interfere with the anti-parallel packing of the N- and C-terminal helices.

In the past year, two groups have succeeded in obtaining high resolution crystallographic data on the helical bundle formed by the N- and C-terminal domains of gp41. Chan et al. (1997) determined the structure of a slightly smaller complex, N36 (residues 555–590)/C34 (residues 637–670), to 2.0 Å resolution. This complex is a six-stranded helical bundle (fig 3B). The center of the bundle consists of a trimeric coiled-coil of three N36 helices arranged in parallel in a left-handed super helix. Three C34 helices are wrapped, antiparallel to the N36 helices, in a left-handed direction around the central coiled-coil trimer (Fig 3C). Overall the complex comprises a cylinder that is 35 Å in diameter and 55 Å in height. Using a somewhat different approach, Weissenhorn et al. (1997) determined the structure of the protease-resistant part of the gp41 ectodomain solubilized with a trimeric GCN4 coiled-coil in place of the amino-terminal fusion peptide. This structure, in which gp41 residues 550–597 form the triple-stranded alpha-helical coiled-coil core and residues 637–674 form the alpha-helices packed in the reverse direction against the outside of the coiled coil, is essentially identical to that of Chan et al. (1997). The C-terminal helices interact with the central coiled-coil core mainly through hydrophobic residues that are located in three grooves on its surface and the residues lining these grooves are highly conserved between HIV and SIV. It is important to note that peptides corresponding to the C-terminal helices do not have helical character in solution (Wild et al., 1995; Wild et al., 1994a) and so the formation and stability of these helices must depend to a large extent on their interaction with the triple-stranded central core. Indeed, it is quite possible that within the native Env complex on the virion, both the N-terminal heptad repeat region and the C-terminal domain are constrained as extended polypeptide chains, and that their transition to helical structure occurs only after receptor binding.

The trimeric coiled-coil complex has implications for the valency of the Env complex itself, for the structural rearrangements of Env involved in membrane fusion and for the development of novel antiretroviral drugs. The HIV Env complex has been described as a tetramer and as a dimer of dimers (Earl et al., 1990; Pinter et al., 1989; Schavaller et al., 1989) as well as a trimer (Gelderblom et al., 1987; Weiss et al., 1990). The trimeric nature of the gp41 coiled-coil complex, even though it is formed in the absence of gp120, clearly suggests that the native Env complex is also trimeric. This would be consistent with other retroviral Env complexes (Einfeld and Hunter, 1988; Einfeld and Hunter, 1990; Kamps et al., 1991) and with the influenza virus HA (Wilson et al., 1981).

The remarkable structural similarity of the HIV gp41 coiled-coil bundle to the fusion pH-induced conformation of influenza virus HA2 (Bullough et al., 1994) and to the transmembrane subunit of Moloney murine leukemia virus (Fass et al., 1997) led both Chan et al. (1997) and Weissenhorn et al. (1997) to propose that this structure represents the fusion-active form of the gp41 ectodomain. A number of factors support this conclusion. First the coiled-coil bundle is very thermostable, in contrast to the native Env complex, indicative of a terminal low energy conformation. Second, mutations that block fusion and infectivity map to residues that stabilize the coiled-coil bundle. For example the isoleucine at position 582, which we have shown is critical for fusion and for stability of the core triple helix (Wild et al., 1994b), makes a homotrimeric contact at the center of the coiled coil but also interacts with Trp640 in the outer helix. Similarly, Asn665 is in an a position (relative to the a-g positions of a heptad repeat) on the outer helix and packs into the groove on the central core. Mutation of this residue would be predicted to disrupt an extensive hydrogen bonding network that stabilizes the outer helix and, as might be expected if this were the case, abrogates fusion (Cao et al., 1993). Finally peptides corresponding to the N- and C-terminal helical domains that make up the coiled-coil bundle have potent anti-viral activity and block HIV fusion (Jiang et al., 1993; Lu et al., 1995; 1994b; Wild et al., 1993; 1994a). Given the stability of the coiled-coil structure, these peptides would not be expected to exert their antiviral effect once it has formed. In contrast, the structure supports the concept that the inhibitory peptides function by interacting with gp41 during the conformational changes to the fusion-active state. Thus the N-terminal peptides might block the formation of the internal coiled-coil or interact (as oligomeric peptides) with the C-terminal region, preventing its interaction with the N-terminal domain. C-terminal peptides might act competitively to prevent binding of this region of gp41 into the grooves of the central core. Indeed, a peptide comprising residues 647–682 (DP178) can bind to a bacterially produced gp41 ectodomain that has been deleted in this region. Moreover, the formation of this "molecular clasp"

reconstituted an epitope recognized by a monoclonal antibody prepared against the intact ectodomain of gp41 (Chen et al., 1995).

Taken together with the high resolution structural information, the data from mutations and inhibitory peptides points to a model for Env-induced fusion which involves two interdependent conformational changes in gp41 (Fig 4). In the first stage, following binding of the receptor complex, the constrained N-terminal heptad-repeat region of gp41 would be freed to assemble into the central coiled-coil, moving the fusion peptide in the process to a position where it could interact with the target cell membrane. The second stage would then be the formation of the coiled-coil bundle. Since this latter structure is arranged such that the N-terminal fusion peptide and the C-terminal membrane-spanning domain of gp41 are at the same end of the bundle, its formation would force the target cell membrane and the viral membrane in close proximity. This mechanical process thus might facilitate the initiation of the process where lipid mixing occurs. Neither the source of energy for the second conformational change and membrane apposition, which would be expected to have a high DG, nor the time frame over which such a process might occur is obvious. However, the fact that peptides such as DP178 can inhibit the fusion process at very low concentrations (pM), suggests that the formation of the coiled-coil bundle might be a rate limiting step in fusion.

e) gp41/fusion as an anti-viral target

It is clear from the work of several groups that peptides corresponding to the N-terminal heptad repeat (DP107) and the C-terminal helical domain (DP178, C34) represent potent inhibitors of HIV-1 entry and replication (Jiang et al., 1993; Lu et al., 1995; 1994b; Wild et al., 1993; 1994a). The possible mechanism of action of these peptides is discussed above and suggests that the conformational changes in gp41 are a novel target for anti-HIV therapeutics. A particularly important development of this concept, therefore, has been the initial clinical trial of the C-terminal (DP178) peptide in HIV-1 infected patients. In this study, 5 of 5 patients, receiving 100mg of the peptide twice daily for a period of 17 days, showed greater than 100-fold reductions in viral load (Saag et al., 1997) - an anti-viral effect greater than several RT inhibitors. Thus it appears that the fusion/entry process is a viable *in vivo* therapeutic target. The presence of the highly conserved, deep cavities on the N-terminal coiled-coil trimer with which conserved C-terminal residues interact, raises the possibility that novel small molecule or peptidomimetic inhibitors of HIV infection might be developed (Chan et al., 1997; Weissenhorn et al., 1997).

f) Requirement for a membrane anchor

The membrane-spanning anchor region of the TM protein is an important structural domain, as might be expected if the initial result of fusion peptide insertion is an oligomeric structure that acts as a bridge between the viral and cell membranes. The region of gp41 that spans the membrane to yield a fusion-active form of the HIV-1 Env remains ill-defined, but C-terminal truncations indicate residues 692-716 can function as a membrane anchor (Dubay et al., 1992; Helseth et al., 1990; Johnston et al., 1993). Point mutations and small deletion mutations within this region can abrogate fusion without affecting the membrane-anchor function (Helseth et al., 1990; Owens et al., 1994). Nevertheless, because the membrane-spanning and cytoplasmic domains of the HIV-1 glycoprotein can be replaced by those of CD4 or CD28 without loss of fusogenicity (Vincent et al., 1993; Wilk et al., 1996), there appears to be no requirement for a specific amino acid sequence within the transmembrane domain for Env to mediate the fusion reaction. In contrast, substitution of a glycosylphosphatidylinositol (GPI) anchor for the HIV-1 protein anchor abrogates fusion and infectivity, even though this GPI-linked protein is efficiently transported to the plasma membrane and incorporated into virions (Salzwedel et al., 1993; Weiss and White, 1993). A similar result has been reported for a GPI-linked form of the influenza virus HA protein, although in this case lipid mixing of the outer leaflet of the bilayer (hemi-fusion) was seen in the absence of complete fusion and mixing of cell contents (Kemble et al., 1994). Thus a protein component that spans both leaflets of the lipid bilayer appears to be critical for complete mixing of the viral and target membranes. Weissenhorn et al. (1997) have suggested that the rod-shaped gp41 coiled-coil bundles might cluster at their hydrophobic tips (which includes the membrane-spanning domain), thereby forming asters with centers that could be the points of initial membrane fusion.

Role of the HIV/SIV Cytoplasmic Domain

The TM proteins of lentiviruses have unusually long cytoplasmic domains. In HIV and SIV they exceed 150 amino acids and in the case of Visna virus it is more than 200 amino acids in length. This contrasts with most retroviral TMs where the cytoplasmic domain ranges from 20–40 amino acids (Hunter and Swanstrom, 1990). In an attempt to define the biological role of the additional residues in the cytoplasmic domain, several investigators have introduced premature termination codons in both the HIV-1 and SIV Env coding sequence to progressively truncate the cytoplasmic domain. The results of these experiments indicate that the cytoplasmic domain is dispensable for HIV or SIV Env-mediated cell-cell fusion. In the case of HIV-1, however, truncation of the cytoplasmic domain by as few as 20 amino acids results in significantly decreased virus replication in most cell types, although the virus is still viable in the highly permissive MT4 cell line (Dubay et al., 1992; Freed and Martin, 1995; Gabuzda et al., 1992; Wilk et al., 1992). SIV replication, in contrast, appears to have no requirement for the cytoplasmic domain of Env (Johnston et al., 1993; Ziegler and Littman, 1993). Indeed growth of SIV in human cells results in the selection for a viral Env with a spontaneous truncation which broadens the host range of the virus, probably by increasing the number of Env molecules incorporated into virions (Johnston et al., 1993) (Ziegler and Littman, 1993). Nevertheless, reversion back to wild-type virus has been observed following inoculation of virus encoding the Env truncation into macaques, suggesting that *in vivo* this region plays an important role (P. Luciw, personal communication; Kodama et al., 1989). A number of structural elements within the cytoplasmic domain may contribute to this *in vivo* function.

Endocytosis/Basolateral Targeting Signals

Cytoplasmic amino-acid motifs can function to direct endocytosis of a protein from the plasma membrane (reviewed in Trowbridge et al., 1993). Over the past few years, the sequences of these internalization/targeting signals have been identified for several constitutively recycling cell surface receptors and have been shown to involve tyrosine residues within a tight turn of the peptide backbone (Robinson et al., 1996; Trowbridge et al., 1993). For the transferrin receptor (TR) this sequence is Tyr-X-Arg-Phe and a general consensus for this type of internalization motif is Tyr-X-Arg/Lys-Y, where Y is usually a hydrophobic amino acid (Collawn et al., 1990). We noted previously (Hunter, 1994) that a majority of retroviral cytoplasmic domains contained a similar motif. In HIV and SIV, a Tyr-X-Pro-Leu/Val sequence, located membrane-proximally (residues 721–724) within the cytoplasmic domain, is highly conserved. Hoxie and colleagues have described a spontaneous mutant of an SIV Env protein in which the tyrosine in this motif was substituted by a cysteine. This mutant was poorly endocytosed, and was expressed on the cell surface and incorporated into virions at high levels (LaBranche et al., 1995; Sauter et al., 1996). Similarly, Siliciano and colleagues have shown that this tyrosine containing motif is important for efficient endocytosis of the HIV-1 Env molecule (Rowell et al., 1995). Additionally, Sauter et al. (1996) were able to demonstrate directly the role of the SIV motif in endocytosis by substituting the cytoplasmic domain of Env for that of CD4.

What is the rationale for conserving signals that result in the internalization and degradation of a protein that must be incorporated into virions for infectivity? It has been proposed that this process might reduce immunological recognition of infected cells; alternatively, the motifs might play a role in correct intracellular targeting. In epithelial cells, where the apical and basolateral membranes have distinct protein populations, retroviral glycoproteins are specifically targeted to the basolateral surface (Roth et al., 1983). When HIV-1 Gag and Env are co-expressed in epithelial cells, it is at this surface that virus assembly and release occurs (Lodge et al., 1994; Owens et al., 1991). Mutation of the same tyrosine-containing motif that directs endocytosis abrogates this basolateral targeting of HIV-1 Env and basolateral assembly of virions (Lodge et al., 1997).

Interestingly, the rate of internalization of Env protein from the surfaces of HIV-1-infected cells is slow, and the presence of the Pr55gag precursor protein appears to be necessary and sufficient to inhibit Env internalization (Egan et al., 1996). Thus an interaction of the matrix domain of the Gag protein with Env during assembly may preclude the interaction of Env with host adaptin molecules that recruit plasma membrane molecules into clathrin-coated pits and thereby effect incorporation into virions.

Calmodulin-binding domains

The cytoplasmic domains of HIV-1, HIV-2 and SIV contain two regions that would be predicted to fold into amphipathic helical segments which closely resemble those found in calmodulin-activated enzymes. Synthetic peptides corresponding to these regions have been shown to bind to calmodulin (CaM) with high affinity in the presence of calcium (Miller et al., 1993; Srinivas et al., 1993). These amphipathic peptides inhibit CaM-regulated activation of bovine brain phosphodiesterase *in vitro*, and have also been shown to inhibit mitogen-induced lymphocyte activation, a property shared by CaM antagonists. These properties are not an artifact of peptides since purified HIV-1 gp160 as well as full-length gp41 binds to CaM and expression of HIV Env in cells results in CaM co-localization (Radding et al., 1996). Moreover, while CaM binding by the C-terminal amphipathic helix is very sensitive to amino acid changes, the natural variation found in this region of Env leaves CaM-binding intact (Tencza et al., 1997; Tencza et al., 1995). Since calmodulin is intimately involved in the metabolism and functioning of T-cells, these studies point to a novel mechanism of viral cytopathogenesis mediated by the interaction of the HIV TM protein with cellular CaM, that could in turn uncouple critical cellular signal transduction pathways (Miller et al., 1993; Srinivas et al., 1993).

Conclusions

Despite the wealth of new information summarized above, much remains to be learned about the role of gp41 in HIV/SIV replication and pathogenesis. The similarities between the HIV Env complex and the HA oligomer of influenza virus has allowed functional extrapolations to be drawn from the limited structural information obtained from bacterially expressed proteins. Nevertheless, even moderate resolution structural data on the native Env complex would provide information critical to understanding the extent of the conformational changes that have only been inferred to-date. High resolution three-dimensional structures of the native molecule would clearly assist in the development of small molecule inhibitors that could block the postulated conformational changes in gp41 following receptor binding and would thus interfere with the process of membrane fusion.

It seems likely that the long cytoplasmic domain of the lentiviruses plays a key role in viral pathogenesis. In SIV the cytoplasmic domain can be truncated without affecting virus replication *in vitro*, yet the conservation of endocytosis signals, CaM binding domains, as well as structural elements that interact with the assembling capsid precursors, suggests that this region might play multiple roles in both virus replication and pathogenesis *in vivo*. Experiments aimed at defining the structure of the cytoplasmic domain and the cell-encoded molecules with which it likely interacts could yield important information in this regard.

The HIV/SIV TM protein provides an excellent example of how multiple, complex functions can be packaged in a small molecule. Although much has been learned in the last few years about the function of gp41, even more remains to be uncovered. The next few years promise to yield exciting new information both on it and the molecular events involved in virus-cell membrane fusion.

References

- Allan, J. S., Strauss, J., and Buck, D. W. (1990). Enhancement of SIV infection with soluble receptor molecules. *Science* 247:1084-8.
- Allan, J. S., Whitehead, E. M., Strout, K., Short, M., Kanda, P., Hart, T. K., and Bugelski, P. J. (1992). Strong association of simian immunodeficiency virus (SIVagm) envelope glycoprotein heterodimers: possible role in receptor-mediated activation. *AIDS Res. Hum. Retroviruses* 8:2011-20.
- Arthur, L., Bess, J., Jr, Sowder, R., II, Beaveniste, R., Mann, D., Chermann, J., and Henderson, L. (1992). Cellular proteins bound to immunodeficiency viruses: Implications for pathogenesis and vaccines. *Science* 258:1935-1938.
- Berger, E. A. (1997). HIV entry and tropism: the chemokine receptor connection. *AIDS* 11 (suppl A), S3-S16.

- Bernstein, H. B., Tucker, S. P., Kar, S. R., McPherson, S. A., McPherson, D. T., Dubay, J. W., Lebowitz, J., Compans, R. W., and Hunter, E. (1995). Oligomerization of the hydrophobic heptad repeat of gp41. *J. Virol.* 69:2745-2750.
- Blacklow, S. C., Lu, M., and Kim, P. S. (1995). A trimeric subdomain of the simian immunodeficiency virus envelope glycoprotein. *Biochemistry* 34:14955-62.
- Bosch, M. L., Earl, P. L., Fargnoli, K., Picciafuoco, S., Giombini, F., Wong-Staal, F., and Franchini, G. (1989). Identification of the fusion peptide of primate immunodeficiency viruses. *Science* 244:694-7.
- Bosch, V., and Pawlita, M. (1990). Mutational analysis of the human immunodeficiency virus type 1 env gene product proteolytic cleavage site. *J. Virol.* 64:2337-44.
- Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature (London)* 371:37-43.
- Cao, J., Bergeron, L., Helseth, E., Thali, M., Repke, H., and Sodroski, J. (1993). Effects of amino acid changes in the extracellular domain of the human immunodeficiency virus type 1 gp41 envelope glycoprotein. *J. Virol.* 67:2747-55.
- Carr, C. M., and Kim, P. S. (1993). A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* 73:823-832.
- Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 89:263-73.
- Chang, D. K., Cheng, S. F., and Chien, W. J. (1997). The amino-terminal fusion domain peptide of human immunodeficiency virus type 1 gp41 inserts into the sodium dodecyl sulfate micelle primarily as a helix with a conserved glycine at the micelle-water interface. *J. Virol.* 71:6593-602.
- Chen, C. H., Matthews, T. J., McDaniel, C. B., Bolognesi, D. P., and Greenberg, M. L. (1995). A molecular clasp in the human immunodeficiency virus (HIV) type 1 TM protein determines the anti-HIV activity of gp41 derivatives: implication for viral fusion. *J. Virol.* 69:3771-7.
- Chen, S. S. (1994). Functional role of the zipper motif region of human immunodeficiency virus type 1 transmembrane protein gp41. *J. Virol.* 68:2002-10.
- Chen, S. S., Lee, C. N., Lee, W. R., McIntosh, K., and Lee, T. H. (1993). Mutational analysis of the leucine zipper-like motif of the human immunodeficiency virus type 1 envelope transmembrane glycoprotein. *J. Virol.* 67:3615-9.
- Clapham, P. R., Blanc, D., and Weiss, R. A. (1991). Specific cell surface requirements for the infection of CD4-positive cells by human immunodeficiency virus types 1 and 2 and by Simian immunodeficiency virus. *Virology* 181:703-15.
- Collawn, J. F., Stangel, M., Kuhn, L. A., Esekogwu, V., Jing, S. Q., Trowbridge, I. S., and Tainer, J. A. (1990). Transferrin receptor internalization sequence YXRF implicates a tight turn as the structural recognition motif for endocytosis. *Cell* 63:1061-1072.
- Daar, E. S., Li, X. L., Moudgil, T., and Ho, D. D. (1990). High concentrations of recombinant soluble CD4 are required to neutralize primary human immunodeficiency virus type I isolates. *Proc. Natl. Acad. Sci. USA* 87:6574-8.
- Dai, L. C., West, K., Littau, R., Takahashi, K., and Ennis, F. A. (1992). Mutation of human immunodeficiency virus type 1 at amino acid 585 on gp41 results in loss of killing by CD8+ A24-restricted cytotoxic T lymphocytes. *J. Virol.* 66:3151-4.
- Dalglish, A. G., Chan, T. C., Kennedy, R. C., Kanda, P., Clapham, P. R., and Weiss, R. A. (1988). Neutralization of diverse HIV-1 strains by monoclonal antibodies raised against a gp41 synthetic peptide. *Virology* 163:209-15.
- Daniels, R. S., Douglas, A. R., Skehel, J. J., and Wiley, D. C. (1983). Analyses of the antigenicity of influenza haemagglutinin at the pH optimum for virus-mediated membrane fusion. *J. Gen. Virol.* 64:1657-1662.
- Decroly, E., Benjannet, S., Savaria, D., and Seidah, N. G. (1997). Comparative functional role of PC7 and furin in the processing of the HIV envelope glycoprotein gp160. *FEBS Lett* 405:68-72.

- Delwart, E. L., and Mosialos, G. (1990). Retroviral envelope glycoproteins contain a leucine zipper like repeat. *AIDS Res. Hum. Retroviruses* 6:703-706.
- Doms, R. W., Earl, P. L., Chakrabarti, S., and Moss, B. (1990). Human immunodeficiency virus types 1 and 2 and simian immunodeficiency virus env proteins possess a functionally conserved assembly domain. *J. Virol.* 64:3537-40.
- Doms, R. W., and Peiper, S. C. (1997). Unwelcomed guests with master keys: how HIV uses chemokine receptors for cellular entry. *Virology* 235:179-90.
- Dubay, J. W., Dubay, S. R., Shin, H. J., and Hunter, E. (1995). Analysis of the cleavage site of the human immunodeficiency virus type 1 glycoprotein: requirement of precursor cleavage for glycoprotein incorporation. *J. Virol.* 69:4675-82.
- Dubay, J. W., Roberts, S. J., Brody, B., and Hunter, E. (1992). Mutations in the leucine zipper of the human immunodeficiency virus type 1 transmembrane glycoprotein affect fusion and infectivity. *J. Virol.* 66:4748-56.
- Dubay, J. W., Roberts, S. J., Hahn, B. H., and Hunter, E. (1992). Truncation of the human immunodeficiency virus type 1 transmembrane glycoprotein cytoplasmic domain blocks virus infectivity. *J. Virol.* 66:6616-25.
- Earl, P. L., Doms, R. W., and Moss, B. (1990). Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc. Natl. Acad. Sci. USA* 87:648-652.
- Earl, P. L., and Moss, B. (1993). Mutational analysis of the assembly domain of the HIV-1 envelope glycoprotein. *AIDS Res. Hum. Retroviruses* 9:589-94.
- Eberle, J., Loussert-Ajaka, I., Brust, S., Zekeng, L., Hauser, P. H., Kaptue, L., Knapp, S., Diamond, F., Saragosti, S., Simon, F., and Gurtler, L. G. (1997). Diversity of the immunodominant epitope of gp41 of HIV-1 subtype O and its validity for antibody detection. *J. Virol. Methods* 67:85-91.
- Eckhart, L., Raffelsberger, W., Ferko, B., Klima, A., Purtscher, M., Katinger, H., and Ruker, F. (1996). Immunogenic presentation of a conserved gp41 epitope of human immunodeficiency virus type 1 on recombinant surface antigen of hepatitis B virus. *J. Gen. Virol.* 77:2001-8.
- Egan, M. A., Carruth, L. M., Rowell, J. F., Yu, X., and Siliciano, R. F. (1996). Human immunodeficiency virus type 1 envelope protein endocytosis mediated by a highly conserved intrinsic internalization signal in the cytoplasmic domain of gp41 is suppressed in the presence of the Pr55gag precursor protein. *J. Virol.* 70:6547-56.
- Einfeld, D., and Hunter, E. (1988). Oligomeric structure of a prototype retrovirus glycoprotein. *Proc. Natl. Acad. Sci. USA* 85:8688-8692.
- Einfeld, D., and Hunter, E. (1990). Oligomeric structure of retroviral envelope glycoproteins. In: *The use of X-ray crystallography in the design of antiviral agents*, W. G. Laver and G. M. Air, eds.: Academic Press, pp. 335-343.
- Einfeld, D. A., and Hunter, E. (1994). Expression of the TM protein of Rous sarcoma virus in the absence of SU shows that this domain is capable of oligomerization and intracellular transport. *J. Virol.* 68:2513-20.
- Einfeld, D. A., and Hunter, E. (1997). Mutational analysis of the oligomer assembly domain in the transmembrane subunit of the Rous sarcoma virus glycoprotein. *J. Virol.* 71:2383-2389.
- Evans, D. J., McKeating, J., Meredith, J. M., Burke, K. L., Katak, K., John, A., Ferguson, M., Minor, P. D., Weiss, R. A., and Almond, J. W. (1989). An engineered poliovirus chimera elicits broadly reactive HIV-1 neutralizing antibodies. *Nature* 339:385-388.
- Fass, D., Harrison, S. C., and Kim, P. S. (1997). Retrovirus envelope domain at 1.7 angstrom resolution. *Nature* 388:691-3.
- Freed, E. O., and Martin, M. A. (1995). Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. *J. Virol.* 69:1984-9.
- Freed, E. O., and Myers, D. J. (1992). Identification and characterization of fusion and processing domains of the human immunodeficiency virus type 2 envelope glycoprotein. *J. Virol.* 66:5472-8.

- Freed, E. O., Myers, D. J., and Risser, R. (1990). Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. *Proc. Natl. Acad. Sci. USA* 87:4650-4.
- Gabuzda, D. H., Lever, A., Terwilliger, E., and Sodroski, J. (1992). Effects of deletions in the cytoplasmic domain on biological functions of human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* 66:3306-15.
- Gallagher, W. R. (1987). Detection of a fusion peptide sequence in the transmembrane protein of the human immunodeficiency virus. *Cell* 50:327-328.
- Gallagher, W. R., Ball, J. M., Garry, R. F., Griffin, M. C., and Montelaro, R. C. (1989). A general model for the transmembrane proteins of HIV and other retroviruses. *AIDS Res. Hum. Retroviruses* 5:431-440.
- Gelderblom, H. R., Hausmann, E. H., Ozel, M., Pauli, G., and Koch, M. A. (1987). Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology* 156:171-6.
- Gnann, J. W., Jr., Nelson, J. A., and Oldstone, M. B., gp41 epitopes, LA (1987b). Fine mapping of an immunodominant domain in the transmembrane glycoprotein of human immunodeficiency virus. *J. Virol.* 61:2639-41.
- Gnann, J. W., Jr., Schwimmbeck, P. L., Nelson, J. A., Truax, A. B., and Oldstone, M. B., gp41 epitopes, LA (1987c). Diagnosis of AIDS by using a 12-amino acid peptide representing an immunodominant epitope of the human immunodeficiency virus. *J. Infect. Dis.* 156:261-7.
- Hallenberger, S., Moulard, M., Sordel, M., Klenk, H. D., and Garten, W. (1997). The role of eukaryotic subtilisin-like endoproteases for the activation of human immunodeficiency virus glycoproteins in natural host cells. *J. Virol.* 71:1036-45.
- Hammond, S. A., Obah, E., Stanhope, P., Monell, C. R., Strand, M., Robbins, F. M., Bias, W. B., Karr, R. W., Koenig, S., and Siliciano, R. F. (1991). Characterization of a conserved T cell epitope in HIV-1 gp41 recognized by vaccine-induced human cytolytic T cells. *J. Immunol.* 146:1470-7.
- Hart, T. K., Kirsh, R., Ellens, H., Sweet, R. W., Lambert, D. M., Petteway, S., Jr., Leary, J., and Bugelski, P. J. (1991). Binding of soluble CD4 proteins to human immunodeficiency virus type 1 and infected cells induces release of envelope glycoprotein gp120. *Proc. Natl. Acad. Sci. USA* 88:2189-93.
- Helseth, E., Olshevsky, U., Furman, C., and Sodroski, J. (1991). Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J. Virol.* 65:2119-23.
- Helseth, E., Olshevsky, U., Gabuzda, D., Ardman, B., Haseltine, W., and Sodroski, J. (1990). Changes in the transmembrane region of the human immunodeficiency virus type 1 gp41 envelope glycoprotein affect membrane fusion. *J. Virol.* 64:6314-8.
- Horal, P., Sveanerholm, B., Jeansson, S., Rymo, L., Hall, W. W., and Vahine, A. (1991). Continuous epitopes of the human immunodeficiency virus type 1 (HIV-1) transmembrane glycoprotein and reactivity of human sera to synthetic peptides representing various HIV-1 isolates. *J. Virol.* 65:2718-23.
- Hunter, E. (1994). Macromolecular interactions in the assembly of HIV and other retroviruses. *Seminars in Virology* 5:71-83.
- Hunter, E., and Swanstrom, R. (1990). Retroviral envelope glycoproteins. In *Retroviruses - Strategies for Replication*, R. Swanstrom and P. K. Vogt, eds. (Berlin: Springer-Verlag), pp. 187-253.
- Jiang, S., Lin, K., Strick, N., and Neurath, A. R. (1993). Inhibition of HIV-1 infection by a fusion domain binding peptide from the HIV-1 envelope glycoprotein GP41. *Biochem Biophys Res Commun* 195:533-8.
- Johnson, R. P., Trocha, A., Buchanan, T. M., and Walker, B. D. (1992). Identification of overlapping HLA class I-restricted cytotoxic T cell epitopes in a conserved region of the human immunodeficiency virus type 1 envelope glycoprotein: definition of minimum epitopes and analysis of the effects of sequence variation. *J. Exp. Med.* 175:961-71.
- Johnston, P. B., Dubay, J. W., and Hunter, E. (1993). Truncations of the simian immunodeficiency virus transmembrane protein confer expanded virus host range by removing a block to virus entry into cells. *J. Virol.* 67:3077-86.

- Kamps, C. A., Lin, Y. C., and Wong, P. K. Y. (1991). Oligomerization and transport of the envelope protein of Moloney murine leukemia virus-TB and of ts-1, a neurovirulent temperature-sensitive mutant of Mo-MuLV-TB. *Virology* 184:687-694.
- Kemble, G. W., Danieli, T., and White, J. M. (1994). Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. *Cell* 76:383-91.
- Kirsh, R., Hart, T. K., Ellens, H., Miller, J., Petteway, S. I., Lambert, D. M., Leary, J., and Bugelski, P. J. (1990). Morphometric analysis of recombinant soluble CD4-mediated release of the envelope glycoprotein gp120 from HIV-1. *AIDS Res. Hum. Retroviruses* 6:1209-12.
- Kodama, T., Wooley, D. P., Naidu, Y. M., Kestler, H. W. d., Daniel, M. D., Li, Y., and Desrosiers, R. C. (1989). Significance of premature stop codons in env of simian immunodeficiency virus. *J. Virol.* 63:4709-14.
- LaBranche, C. C., Sauter, M. M., Haggarty, B. S., Vance, P. J., Romano, J., Hart, T. K., Bugelski, P. J., Marsh, M., and Hoxie, J. A. (1995). A single amino acid change in the cytoplasmic domain of the simian immunodeficiency virus transmembrane molecule increases envelope glycoprotein expression on infected cells. *J. Virol.* 69:5217-27.
- Lodge, R., Göttinger, H., Gabuzda, D., Cohen, E. A., and Lemay, G. (1994). The intracytoplasmic domain of gp41 mediates polarized budding of human immunodeficiency virus type 1 in MDCK cells. *J. Virol.* 68:4857-4861.
- Lodge, R., Lalonde, J. P., Lemay, G., and Cohen, E. A. (1997). The membrane-proximal intracytoplasmic tyrosine residue of HIV-1 envelope glycoprotein is critical for basolateral targeting of viral budding in MDCK cells. *EMBO J.* 16:695-705.
- Lu, M., Blacklow, S. C., and Kim, P. S. (1995). A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat Struct Biol* 2:1075-82.
- Marsh, M., and Helenius, A. (1989). Virus entry into animal cells. *Adv Virus Res* 36:107-51.
- McClure, M. O., Marsh, M., and Weiss, R. A. (1988). Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism. *EMBO J.* 7:513-8.
- Miller, M. A., Mietzner, T. A., Cloyd, M. W., Robey, W. G., and Montelaro, R. C. (1993). Identification of a calmodulin-binding and inhibitory peptide domain in the HIV-1 transmembrane glycoprotein. *AIDS Res. Hum. Retroviruses* 9:1057-66.
- Moore, J. P. (1997). Coreceptors: implications for HIV pathogenesis and therapy. *Science* 276:51-2.
- Moore, J. P., Jameson, B. A., Sattentau, Q. J., Willey, R., and Sodroski, J. (1993). Towards a structure of the HIV-1 envelope glycoprotein gp120: an immunochemical approach. *Philos Trans R Soc Lond Biol* 342:83-8.
- Moore, J. P., and Klasse, P. J. (1992). Thermodynamic and kinetic analysis of sCD4 binding to HIV-1 virions and of gp120 dissociation. *AIDS Res. Hum. Retroviruses* 8:443-50.
- Moore, J. P., McKeating, J. A., Huang, Y. X., Ashkenazi, A., and Ho, D. D. (1992). Virions of primary human immunodeficiency virus type 1 isolates resistant to soluble CD4 (sCD4) neutralization differ in sCD4 binding and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* 66:235-43.
- Moore, J. P., McKeating, J. A., Norton, W. A., and Sattentau, Q. J. (1991). Direct measurement of soluble CD4 binding to human immunodeficiency virus type 1 virions: gp120 dissociation and its implications for virus-cell binding and fusion reactions and their neutralization by soluble CD4. *J. Virol.* 65:1133-40.
- Moore, J. P., McKeating, J. A., Weiss, R. A., and Sattentau, Q. J. (1990). Dissociation of gp120 from HIV-1 virions induced by soluble CD4. *Science* 250:1139-42.
- Muster, T., Guinea, R., Trkola, A., Purtscher, M., Klima, A., Steindl, F., Palese, P., and Katinger, H. (1994). Cross-neutralizing activity against divergent human immunodeficiency virus type 1 isolates induced by the gp41 sequence ELDKWAS. *J. Virol.* 68:4031-4.
- Muster, T., Steindl, F., Purtscher, M., Trkola, A., Klima, A., Himmler, G., Ruker, F., and Katinger, H. (1993). A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J. Virol.* 67:6642-7.

- Neurath, A. R., Strick, N., Lin, K., and Jiang, S. (1995). Multifaceted consequences of anti-gp41 monoclonal antibody 2F5 binding to HIV type 1 virions. *AIDS Res. Hum. Retroviruses* 11:687-96.
- Oldstone, M. B., Tishon, A., Lewicki, H., Dyson, H. J., Feher, V. A., Assa-Munt, N., and Wright, P. E. (1991). Mapping the anatomy of the immunodominant domain of the human immunodeficiency virus gp41 transmembrane protein: peptide conformation analysis using monoclonal antibodies and proton nuclear magnetic resonance spectroscopy. *J. Virol.* 65:1727-34.
- Owens, R. J., Burke, C., and Rose, J. K. (1994). Mutations in the membrane-spanning domain of the human immunodeficiency virus envelope glycoprotein that affect fusion activity. *J. Virol.* 68:570-4.
- Owens, R. J., Dubay, J. W., Hunter, E., and Compans, R. W. (1991). Human immunodeficiency virus envelope protein determines the site of virus release in polarized epithelial cells. *Proc. Natl. Acad. Sci. USA* 88:3987-3991.
- Pinter, A., Honnen, W. J., Tilley, S. A., Bona, C., Zaghouni, H., Gorny, M. K., and Zolla-Pazner, S. (1989). Oligomeric structure of gp41, the transmembrane protein of human immunodeficiency virus type 1. *J. Virol.* 63:2674-2679.
- Purtscher, M., Trkola, A., Grassauer, A., Schulz, P. M., Klima, A., Dopfer, S., Gruber, G., Buchacher, A., Muster, T., and Katinger, H. (1996). Restricted antigenic variability of the epitope recognized by the neutralizing gp41 antibody 2F5. *AIDS* 10:587-93.
- Radding, W., Pan, Z. Q., Hunter, E., Johnston, P., Williams, J. P., and McDonald, J. M. (1996). Expression of HIV-1 envelope glycoprotein alters cellular calmodulin. *Biochem. Biophys. Res. Comm.* 218:192-197.
- Robinson, M. S., Watts, C., and Zerial, M. (1996). Membrane dynamics in endocytosis. *Cell* 84:13-21.
- Robinson, W., Jr., Kawamura, T., Gorny, M. K., Lake, D., Xu, J. Y., Matsumoto, Y., Sugano, T., Masuko, Y., Mitchell, W. M., Hersh, E., et al. (1990a). Human monoclonal antibodies to the human immunodeficiency virus type 1 (HIV-1) transmembrane glycoprotein gp41 enhance HIV-1 infection in vitro. *Proc. Natl. Acad. Sci. USA* 87:3185-9.
- Robinson, W., Jr., Kawamura, T., Lake, D., Masuko, Y., Mitchell, W. M., and Hersh, E. M. (1990b). Antibodies to the primary immunodominant domain of human immunodeficiency virus type 1 (HIV-1) glycoprotein gp41 enhance HIV-1 infection in vitro. *J. Virol.* 64:5301-5.
- Roth, M. G., Srinivas, R. V., and Compans, R. W. (1983). Basolateral maturation of retroviruses in polarized epithelial cells. *J. Virol.* 45:1065-73.
- Rowell, J. F., Stanhope, P. E., and Siliciano, R. F. (1995). Endocytosis of endogenously synthesized HIV-1 envelope protein. Mechanism and role in processing for association with class II MHC. *J. Immunol.* 155:473-88.
- Saag, M., Alldredge, L., Kilby, M., et al. (1997). A short term assessment of the safety, pharmacokinetics, and antiviral activity of T-20: An inhibitor of gp41 mediated membrane fusion. 35th Meeting of the IDSA, pp. Abstract #771.
- Salzwedel, K., Johnston, P. B., Roberts, S. J., Dubay, J. W., and Hunter, E. (1993). Expression and characterization of glycopospholipid-anchored human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* 67:5279-88.
- Sattentau, Q. J., and Moore, J. P. (1991). Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J. Exp. Med.* 174:407-15.
- Sattentau, Q. J., and Moore, J. P. (1993). The role of CD4 in HIV binding and entry. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 342:59-66.
- Sattentau, Q. J., Moore, J. P., Vignaux, F., Traincard, F., and Poignard, P. (1993). Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J. Virol.* 67:7383-93.
- Sattentau, Q. J., Zolla-Pazner, S., and Poignard, P. (1995). Epitope exposure on functional, oligomeric HIV-1 gp41 molecules. *Virology* 206:713-7.
- Sauter, M. M., Pelchen-Matthews, A., Bron, R., Marsh, M., LaBranche, C. C., Vance, P. J., Romano, J., Haggarty, B. S., Hart, T. K., Lee, W. M., and Hoxie, J. A. (1996). An internalization signal in the

- simian immunodeficiency virus transmembrane protein cytoplasmic domain modulates expression of envelope glycoproteins on the cell surface. *J Cell Biol* 132:795-811.
- Schawaller, M., Smith, G. E., Skehel, J. J., and Wiley, D. C. (1989). Studies with crosslinking reagents on the oligomeric structure of the env glycoprotein of HIV. *Virology* 172:367-369.
- Sekigawa, I., Chamow, S. M., Groopman, J. E., and Byrn, R. A. (1990). CD4 immunoadhesin, but not recombinant soluble CD4, blocks syncytium formation by human immunodeficiency virus type 2-infected lymphoid cells. *J. Virol.* 64:5194-8.
- Shankar, P., Fabry, J. A., Fong, D. M., and Lieberman, J. (1996). Three regions of HIV-1 gp160 contain clusters of immunodominant CTL epitopes. *Immunol. Lett.* 52:23-30.
- Shugars, D. C., Wild, C. T., Greenwell, T. K., and Matthews, T. J. (1996). Biophysical characterization of recombinant proteins expressing the leucine zipper-like domain of the human immunodeficiency virus type 1 transmembrane protein gp41. *J. Virol.* 70:2982-91.
- Sinangil, F., Loyter, A., and Volsky, D. J. (1988). Quantitative measurement of fusion between human immunodeficiency virus and cultured cells using membrane fluorescence dequenching. *Febs. Lett.* 239:88-92.
- Skehel, J. J., Bayley, P. M., Brown, E. B., Martin, S. R., Waterfield, M. D., White, J. M., Wilson, I. A., and Wiley, D. C. (1982). Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. *Proc. Natl. Acad. Sci. USA* 79:968-72.
- Srinivas, S. K., Srinivas, R. V., Azanbaramaiah, G. M., Compans, R. W., and Segrest, J. P. (1993). Cytosolic domain of the human immunodeficiency virus envelope glycoproteins binds to calmodulin and inhibits calmodulin-regulated proteins. *J Biol Chem* 268:22895-9.
- Steffy, K. R., Kraus, G., Looney, D. J., and Wong-Staal, F. (1992). Role of the fusogenic peptide sequence in syncytium induction and infectivity of human immunodeficiency virus type 2. *J. Virol.* 66:4532-5.
- Stein, B. S., Gowda, S. D., Lifson, J. D., Penhallow, R. C., Bensch, K. G., and Engleman, E. G. (1987). pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell* 49:659-68.
- Sullivan, N., Sun, Y., Li, J., Hofmann, W., and Sodroski, J. (1995). Replicative function and neutralization sensitivity of envelope glycoproteins from primary and T-cell line-passaged human immunodeficiency virus type 1 isolates. *J. Virol.* 69:4413-22.
- Tencza, S. B., Mietzner, T. A., and Montelaro, R. C. (1997). Calmodulin-binding function of LLP segments from the HIV type 1 transmembrane protein is conserved among natural sequence variants. *AIDS Res. Hum. Retroviruses* 13:263-9.
- Tencza, S. B., Miller, M. A., Islam, K., Mietzner, T. A., and Montelaro, R. C. (1995). Effect of amino acid substitutions on calmodulin binding and cytolytic properties of the LLP-1 peptide segment of human immunodeficiency virus type 1 transmembrane protein. *J. Virol.* 69:5199-202.
- Trowbridge, I. S., Collawn, J. F., and Hopkins, C. R. (1993). Signal-dependent membrane protein trafficking in the endocytic pathway. *Annu. Rev. Cell Biol.* 9:129-61.
- Vincent, M. J., Raja, N. U., and Jabbar, M. A. (1993). Human immunodeficiency virus type 1 Vpu protein induces degradation of chimeric envelope glycoproteins bearing the cytoplasmic and anchor domains of CD4: role of the cytoplasmic domain in Vpu-induced degradation in the endoplasmic reticulum. *J. Virol.* 67:5538-49.
- Weiss, C. D., Levy, J. A., and White, J. M. (1990). Oligomeric organization of gp120 on infectious human immunodeficiency virus type 1 particles. *J. Virol.* 64:5674-5677.
- Weiss, C. D., and White, J. M. (1993). Characterization of stable Chinese hamster ovary cells expressing wild-type, secreted, and glycosylphosphatidylinositol-anchored human immunodeficiency virus type 1 envelope glycoprotein. *J. Virol.* 67:7060-6.
- Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997). Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 387:426-30.

- Weissenhorn, W., Wharton, S. A., Calder, L. J., Earl, P. L., Moss, B., Aliprandis, E., Skehel, J. J., and Wiley, D. C. (1996). The ectodomain of HIV-1 env subunit gp41 forms a soluble, alpha-helical, rod-like oligomer in the absence of gp120 and the N-terminal fusion peptide. *EMBO J.* 15:1507-14.
- Werner, A., Winkowsky, G., and Kurth, R. (1990). Soluble CD4 enhances simian immunodeficiency virus STVagm infection. *J. Virol.* 64:6252-6.
- White, J., Kielian, M., and Helenius, A. (1983). Membrane fusion proteins of enveloped animal viruses. *Q. Rev. Biophys.* 16:151-95.
- White, J. M. (1992). Membrane fusion. *Science* 258:917-24.
- Wild, C., Dubay, J. W., Greenwell, T., Baird, T., Jr., Oas, T. G., McDanal, C., Hunter, E., and Matthews, T. (1994b). Propensity for a leucine zipper-like domain of human immunodeficiency virus type 1 gp41 to form oligomers correlates with a role in virus-induced fusion rather than assembly of the glycoprotein complex. *Proc. Natl. Acad. Sci. USA* 91:12676-80.
- Wild, C., Greenwell, T., and Matthews, T. (1993). A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell-cell fusion. *AIDS Res. Hum. Retroviruses* 9:1051-3.
- Wild, C., Greenwell, T., Shugars, D., Rimsky-Clarke, L., and Matthews, T. (1995). The inhibitory activity of an HIV type 1 peptide correlates with its ability to interact with a leucine zipper structure. *AIDS Res. Hum. Retroviruses* 11:323-5.
- Wild, C., Oas, T., McDanal, C., Bolognesi, D., and Matthews, T. (1992). A synthetic peptide inhibitor of human immunodeficiency virus replication: correlation between solution structure and viral inhibition. *Proc. Natl. Acad. Sci. USA* 89:10537-41.
- Wild, C. T., Shugars, D. C., Greenwell, T. K., McDanal, C. B., and Matthews, T. J. (1994a). Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc. Natl. Acad. Sci. USA* 91:9770-4.
- Wilk, T., Pfeiffer, T., and Bosch, V. (1992). Retained in vitro infectivity and cytopathogenicity of HIV-1 despite truncation of the C-terminal tail of the env gene product. *Virology* 189:167-77.
- Wilk, T., Pfeiffer, T., Bukovsky, A., Moldenhauer, G., and Bosch, V. (1996). Glycoprotein incorporation and HIV-1 infectivity despite exchange of the gp160 membrane-spanning domain. *Virology* 218:269-74.
- Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature (London)* 289:366-373.
- Xu, J. Y., Gorny, M. K., Palker, T., Karwowska, S., and Zolla-Pazner, S. (1991). Epitope mapping of two immunodominant domains of gp41, the transmembrane protein of human immunodeficiency virus type 1, using ten human monoclonal antibodies. *J. Virol.* 65:4832-8.
- Zingler, K., and Littman, D. R. (1993). Truncation of the cytoplasmic domain of the simian immunodeficiency virus envelope glycoprotein increases Env incorporation into particles and fusogenicity and infectivity. *J. Virol.* 66:2824-2831.

Global Variation in the HIV-1 V3 Region

Priya K. Dighe, Bette T. Korber and Brian T. Foley

MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545

Introduction

Due to the immunogenicity and functional importance of the V3 loop, there has been a great deal of interest in the V3 region of the envelope protein, resulting in a large international effort to obtain V3 region sequences. This section, which includes sequences taken from more than 2000 individuals and complete references, provides an overview of the variation of sequences that span this region.

Sequences

To best summarize the spectrum of international HIV-1 variants, only one representative viral sequence was included per infected individual. A complete set of references accompanies the sequence alignments, and nomenclature was preserved from the original papers so individuals and isolates can be clearly identified. HIV1 was deleted from the sequence names in this section, as all sequences included here are HIV1. Included with the references, when available, are brief descriptions of critical features of the sequences. This includes the health status of the individual from whom the virus was derived, whether or not the virus was cultured, and the year the blood sample was taken.

All sequences are prefaced by a subtype association (see phylogenetic clustering below) and a two letter country code to identify the country that the individual resided in at the time that the blood sample was taken. If the person was a recent immigrant and this information was available, we included the country of origin in the references. The two letter code was developed for Internet (Copyright 1992, Lawrence H. Landwater and the Internet Society), and incorporated here based on a suggestion made by Dr. Francine McCutchan. The key to the country codes follows this introduction. Note that this key was updated for 1997, with several country codes for Eastern European nations added since 1995.

Sometimes only one viral sequence was available from a person: a clone from an isolate, or a direct sequence of PCR amplified peripheral blood DNA. For other individuals, up to 80 viral sequences from PCR amplified DNA or RNA from blood samples were available. Consequently, over 11,000 sequences are represented by the 2087 included in this section. When two sequences were available from a person, one of the two was randomly selected. When three or more sequences were generated from a person, all available sequences were aligned (without regard to different time points of sampling) and either one representative sequence was chosen, or a consensus of the most common base found in each position in the alignment was generated. If there was a tie (e.g., 10 A's, 10 T's), the top base or amino acid in the alignment was used. If a set of sequences from two or more individuals was epidemiologically linked, and genetically very similar, only one sequence from the set was included, preferably the most recently infected. In the sequence description and references section, the short hand "PCR-direct, peripheral blood DNA" is used to signify that viral DNA was amplified from PBMCs, without culturing, and a single "direct" sequence was obtained from the amplification reaction products. The short hand "Consensus, PCR-clones, peripheral blood DNA" signifies that viral DNA was amplified from PBMCs and a set of clones was generated and sequenced from the PCR amplification products. The cloned sequences were aligned and a consensus was generated. In a handful of cases, a particular gp160 clone from an isolate was shown to be expressed and functional using a vaccinia virus T7 expression system. In these cases, the clone rather than the consensus of all sequences from a particular individual is included.

Phylogenetic clustering

Sequences have been organized according to the phylogenetic subtype association (A-J) of their envelope V3 regions only. The original sequence subtype (A-H) designations were defined based on the phylogenetic relationships determined by using both gag and env genes (when possible), are

approximately genetically equidistant in envelope, and have multiple members. The phylogenetic subtype designations and associations have generally been adopted by the HIV research community, and are now often presented with the publication of new sequences. We have either determined the subtype designations here, if not specified in the original manuscript, or else confirmed the subtype designations of the original manuscripts, and then used the subtypes to organize this section. Generally, confirmations were done by aligning a set HIV-1 V3 region sequences with longer env gene sequences (Part IIIC) that have clear subtype associations, and then using parsimony or neighbor joining trees to determine associations. Some of the shorter gene fragments from this region were given a subtype designation based on Hamming distances, using the similarity function of the MASE program (Faulkner DV, and Jurka J. *TIBS* 13:321-322 (1988)); these sequences have ".sh" appended to their name to indicate that they were too short for phylogenetic analysis. Parsimony trees were generated using PAUP (David Swofford, Illinois Natural History Survey), and neighbor-joining trees were generated with Kimura distances and a transition to transversion ratio of 1.3 using PHYLIP (Joseph Felsenstein, University of Washington). All available nucleotide sequence information was used for phylogenetic analysis; longer protein sequences were trimmed to be approximately the same length as the majority of the PCR fragments in this region, for the purposes of presentation. Some sequences were difficult to classify, and are included in the "U", or unclassified, section. In addition, recombination between HIV-1 strains occurs when an individual is infected with more than one strain. A meeting was held in Santa Fe, New Mexico in October, 1995 to discuss the implications of recombination and methods for detecting recombinant sequences. Because inter-subtype as well as intra-subtype recombination is known to occur, the subtype designations reported in this section should be interpreted only as pertaining to the V3 region of the envelope gene. For example HIV-1 MAL from Zaire, is known to be recombinant between subtypes A and D, with the V3 loop of env resembling subtype D. D_ZR-MAL is still listed with other subtype D sequences in this study, but may be moved to the U (uncertain) group in the future. In the V3 region MAL is strongly associated with the D subtype.

The set of sequences used to help resolve subtype associations included at least two sequences from each subtype (A-H), plus a simian immunodeficiency virus outgroup sequence. The sequences were selected based on being "typical" of the subtype they represent based on phylogenetic analysis. The set has changed as more sequences have accumulated. Thus not all subtype designations were based on the same reference set.

Limitations of phylogenetic analyses

Most of the PCR derived sequences contain a sub-optimal length for phylogenetic analyses, given the level of variability in this region – typically on the order of 250 to 300 nucleotides. Due to this limitation, some of the classifications in this section are uncertain and are our best estimate given the available information. Control studies were performed to compare the phylogenetic clustering of the V3 region using available longer sequences, however, and these studies indicate that our subtype designations based on the V3 region are generally reliable. For 146 sequences, we had an approximately 700 base region of env available representing all of the subtypes A-H. (The limitation in length was due to including the H subtype sequences, which did not cover all of gp120.) After removing positions in the alignment which included gaps, 519 bases were left. When a 298 base V3 region fragment was excised from this set, and neighbor joining trees were constructed using both the 519 base and 298 base long sequences, the phylogenetic subtype designations were consistent in each case. Further, when a subset of longer gp120 sequences was analyzed (92 of the 146), including 935 bases after removing positions in the alignment which included gaps, the subtype designations were again clear in neighbor-joining trees. This indicates that the limited V3 region PCR fragments, which include more than the V3-loop, are generally able to serve as a reliable basis for subtype determination, given the limitation that the V3 loop sequence may be embedded in a recombinant genome.

Without detailed analysis, genetic recombination between subtypes may obscure phylogenetic relationships between sequences. A characteristic of recombination is an indeterminate place in phylogenetic analyses, and some of the "Uncertain" category sequences may prove to be recombinant genomes upon further inspection. Also, while a subtype designation based on a gene or gene fragment may be correct, recombination events outside the region examined may have occurred. Therefore,

Introduction

care should be taken to not overinterpret the subtype designations. If one is to discuss the subtype designations of viral isolates based on the data presented here, they should refer to the designation as "B-like over the V3 loop region," rather than as "subtype B".

Limitations of V3 amino acid consensus sequences

The V3 amino acid consensus sequences generated for each subtype have interesting features; however, one should be wary about assuming that any of the consensus sequences may broadly represent their subtype. Certainly many V3 loop variants in each of the subtypes are extremely divergent from the consensus sequences. These divergent forms may have very different biological and immunological characteristics from viruses which are similar to the consensus. Additionally, because of the relatively small sample size of most of the subtypes, consensus sequences can be dominated by a small group of highly similar sequences, which may in turn be a sampling artifact. Hence, these consensus sequences are "evolving" as new sequences from each subtype become available.

	V3 LOOP									
ALL_CONSENSUS	IIIRSENITDKAKTIVQLNKEVKIM	CTAPNDRNR.SIML..GPGQ..AFYATGD..I.I.GDIRQ.AMC	NISSENNN.TLQVANKL..RE..RP....NR...T.IIF..FS.SGGGL.EITTHSF.NCOO.EFFYCVT							
A_CONSENSUS_97	VM-----N-----V-P-X-	-----VR-----	-----V-----TQ-----K-Y-----							
B_CONSENSUS_97	VV-----P-----	-----R-----T-E-----	-----L-AK-----K-IV-----Q-G-----V-NQ-----P-----VM-----							
C_CONSENSUS_97	-----L-N-V-----H-----V	-----R-----T-----	-----KOK-B-----R-G-----A-----P-----E-A-----R-----							
D_CONSENSUS_97	-----L-N-----I-----T-----	-----Y-Q-RT-----L-T-----	-----GA-K-----K-----GD-LL-----T-----K-----P-----							
E_CONSENSUS_97	-----L-N-----H-K-----	-----S-T-----T-----V-R-----D-----K-Y	E-HG-K-E.V-K-TE-----K-----H-----Q-P-----M-H-----R-----							
F_CONSENSUS_97	-----Q-S-T-----H-T-Q-----	-----T-----D-----K-----	V-G-Q-----R-RA-----ES-----P-----A-----K-NS-----M-----R-----							
G_CONSENSUS_97	-----L-----V-----K-I-----	-----TF-----	V-K-B-M-N-TAQ-----KK-I-----H-T-NS-A-----R-----							
H_CONSENSUS_97	V-----X-----T-N-----KSP-E-----	-----S-----	T-ED-RR-----HE-VQQ-----H-----Q-----B-----M-----M-T-----R-----							
I_CONSENSUS_97	VV-----X-----T-N-----AKA-X-----	-----G-----V-----TW-----B-----	GND-D-----KVISSE-----KA-L-P-----R-A-P-----							
J_CONSENSUS_97	V-----X-----KT-T-V-----	V-A-----G-----VL-----B-----	GR-----SR-VA-----Y-----T-----Q-K-AMP-----N-T-----							
Q_CONSENSUS_97	-R-NQK--S-SQ-N--T--STDMT	-E-G-Q-VQ-B-K--KA-NTSGLA.AGNNNS-A-Y	TINA-D-EK.A-NOTAE--Y--LELV--M--SNVT-TNF.NRESGGDAEVTHLFTVCHG-----							
Q_CONSENSUS_97	-----N-----Q-----	-----R-----R-----T-----	-----Q-W-----X-----P-----							

Subtype consensus sequences. This V3 region alignment shows a consensus sequence generated for each of the eleven subtypes. The subtype consensus sequences indicate the most common amino acid found in each position among the sequences associated with each subtype. The sequences are aligned to a consensus based on the most common amino acid in the subtype consensus sequences, which approximates a "global" consensus. It was generated in this way (rather than by using all 1651 sequences) to avoid over-representation of the B subtype, which has by far the largest number of available sequences. As is the convention in this compendium, a dash (-) indicates concurrence with the top sequence in the alignment; a period (.) indicates a deletion. The carets show where the N-linked glycosylation sites are found in the consensus. The V3 loop is set off from the surrounding sequence by a space on either side to facilitate viewing. Interesting features of the consensus alignment are: 1) Only in the B subtype is GPGR

the most common tip of the V3 loop; globally, GPGQ is more prevalent. 2) A highly conserved N-linked glycosylation site is constitutively absent in the C subtype, proximal to the first cysteine (C) in loop. 3) The D subtype consensus has 34 amino acids from cysteine (C) to cysteine (C) rather than the more common 35; at the point where the deletion occurs, it is not uncommon to find insertions of 2 to 4 amino acids, as can be observed in the sequence alignments. 4) The D subtype has two arginine (R) residues in the V3 loop that are uncharacteristic relative to the other consensus sequences; positively charged amino acids in these positions may result in a syncytia inducing, non-monocytropic phenotype (Fouchier RAM, et al., *J. Virol.* 66:3183-3187 (1992)). 5) A higher degree of variation is seen in the region just downstream of the V3 loop than within it. This difference is also observed internally among the sequences of the different subtypes. 6) The A, C, G and H consensus sequences have very similar V3 loop sequences.

V3 Loop Amino Acids

The following pages present amino acid alignments of the V3 loop, arranged by phylogenetic subtype. For each subtype, the number of sequences used to construct the alignment is indicated. The top line in each alignment represents the consensus sequence for that subtype, where consensus simply means the most common amino acid found in each position among the sequences of the given subtype. The subscripts record the frequency with which that amino acid is observed at that location among members of the subtype. An amino acid which is conserved 100% is shown with no subscript. Directly beneath the most common amino acid in each position are the other amino acids observed in that position, listed from most common to least common. An asterisk (*) subscript means less than 0.5% of the sequences had the indicated amino acid at that location. A dash (-) indicates a gap inserted to maintain the alignment. Percentages were rounded to the nearest whole number.

For this year's alignment, the HMMER (version 1.8) hidden Markov model software (Sean Eddy, Dept. of Genetics, Washington U. School of Medicine, St. Louis, MO 63110; eddy@genetics.wustl.edu) was used to objectively align all 2035 sequences. The frequency counts are derived from this alignment. Because each subtypes required different numbers and positions of gaps in order to create the full multiple sequence alignment, some sequences with unusual insertions were trimmed from the HMMER alignment, and a few positions were adjusted by hand, using MASE, prior to printing the full alignment which appears following the country codes description. The sequences which were culled from the alignment after counting frequencies, are appended.

Both the untouched HMMER alignment, and the edited version, will be available via ftp from the LANL HIV database (<http://hiv-web.lanl.gov>). Questions about these alignments should be directed to (btf@t10.lanl.gov) (505-665-1970).

A subtype (320 sequences)

N₁ C T₂ R₃ P₄ N₅ N₆ T₇ R₈ K₉ S₁₀ V₁₁ R₁₂ L₁₃ G₁₄ P₁₅ G₁₆ Q₁₇ A₁₈ F₁₉ Y₂₀ A₂₁ T₂₂ G₂₃ D₂₄ L₂₅ L₂₆ G₂₇ D₂₈ L₂₉ R₃₀ Q₃₁ A₃₂ H₃₃ C N₃₄
T₁ Y₂ L₃ T₄ H₅ G₆ D₇ T₈ K₉ S₁₀ R₁₁ G₁₂ L₁₃ H₁₄ L₁₅ A₁₆ L₁₇ R₁₈ T₁₉ I₂₀ F₂₁ T₂₂ R₂₃ S₂₄ E₂₅ V₂₆ T₂₇ E₂₈ N₂₉ T₃₀ K₃₁ K₃₂ S₃₃ Y₃₄
D₁ S₂ V₃ R₄ L₅ S₆ K₇ Y₈ I₉ K₁₀ T₁₁ R₁₂ M₁₃ P₁₄ M₁₅ K₁₆ S₁₇ L₁₈ H₁₉ S₂₀ D₂₁ A₂₂ M₂₃ V₂₄ K₂₅ V₂₆ T₂₇ R₂₈ T₂₉ F₃₀
I₁ E₂ K₃ F₄ M₅ A₆ Y₇ T₈ V₉ L₁₀ A₁₁ E₁₂ D₁₃ N₁₄ M₁₅ T₁₆ A₁₇ C₁₈ S₁₉ Y₂₀ K₂₁
X₁ H₂ X₃

B subtype (1078 sequences)

N₁ C T₂ R₃ P₄ N₅ N₆ T₇ R₈ K₉ S₁₀ L₁₁ H₁₂ L₁₃ G₁₄ P₁₅ G₁₆ R₁₇ A₁₈ F₁₉ Y₂₀ T₂₁ T₂₂ G₂₃ E₂₄ L₂₅ L₂₆ G₂₇ D₂₈ L₂₉ R₃₀ Q₃₁ A₃₂ H₃₃ C N₃₄
T₁ H₂ D₃ K₄ S₅ E₆ L₇ Y₈ V₉ T₁₀ T₁₁ S₁₂ T₁₃ Y₁₄ K₁₅ I₁₆ I₁₇ V₁₈ M₁₉ T₂₀ D₂₁ K₂₂ S₂₃ F₂₄ Q₂₅ P₂₆ W₂₇ Q₂₈ K₂₉ E₃₀ S₃₁ S₃₂ V₃₃ V₃₄ G₃₅ I₃₆ R₃₇ S₃₈ D₃₉ G₄₀ L₄₁ M₄₂ R₄₃ Q₄₄ K₄₅ T₄₆ H₄₇ A₄₈
D₁ K₂ S₃ E₄ L₅ Y₆ V₇ T₈ H₉ G₁₀ D₁₁ T₁₂ K₁₃ I₁₄ I₁₅ V₁₆ M₁₇ T₁₈ D₁₉ K₂₀ S₂₁ F₂₂ Q₂₃ P₂₄ W₂₅ Q₂₆ K₂₇ E₂₈ S₂₉ S₃₀ V₃₁ V₃₂ G₃₃ I₃₄ R₃₅ S₃₆ D₃₇ G₃₈ L₃₉ M₄₀ R₄₁ Q₄₂ K₄₃ T₄₄ H₄₅ A₄₆
I₁ E₂ K₃ F₄ M₅ A₆ Y₇ T₈ V₉ L₁₀ A₁₁ E₁₂ D₁₃ N₁₄ M₁₅ T₁₆ A₁₇ C₁₈ S₁₉ Y₂₀ K₂₁
X₁ H₂ X₃

HIV-1 V3 Region

F subtype (80 sequences)

G subtype (28 sequences)

N_3 C T_{36} R P_{36} N_{35} N N T R_{36} K_{36} S_{36} I_{36} T_{36} F_{36} G_{36} P_{36} G Q_{36} A_{36} F_{36} Y_{36} A_{36} T G_{36} D_{36} I_{36} I_{36} G D_{36} I R_{36} Q_{36} A H_{36} C N_{36}
 T_{36} I_7 T_4 S_5 I_4 R_4 R_4 K_4 R_{18} L_{22} A_7 T_7 L_4 H_7 V_7 L_7 G_4 S_4 S_{11} E_{21} V_4 T_4 N_{21} E_{21} K_4 L_4 Y_{21} K_4
 M_{11} V_7 I_{11} V_{11} R_4 P_7 K_7 S_7 N_7 D_7 T_4 N_{11} Q_7 G_7 S_7

Summary of variations in the tetrameric tip of the V3 loop. This table is a tally of the different tetramers observed in the 1651 individuals analyzed. This tip is thought to form a turn, and is the focal point of the potent neutralizing antibody epitopes that have been mapped to the V3 loop, as well as of T cell epitopes. Each column shows the number of occurrences of a given tetramer in either the entire 976 sequences (combined), or in subsets consisting of subtypes A–O, and the unclassified sequences (U). Underneath the column heading is the number of sequences in each category. The most common form found in each subtype is highlighted in bold lettering. In the B subtype, GPGR is the predominant form, however globally GPGQ is more common.

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	2035	320	1078	233	119	135	81	28	3	1	5	13	19
GPGQ	838	284	93	232	38	104	54	21	1	1	5		6
GPGR	800	12	724	1	16	19	20		1				7
GPGR	88	5	82										1
GWGR	36		36										
GPQS	28	1	27										
GPGG	25	2	22					1					
GLGR	25	4	17		1	1	1						1
GLGQ	22				21								1
APGR	18	1	17										
GSGQ	17	3			12		1						1
GQGQ	13		1		12								
GPQH	11	1	1			6	2	1					
GQGR	9		6		1	2							
GPMA	9											9	
GPGR	6		6										
GTGQ	6				6								
GRGQ	6	1			4				1				
GVGR	4		2		1		1						
GSGR	4	1	3										
GLGH	4	2				1		1					
APGQ	4	1						1	2				
GPRR	3		3										
GPGA	3		3										
EPGR	3		3										
APGS	3		3										
GTGR	3							2					1
RPGR	3		3										
GPKR	3		3										
GWGR	2		2										
GGGR	2		2										
GAGR	2		2										
AGGR	2		2										
GRGR	2		1										
GLGS	1		1										
GTGG	1		1										
GPWR	1											1	
GLRQ	1				1								
AQGR	1				1								
GPMS	1											1	
GPLR	1											1	
GARR	1		1										

V3 Loop Variation

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	2035	320	1078	233	119	135	81	28	3	1	5	13	19
QGGQ	1					1							
GPWG	1		1										
QIQQ	1				1								
QGGI	1					1							
GPLS	1											1	
GPQN	1		1										
GPGE	1		1										
GPEK	1		1										
GLGK	1		1										
AMGR	1		1										
APGJ	1		1										
AGGK	1		1										
APQE	1				1								
GPJL	1	1											
GPJL	1	1											
GPJL	1	1											
GPJP	1	1											
GMWQ	1				1								
IGGR	1				1								
QPCR	1		1										
TPGR	1		1										
*PGR	1						1						

Divergent patterns of V3 loop evolution in M-group subtypes.

Introduction

Individual isolates or clones of HIV-1 from within any one subtype exhibit differences in cytopathic effect and cell tropism. Although subtype B is the most thoroughly studied, syncytium-inducing (SI) and non-syncytium inducing (NSI) isolates have been described for other subtypes as well. Despite this intraclade diversity of phenotypes, to date there has been no definitive indication of intersubtype differences in pathogenicity or transmissibility. This may be due in part to a lack of controlled studies which could identify such differences. Subtype B virus infecting North Americans and Europeans cannot be directly compared to subtype A virus infecting central Africans, because of differences in factors such as diet, access to health care, and exposure to secondary pathogens. It is difficult to do longitudinal cohort studies in the developing world, where several subtypes are prevalent in the same population. The wide range of phenotypic variability within a single subtype, as well as ranges of human responses to infection, requires a large sample size for study. Finally, long-term studies including sero-incident cases are critical for detecting differences in pathogenicity, and this type of study requires many years to complete. This analysis and discussion is focussed on identifying differences in evolutionary trends of the envelope protein V3 region between the HIV-1 M group subtypes. It is hoped that this may help to identify subtypes of HIV-1 which are more likely to show differences in phenotype.

The evolution of HIV-1 is influenced not only by mutations in the viral genome and selective pressure by individual human hosts, but also by the overall temporal and geographical patterns of the epidemic. The rapid spread of subtype B into large serologically naive populations in the urban centers of North America and Europe resulted in a star phylogeny of subtype B, suggestive of large amounts of horizontal transfer of the virus in a relatively short period of time. Conversely, the evolution of subtype A has been influenced by less geographical spread, remaining largely in the sub-Saharan region of the African continent, within more rural populations, where dual infection with other subtypes of HIV-1 (and even HIV-2) has occurred. Thus the subtype A phylogenetic tree is more "bushy" and intersubtype recombinants containing subtype A are more common than those containing subtype B.

Differences in modes of transmission might lead to different evolutionary trends for HIV-1. Transmission via contaminated needles may put different selection pressures on the virus than sexual transmission for example. Thus, the analysis of HIV-1 sequence variability should be accompanied by a consideration of the natural history of the various lineages of HIV-1, as well as an understanding of the biased sampling of HIV-1 isolates submitted to sequence analysis.

A quick perusal of the tabulation of V3 loop amino acid sequences divided by subtype (pages 77-80) illustrates the variability of evolutionary patterns between subtypes. For example, while subtype C shows near perfect conservation of the GPGQ amino acid sequence at the tip of the V3 loop, subtype D has retained the proline in less than 50% of sequences sampled. An elevated rate of nonsynonymous substitutions in the third variable (V3) loop of subtype D viruses has previously been noted [Korber et al.(1994)]. Serological studies also indicated that the V3 loop of the D subtype was more diverse than other subtypes [Barin et al.(1996)]. The V3 loop is a functionally and immunogenically important domain of the viral envelope. Positively charged amino acids in certain positions in the V3 loop correlate with a syncytium-inducing phenotype in culture [de Jong et al.(1992), De Wolf et al.(1994)], and usage of the CXCR4 second receptor [Cocchi et al.(1996), Bieniasz et al.(1997)].

We examined differences in patterns of envelope V3 region variability using the V3 data set from the 1996 HIV sequence compendium [Blouin et al.(1996)], with many more sequences available than when we first attempted this study in 1994 [Korber et al.(1994)]. The data was explored using the following method: V3 region amino acid sequences were split into two regions: the V3 loop, typically including 35 amino acids between the cysteine residues which form a disulfide bond to produce the loop; and the 35 amino acids flanking the loop, including 20 amino acids from the region immediately adjacent to the NH2-terminal side of the loop (NFTDNARVIVQLNESVEIN for example) and 15 amino

V3 Loop Evolution

acids from the region immediately adjacent to the COOH-side of the loop (NLSSSTKWNNLTKQIT for example). Each of these two regions (loop or flanks) was aligned and scored for pairwise similarity using the PIMA software package [Smith & Waterman(1992)], with the protein similarity scoring scheme based on amino acid substitution matrices from protein blocks [Henikoff & Henikoff(1992), Henikoff & Henikoff(1993)]. The sequences in this set are compiled such that only a single sequence from each individual patient, or from each group of epidemiologically linked patients, is represented, thus eliminating the bias of multiple sequences from the same patient.

The results from these analyses are shown in Figures 1 and 2. The median of the similarities found in the V3 loop of subtype D is much lower than in the other subtypes, an indication of much greater diversity, while the interquartile range (a measure of the variation around the median) is larger. The diversity in the flanking regions, in contrast, is very similar to that in the other subtypes. This result suggests that the loop of subtype D is unique relative to the other subtypes, and is either under greater positive selective pressure for change, or alternatively, free from evolutionary constraints that restrict the variation of other subtypes. This effect is highly statistically significant, and is associated with the charge on the V3 loop (see below for discussion, and Figure 3). For example, using the non-parametric Wilcoxon or the Kolmogorov-Smirnov tests to compare the distributions of subtype A and D protein similarity scores for the V3 loop, gave p-values of < 0.00001 . (To do the statistical test, the average value of each sequence compared to all others of that subtype was used to represent that sequence, so each sequence was only counted once for the statistical comparison, although all pairwise comparisons were made to construct the figures.) The A versus D comparison is of particular interest as the samples tend to come from the same studies conducted in the same geographic regions. Another way to view this data is through comparing the cumulative distributions of the amino acid similarity scores for the different subtypes (Figure 4). In the V3 flanking region, the distributions of similarity scores for subtypes A, B, C, and D tend to be very similar. Subtype E sequences are highly conserved, paralleling the short phylogenetic branch lengths of these closely related sequences [Yu et al.(1995)], presumably reflecting the relative youth of the main (Southeast Asian) subtype E epidemic when many of the sequences were obtained. In the V3 loop, one can see profoundly different behaviors in different subtypes. The subtype D V3 loops are radically divergent. The subtype C V3 loops tend to be more conserved.

The distribution of the similarities in subtype E V3 loops, originating generally either from Thailand or the Central African Republic, is different from the other subtypes, in that the distribution in the subtype E V3 loop similarity scores seems to be bimodal (Figures 1 and 3), reflecting the presence of two subpopulations of sequences. A highly conserved set of protein domains are apparent for the Thai subtype E sequences, but it is interesting to note that some of the phylogenetically similar Thai sequences are in the more divergent group of V3 loop protein sequences ([Yu et al.(1995)], and the data underlying Figures 1 and 3), suggesting a potential for rapid divergence in this domain in subtype E.

To further analyze the greater divergence in subtype D, the sequences were subdivided into two groups, one that retains proline in the tip of the loop (GPGR or GPGQ), the other consisting of all sequences that do not have proline (GLGQ, GSGQ, etc.). As can be seen in Figure 2 (top panel), the distribution of similarity scores in the proline-containing group is very similar to that of other subtypes, while scores in the other group are much lower. Additionally, there appears to be an association between this increased variability and increased positive charge in the V3 loop sequences. Because increased positive charge is associated with the SI phenotype (at least in subtype B), this phenomenon may be related either to greater freedom to diverge or greater evolutionary pressure to change in the SI (CXCR4 using) forms.

It is possible that the observed differences in V3 variation between the subtypes result from systematic sampling artifacts from the different subtypes. This seems unlikely in light of the large number of individuals now sampled, and the efforts of the UNAIDS and DAIDS to collect international samples from recent seroconverters. The data suggest that, in addition to the phylogenetic distinctions, there are distinctions between the subtypes resulting from different evolutionary pressures. Thus biological differences do exist, either in the history of the epidemic of the different subtypes, or in different selective pressures. It is worthy of note that subtypes B and D appear to have shared a common ancestor, after some period of divergence from the other subtypes. Although phenotypic distinctions have not

been clearly detected in the data accrued so far, potential differences are indicated by differences in genetic makeup and rate of evolution, at least within the functionally important V3 loop of the different subtypes. Despite these interesting distinctions, it will be important in the future not to focus the search for biological differences in viruses too narrowly on HIV-1 subtypes. It has already been found that important phenotypic distinctions are found between different lineages within one subtype (such as SI and NSI isolates) and too much emphasis on the subtype distinctions could conceivably result in myopia with respect to these other potential differences.

V3 Loop Evolution

Figure 1

Figure 2

V3 Loop Evolution

Figure 3

Figure 4

COUNTRY CODES

It is becoming increasingly useful to name viral isolates and samples with a country code. The following code was captured from Internet files:

`ftp://ftp.ripe.net/iso3166-countrycodes`

`gopher://sanagn.itu.ch:70/00/.1/ITU-Databases/.1/CtryCodes/.full-list.txt`

for ISO two-letter codes and for ITU three-letter codes, respectively.

This is a list based on the International Organization for Standardization (ISO) 3166:1993 standard, updated from a list prepared by Mark Horton. Note that the original standard has this same information sorted into about 6 different orders, both in English and French, therefore this is an abbreviated version not to be taken as the entire standard. While it has been checked against the standard, it may possibly contain errors; the standard and registration newsletters should be verified for any critical application.

This copy has been updated and is believed to be current through September 1997.

Table of Country Codes from ISO 3166

Country	A 2	A 3	Number
AFGHANISTAN	AF	AFG	004
ALBANIA	AL	ALB	008
ALGERIA	DZ	DZA	012
AMERICAN SAMOA	AS	ASM	016
ANDORRA	AD	AND	020
ANGOLA, REPUBLIC OF	AO	AGO	024
ANGUILLA	AI	AIA	660
ANTARCTICA	AQ	ATA	010
ANTIGUA AND BARBUDA	AG	ATG	028
ARGENTINA	AR	ARG	032
ARUBA	AW	ABW	533
AUSTRALIA	AU	AUS	036
AUSTRIA	AT	AUT	040
AZERBAIJAN	AZ	AZE	031
BAHAMAS	BS	BHS	044
BAHRAIN	BH	BHR	048
BANGLADESH	BD	BGD	050
BARBADOS	BB	BRB	052
BELGIUM	BE	BEL	056
BELIZE	BZ	BLZ	084
BENIN	BJ	BEN	204
BERMUDA	BM	BMU	060
BHUTAN	BT	BTN	064
BOLIVIA	BO	BOL	068
BOTSWANA	BW	BWA	072
BOUVET ISLAND	BV	BVT	074
BOSNIA AND HERZEGOWINA	BA	BIH	070
BRAZIL	BR	BRA	076
BRITISH INDIAN OCEAN TERRITORY	IO	IOT	086
BRUNEI DARUSSALAM	BN	BRN	096
BULGARIA	BG	BGR	100
BURKINA FASO	BF	BFA	854
BURUNDI	BI	BDI	108
BYELORUSSIAN SSR	BY	BYS	112

Country Codes

CAMBODIA	KH	KHM	116
CAMEROON	CM	CMR	120
CANADA	CA	CAN	124
CAPE VERDE	CV	CPV	132
CAYMAN ISLANDS	KY	CYM	136
CENTRAL AFRICAN REPUBLIC	CF	CAF	140
CHAD	TD	TCD	148
CHILE	CL	CHL	152
CHINA	CN	CHN	156
CHRISTMAS ISLAND	CX	CXR	162
COCOS (KEELING) ISLANDS	CC	CCK	166
COLOMBIA	CO	COL	170
COMOROS	KM	COM	174
CONGO	CG	COG	178
COOK ISLANDS	CK	COK	184
COSTA RICA	CR	CRI	188
COTE D'IVOIRE	CI	CIV	384
CROATIA	HR	HRV	191
CUBA	CU	CUB	192
CYPRUS	CY	CYP	196
CZECH REPUBLIC	CZ	CZE	203
DENMARK	DK	DNK	208
DJIBOUTI	DJ	DJI	262
DOMINICA	DM	DMA	212
DOMINICAN REPUBLIC	DO	DOM	214
EAST TIMOR	TP	TMP	626
ECUADOR	EC	ECU	218
EGYPT	EG	EGY	818
EL SALVADOR	SV	SLV	222
EQUATORIAL GUINEA	GQ	GNQ	226
ERITREA	ER	ERI	232
ESTONIA	EE	EST	233
ETHIOPIA	ET	ETH	231
FALKLAND ISLANDS (MALVINAS)	FK	FLK	238
FAROE ISLANDS	FO	FRO	234
FIJI	FJ	FJI	242
FINLAND	FI	FIN	246
FRANCE	FR	FRA	250
FRANCE, METROPOLITAN	FX		249
FRENCH GUIANA	GF	GUF	254
FRENCH POLYNESIA	PF	PYF	258
FRENCH SOUTHERN TERRITORIES	TF	ATF	260
GABON	GA	GAB	266
GAMBIA	GM	GMB	270
GEORGIA	GE	GEO	268
GERMANY	DE	DEU	276
GHANA	GH	GHA	288
GIBRALTAR	GI	GIB	292
GREECE	GR	GRC	300
GREENLAND	GL	GRL	304
GRENADA	GD	GRD	308
GUADELOUPE	GP	GLP	312
GUAM	GU	GUM	316
GUATEMALA	GT	GTM	320

GUINEA	GN	GIN	324
GUINEA-BISSAU	GW	GNB	624
GUYANA	GY	GUY	328
HAITI	HT	HTI	332
HEARD AND MCDONALD ISLANDS	HM	HMD	334
HONDURAS	HN	HND	340
HONG KONG	HK	HKG	344
HUNGARY	HU	HUN	348
ICELAND	IS	ISL	352
INDIA	IN	IND	356
INDONESIA	ID	IDN	360
IRAN (ISLAMIC REPUBLIC OF)	IR	IRN	364
IRAQ	IQ	IRQ	368
IRELAND	IE	IRL	372
ISRAEL	IL	ISR	376
ITALY	IT	ITA	380
JAMAICA	JM	JAM	388
JAPAN	JP	JPN	392
JORDAN	JO	JOR	400
KAZAKHSTAN	KZ	KAZ	398
KENYA	KE	KEN	404
KIRIBATI	KI	KIR	296
KOREA, DEMOCRATIC PEOPLE'S REPUBLIC OF	KP	PRK	408
KOREA, REPUBLIC OF	KR	KOR	410
KUWAIT	KW	KWT	414
KYRGYZSTAN	KG	KGZ	417
LAO PEOPLE'S DEMOCRATIC REPUBLIC	LA	LAO	418
LATVIA	LV	LVA	428
LEBANON	LB	LBN	422
LESOTHO	LS	LSO	426
LIBERIA	LR	LBR	430
LIBYAN ARAB JAMAHIRIYA	LY	LBY	434
LIECHTENSTEIN	LI	LIE	438
LITHUANIA	LT	LTU	440
LUXEMBOURG	LU	LUX	442
MACAU	MO	MAC	446
MACEDONIA, FORMER YUGOSLAV REPUBLIC OF	MK	MKD	807
MADAGASCAR, REPUBLIC OF	MG	MDG	450
MALAWI	MW	MWI	454
MALAYSIA	MY	MYS	458
MALDIVES	MV	MDV	462
MALI	ML	MLI	466
MALTA	MT	MLT	470
MARSHALL ISLANDS	MH	MHL	584
MARTINIQUE	MQ	MTQ	474
MAURITANIA	MR	MRT	478
MAURITIUS	MU	MUS	480
MAYOTTE	YT	MYT	175
MEXICO	MX	MEX	484
MICRONESIA	FM	FSM	583
MOLDOVA	MD	MDA	498
MONACO	MC	MCO	492
MONGOLIA	MN	MNG	496
MONTSERRAT	MS	MSR	500

Country Codes

MOROCCO	MA	MAR	504
MOZAMBIQUE	MZ	MOZ	508
MYANMAR	MM	MMR	104
NAMIBIA	NA	NAM	516
NAURU	NR	NRU	520
NEPAL	NP	NPL	524
NETHERLANDS	NL	NLD	528
NETHERLANDS ANTILLES	AN	ANT	530
NEUTRAL ZONE	NT	NTZ	536
NEW CALEDONIA	NC	NCL	540
NEW ZEALAND	NZ	NZL	554
NICARAGUA	NI	NIC	558
NIGER	NE	NER	562
NIGERIA	NG	NGA	566
NIUE	NU	NIU	570
NORFOLK ISLAND	NF	NFK	574
NORTHERN MARIANA ISLANDS	MP	MNP	580
NORWAY	NO	NOR	578
OMAN	OM	OMN	512
PAKISTAN	PK	PAK	586
PALAU	PW	PLW	585
PANAMA	PA	PAN	590
PAPUA NEW GUINEA	PG	PNG	598
PARAGUAY	PY	PRY	600
PERU	PE	PER	604
PHILIPPINES	PH	PHL	608
PITCAIRN	PN	PCN	612
POLAND	PL	POL	616
PORTUGAL	PT	PRT	620
PUERTO RICO	PR	PRI	630
QATAR	QA	QAT	634
REUNION	RE	REU	638
ROMANIA	RO	ROM	642
RUSSIAN FEDERATION	U	RUS	643
RWANDA	RW	RWA	646
ST. HELENA	SH	SHN	654
SAINT KITTS AND NEVIS	KN	KNA	659
SAINT LUCIA	LC	LCA	662
ST. PIERRE AND MIQUELON	PM	SPM	666
SAINT VINCENT AND THE GRENADINES	VC	VCT	670
SAMOA	WS	WSM	882
SAN MARINO	SM	SMR	674
SÃO TOME AND PRÍNCIPE	ST	STP	678
SAUDI ARABIA	SA	SAU	682
SENEGAL	SN	SEN	686
SEYCHELLES	SC	SYC	690
SIERRA LEONE	SL	SLE	694
SINGAPORE	SG	SOP	702
SLOVAKIA	SK	SVK	703
SLOVENIA	SI	SVN	705
SOLOMON ISLANDS	SB	SLB	090
SOMALIA	SO	SOM	706
SOUTH AFRICA	ZA	ZAF	710
SOUTH GEORGIA AND THE SOUTH SANDWICH ISLANDS	GS		239

Country Codes

SPAIN	ES	ESP	724
SRI LANKA	LK	LKA	144
SUDAN	SD	SDN	736
SURINAME	SR	SUR	740
SVALBARD AND JAN MAYEN ISLANDS	SJ	SJM	744
SWAZILAND	SZ	SWZ	748
SWEDEN	SE	SWE	752
SWITZERLAND	CH	CHE	756
SYRIAN ARAB REPUBLIC	SY	SYR	760
TAIWAN, PROVINCE OF CHINA	TW	TWN	158
TAJKISTAN	TJ	TJK	762
TANZANIA, UNITED REPUBLIC OF	TZ	TZA	834
THAILAND	TH	THA	764
TOGO	TG	TGO	768
TOKELAU	TK	TKL	772
TONGA	TO	TON	776
TRINIDAD AND TOBAGO	TT	TTO	780
TUNISIA	TN	TUN	788
TURKEY	TR	TUR	792
TURKMENISTAN	TM	TKM	795
TURKS AND CAICOS ISLANDS	TC	TCA	796
TUVALU	TV	TUV	798
UGANDA	UG	UGA	800
UKRAINE	UA	UKR	804
UNITED ARAB EMIRATES	AE	ARE	784
UNITED KINGDOM	GB	GBR	826
UNITED STATES	US	USA	840
UNITED STATES MINOR OUTLYING ISLANDS	UM	UMI	581
URUGUAY	UY	URY	858
UZBEKISTAN	UZ	UZB	860
VANUATU	VU	VUT	548
VATICAN CITY STATE (HOLY SEE)	VA	VAT	336
VENEZUELA	VE	VEN	862
VIET NAM	VN	VNM	704
VIRGIN ISLANDS (BRITISH)	VG	VGB	092
VIRGIN ISLANDS (U.S.)	VI	VIR	850
WALLIS AND FUTUNA ISLANDS	WF	WLF	876
WESTERN SAHARA	BH	ESH	732
YEMEN	YE	YEM	887
YUGOSLAVIA	YU	YUG	891
ZAIRE	ZR	ZAR	180
ZAMBIA	ZM	ZMB	894
ZIMBABWE	ZW	ZWE	716

V3 LOOP															
A_CONSENSUS_97	VVIRSENIHMAETIIVQLVEPUNIN	CTRFNNWT	RKSV	RTNGNQDFT	ATG	DIIG	DI	ROANG	MUSRTENMT	LQOATOLL	RKYFKN	NRTE	TWITFANG	SGDDLEITTHSP	NCQGGEFFVQNI
ALL_CONSENSUS	N-S-E	S-G	I	T					I	K	EH		P		T
A_BJ1.1	IT	NO	S-G	I	T				IM	K	EH		P		T
A_BJ1.193	I	F	N-S-E-D	I	I			L-Y	T-N	KR	VAK	QQ			
A_BJ1.210	IN	K	K-R	S	I	T	E		IM	T	R	M			
A_BJ1.23	QV	F-D-T-M	A-A-S	I	I	S	A-E	K	IMGR	H	SK	EQ			
A_BJ1.231	M	L	I-A	N-T	I	GI	H	AD	EV	M	Y	S	QN		
A_BJ1.234	M	L	I-A	T-D	I	I	T		Y	GA	EA	SK	E		
A_BJ1.241	I		AT-R	D	I	G	M	T	Y	H	K	I	EH		
A_BJ1.249	I	A-L	N-A-P-F	S	I	H		I	Y	T	W	S	R	VAK	QQ
A_BJ1.251	M	L	F	D	I	G	H		M	K	Y	DI	D	KA	CH
A_BJ1.252	A	K	D-L	AA	G			T		G		V	EH		
A_BJ1.253	I	D	S-M	AK	G			T	A	K	W			EL	
A_BJ1.256	I		F-T	S	I	R	T	T		RG	D	W	R	VAK	E
A_BJ1.260	I	S	I	K-N-T	I	G	H	T	G	Y	KVK	KG	R	G	M
A_BJ1.281	I		T-R	D	I	G	H	R	T		K			H	
A_BJ1.33	I	I	K-R	D	I	G	H	T		K				R	
A_BJ1.36	II	K	H	DK	R			T		G	K	B	K	K	
A_BJ1.41	M	A	F	N-B-F		I				E	K	I	R	SK	NH
A_BJ1.44	I	K		T	G			T		G	K	A			
A_BY.BLKL0A	I		V	N-T-T		T	I		T	K	T	C	I	NG	V
A_CA.WVCL1				S-Q		I			A	Y			GS	R	VA
A_CF1.11421	See sequences with inserts														
A_CF1.11699	IR	F	LK	D-R	S				Y	I	RR		R	E	R
A_CF1.1189	See sequences with inserts														
A_CF1.1286	See sequences with inserts														
A_CF1.4010	S	L	M	T	I		I	H	N	K	D	K	VQK	RE	W
A_CF1.4018	R	FSS	S	DK	IN	T	I			T	R	M	T	K	
A_CF1.4023	M	D		DS	I		R	M	K	T	F				
A_CF1.4031	R	S	F	S			I			T					
A_CF1.4044	Q	PSB	F	P	TK		S	K	T	I	H	A	R	A	
A_CF1.4050	S	D	L		M	T	I		R						
A_CF1.4054	R	F		N	T	R	H		I						
A_CF1.4055	R	F		N	T	R	S		I						
A_CF1.4058	R	F	D		M	T	R	I	H	S	R				
A_CF2.GAN	I			K	N	T	I	ST		RM	H	S		ID	G
A_CF2.SAS	IN	FS		MOS	E			RM	H	R	I	D	A	V	
A_C11.17con	I			FM	S			G			T				
A_C11.39con	I	S			R		S	GI	H		D	R			
A_C11.41con	IN	L		N	Q		I	G		D	H		E	T	
A_C11.46con	I			PHK	G		S		D	H		G	T	E	T
A_C11.46con	I			D	T		I			T		E	E		
A_C11.51con	I			K	T		G		H		T	F		M	K
A_C12.C114	I			K	R		S		P		D				
A_C12.C12	R			D	H		I	S		N					
A_C12.C120	IA			D	H		S		S	I					
A_C12.C13	I			N			G								
A_C12.C1326	See sequences with inserts														
A_C12.C1327	I			N	T		I	G		G	H				
A_C12.C1329	T			X			G		L		T				
A_C12.C1330	N			D	T		I		H						
A_C12.C142	M			SLAD	T		I		H						
A_C12.C145	A			K	L		E			T					
A_C12.C147	See sequences with inserts														
A_C13.2049	I			TK			G			T					
A_C13.2091	L									V					
A_C13.2130	S									I	H				
A_C13.2132	S									N					
A_C13.2257	K			N	T		I			T	R	A			

V3 LOOP

[illegible]

[illegible]

A_CONSENSUS_97	VVRSENITHNARTYIVOLVSPVKIN	CTRENNY.HKSV.HIGPOQQAIFY	ATG.DIIG.DI.ROANC	AVSRTEWNT.LQVATOLL	HKYFKM.NRTE.TNIIFAN	SOGELEITHSF	MOGGEFFTCNN
ALL_CONSENSUS	II-----N-S-E-----	-----I-----	-----I-----	I-----K-----	-----E-----	-----P-----	-----T-----
A_KL.KEN98:1	II-----V-----	-----S-----T-----	-----	-----S-----K-----X-----D-----	-----	-----	-----
A_KL.KEN98:2	-K-----N-----FNN-----	-----H-----	-----M-----	-----A-----X-----VS-----	-----T-----	-----	-----
A_KL.NA111	-R-----N-----TK-----R-----	-----H-----	-----N-----	V-X-----K-----GE-----X-----	-----T-----Q-----	-----	-----T-----
A_KL.NA112	IK-----N-----FOK-----I-----	-----S-----S-----I-----	-----A-----	-S-----T-----HGB-----T-----K-----GGR-----	-----T-----V-----	-----	-----T-----
A_KL.NA114	-H-----RN-----FK-----A-----	-----H-----	-----F-----A-----	-----R-----KE-----	-----IR-----	-----G-----	-----
A_KL.NA115	IK-----N-----K-----	-I-----XT-----I-----S-----K-----	-F-----S-----N-----	-----KLD-----T-----E-----Q-----	-----E-----R-----	-----K-----	-----
A_KL.NA117	-R-----N-----V-----D-----	-S-----T-----	-----?	-----L-----KA-----	-----VE-----	-----T-----S-----	-----T-----
A_KL.NA118	-I-----D-----V-----DK-----I-----	-----I-----	-----?S-----M-----	-----SL-----DE-----FK-----K-----	-----GB-----	-----N-----M-----	-----
A_KL1.AO9246	IH-----PS-----AQ-----Q-----	-----R-----H-----	-----AV-----N-----	-I-X-G-TE-----R-----EK-----	G-VSN-----IT-----A-----	-----	-----
A_KL1.GH9023	I-----L-----A-----G-----	-----S-----R-----I-----	-----	-----MG-----T-----	-----EH-----	-----DS-----	-----
A_KL1.GH9152	-H-----DL-----T-----	-----SD-----R-----	-----	-----M-----K-----	-----I-----	-----T-----	-----
A_KL1.GH9299	-----N-----	-----CK-----T-----K-----	-----T-----T-----	-----G-----	-----V-----D-----E-----	-----	-----
A_KL1.GH94014	-----S-----I-----	-----R-----	-----T-----	-----Y-----B-----ME-----V-----K-----	-----Y-----	-----N-----N-----	-----
A_KL1.GH95012	I-----N-----	-----DK-----R-----	-----G-----A-----I-----HM-----	-----P-----	-----KD-----Q-----	-----R-----	S-M-E-----A-----
A_KL1.KEN9135	-R-----V-----N-----X-----T-----	-----D-----I-----	-----I-----	-----T-----	-----S-----D-----	-----G-----K-----	-----
A_KL1.RM8903	-R-----V-----N-----D-----S-----T-----	-----	-----T-----	-----	-----GAK-----	-----E-----K-----	G-S-----R-----
A_KL1.RM8935	-K-----N-----N-----PHQ-----E-----	-----Q-----H-----	-----E-----G-----	-----Y-----TI-----K-----BP-----	-----E-----X-----	-----ON-----E-----T-----	-----
A_KL1.T2925825	-R-----SD-----N-----T-----T-----	-----T-----I-----	-----A-----	-----	-----SR-----R-----	-----EK-----VV-----	-----HW-----
A_KL1.T294016	-H-----N-----N-----FWNA-----Q-----	-----S-----	-----H-----	-----G-----	-----Q-----	-----K-----IK-----	BT-----R-----V-----
A_KL1.UO9283	-N-----N-----L-----X-----	-----RRJ-----P-----	-----H-----R-----NV-----	-----K-----	-----SS-----SK-----	-----AA-----	G-H-G-----T-----V-----G-----
A_KL1.ZR93033	-AV-----N-----L-----FT-----R-----	-----S-----	-----I-----	-----G-----	-----	-----MD-----K-----	YS-----N-----F-----NV-----
A_RU1.GAN1	-----N-----E-----	-----S-----	-----GI-----H-----	-----	-----I-----R-----Q-----	-----	-----
A_RU1.LV45	-----K-----N-----T-----	-----G-----	-----	-----T-----	-----T-----N-----	-----	-----
A_RU1.RM470	-I-----K-----SB-----TNN-----RT-----H-----	-I-----LA-----	-----L-----N-----	-I-----TN-----	-----	-----I-----DQ-----D-----	YSI-----N-----M-----
A_RU1.SHL9	-----K-----E-----TNN-----KT-----Q-----	-----G-----	-----L-----H-----	-I-----E-----M-----R-----	-----	-I-----AGL-----E-----YE-----	-----H-----N-----
A_RU3.2729B	-----D-----G-----I-----T-----N-----T-----	-I-----G-----	-----T-----I-----	-----T-----V-----	-----K-----Y-----	-I-----AA-----S-----	-----KI-----
A_RU3.323	-----D-----G-----I-----T-----N-----T-----	-I-----G-----	-----T-----I-----	-----T-----V-----	-----K-----Y-----	-I-----AA-----S-----	-----KI-----
A_RU3.6525B	-----D-----G-----I-----T-----N-----T-----	-I-----G-----	-----T-----I-----	-----P-----T-----	-----V-----	-I-----AA-----S-----	-----KI-----
A_RU3.7517	-----D-----S-----I-----T-----N-----T-----	-I-----S-----	-----T-----I-----	-----T-----	-----SV-----	-I-----AA-----S-----	-----KI-----
A_RU3.7624B	-----D-----G-----I-----T-----N-----T-----	-I-----G-----	-----T-----I-----	-----T-----	-----VF-----	-I-----R-----AA-----S-----	-----KI-----
A_RU3.8722B	-----D-----G-----I-----T-----N-----T-----	-I-----G-----	-----T-----I-----	-----T-----	-----V-----	-I-----AA-----S-----	-----KI-----
A_RU3.9022	-----D-----G-----IT-----T-----N-----T-----	-I-----G-----	-----T-----I-----G-----	-----T-----	-----V-----	-I-----AA-----S-----	-----KI-----
A_RU3.936	-----D-----G-----I-----T-----N-----T-----	-I-----G-----	-----T-----I-----	-----T-----	-----V-----	-I-----AA-----S-----	-----KI-----
A_RU3.947	-----D-----G-----I-----T-----N-----T-----	-I-----G-----A-----	-----T-----I-----	-----T-----	-----V-----	-I-----AV-----S-----	-----KI-----
A_RU3.9506	-----D-----G-----I-----T-----N-----T-----	-I-----G-----	-----T-----I-----	-----T-----	-----V-----	-I-----AA-----S-----	-----KI-----
A_RU3.9511	-----D-----G-----I-----T-----N-----T-----	-I-----G-----	-----T-----I-----	-----T-----	-----V-----	-I-----AA-----S-----	-----KI-----
A_RU3.9704B	-----D-----G-----I-----T-----N-----T-----	-I-----G-----	-----T-----I-----	-----T-----	-----V-----	-I-----AA-----S-----	-----KI-----
A_RM.964C	-R-----A-----D-----M-----H-----S-----P-----Q-----	-----	-----T-----	-----E-----N-----	-I-----	-----LN-----A-----Q-----	-----K-----V-----G-----G-----T-----R-----T-----
A_RM.SF1703	-E-----N-----	-----NKT-----S-----	-----	-----T-----	-----	-----AU-----K-----	-----S-----M-----KS-----S-----S-----
A_RM1.W2RM008	-----V-----	-----K-----	-----T-----I-----	-----S-----M-----	-----	-----S-----BA-----R-----VE-----	-----GH-----G-----M-----T-----
A_RM1.W2RM016	-----DRA-----N-----T-----	-----I-----	-----P-----R-----	-----N-----	-----	-----S-----BA-----R-----VE-----	-----GH-----G-----M-----T-----
A_RM1.W2RM02	-----DRA-----	-----	-----	-----	-----	-----AK-----D-----	-----RG-----KK-----
A_RM1.W2RM024	-I-----I-----L-----R-----	-----T-----	-----T-----	-----R-----Y-----	-----	-----YIN-----	-----N-----G-----B-----GR-----
A_RM1.W2RM025	-----N-----	-----K-----	-----V-----	-----T-----B-----	-----R-----F-----	-----T-----NES-----E-----R-----	-----G-----H-----P-----K-----T-----
A_RM1.W2RM026	-----S-----	-----R-----I-----	-----T-----	-----K-----			

V3 LOOP

```

A_CONS26250_97
ALL_CONSERVED
A_R04_074
A_R04_081cons
A_R04_082cons
A_R04_226
A_R04_439cons
A_R04_530cons
A_R04_564cons
A_R04_566cons
A_R04_618cons
A_R04_730cons
A_T2_TAM142
A_T2_TAM15CON
A_T2_TAM8
A_T2_TAM9CON
A_T22_016
A_T22_017
A_T23_1574
A_T23_1576
A_T23_1577
A_T23_1584
A_U0_1033
A_U0_9200037
A_U0_964
A_U0_U455
A_U0_U506
A_U01_W200029
A_U01_W20003
A_U01_W200037
A_U02_11
A_U02_115
A_U02_116
A_U02_117
A_U02_118
A_U02_119
A_U02_124
A_U02_72
A_U02_78
A_U02_82
A_U02_84
A_U04_08273
A_U04_09275
A_U05_9400003
A_U05_9400004
A_U05_9400009
A_U05_9400011
A_U05_9400012
A_U05_9400016
A_U05_9400018
A_U05_9400019
A_U06_95052
A_U06_95053
A_U06_95058
A_U06_95177
A_U06_95217
A_U06_95303
A_U06_95321
A_U06_95329
A_U06_95337
A_U06_95347
VVRSENITHARTIVQLVAFVKH
II
-QV-----M-----DKA-E-----
-R-----NQ--E-----
-K-----I-T-----S-----
-K-----D-T-I-----MKS-E-I
IR-----N-L--R--D-----S-----
-K-C-----D-A-----H-----
-R-A-----M-H-NTS-Q-----
INV-F-----D-S-----FDRA-E-----
-R-----V-L--K-----T-I-----
-Q-----FM-T-R-----H-----
-R-----N-PA-----H-----
-K-----N-T-E-----I-----TS-
-M-----N-PIK-E-----N-----
-R-----N-FSR-----R-H-----
-M-----V-N-FT-----S-----
-H-----V-N-T-Q-----S-----
-V-----L-TQ-E-----I-----A-----
-----HGEH-M-E-----K-----
-----N-L-FAK-E-----F-----T-H-----F-----A-----K-----
-----V-----D-E-----I-K-----
M-----N-FT-----I-HL-R-----
-H-----V-N-N-S-T-----R-----T-----
-T-A-----N-FAR-R-----E-----
see sequences with inserts
-K-----SD-----TK-T-----YKRV-ARI-H-----R-S-----SM-----V-----GQ-K-----V-K-----E-W-----T-----N-SXP-----V-----
Sea sequences with inserts
-----S-----H-N-B-----I-----M-----T-K-B-K-----T-V-EB-----W-----T-----
-----V-N-N-S-T-----R-----T-----R-----GQ-K-----H-VE-----M-----M-----MS-----
-M-----L-N-TE-----T-----H-----R-----KV-K-----VS-----T-----T-----P-----
-N-----N-T-----I-----I-H-----R-----N-----KA-K-----VS-----T-----T-----P-----
-N-----N-L-FNTS-Q-----G-H-----E-----E-----ESLG-----DK-----P-----
IN-----N-L-FN-T-Q-----T-----N-----I-QAK-K-----KK-----GDLN-KT-----MT-----
-R-K-----N-FRT-----G-H-----M-----R-----E-N-----R-----E-K-----TS-G-----T-----P-----
II-----L-N-KN-I-----Q-K-X-----Q-V-----Y-----AA-K-----G-L-----H-B-----T-----P-----
A_U02_124
-K-----SD-----FN-S-T-----K-REA-K-----D-E-N-----T-----
-K-----N-FTE-----R-H-----A-S-----R-K-----E-VK-----T-G-----
A_U02_78
-K-K-----SD-----L-TK-----G-----R-----T-----S-KE-K-VK-----THM-----T-----
A_U02_82
-F-----N-FTR-M-T-----I-----T-----T-----AD-M-----K-Y-E-KK-----R-----E-V-----TE-R-----T-----P-----
A_U02_84
IN-----N-PT-----I-----I-G-----T-----AD-M-----G-K-K-----B-K-----EM-GR-----P-----
-K-R-----SD-----L-TT-T-----G-----T-----S-----I-S-KE-K-VK-----THM-----T-----
-K-----N-L-TT-T-----V-----G-H-----Y-----E-M-G-K-E-----E-M-KT-R-N-----T-----P-----
A_U05_9400003
-----I-----T-I-H-----T-----T-----T-MG-U-----H-----TTN-V-----S-K-R-----I-T-----L-----
A_U05_9400009
-----E-S-----I-----M-----AK-E-----K-R-N-----TH-G-----GR-L-----I-----M-G-----
A_U05_9400011
-----E-----I-HL-----A-----G-KAK-H-----TK-----TF-E-----V-----L-IV-----L-----
A_U05_9400012
-----T-----G-----I-B-----K-----999-R-----EXMGK-----ABHTP-----A-MS-----V-----E-----
A_U05_9400016
-----T-----G-----I-B-----K-----S-ONG-----G-VE-----T-M-N-----T-----I-----G-----L-----
A_U05_9400018
-----T-----G-----K-----HKA-----KI-X-----GT-----D-----VL-----A-----
A_U05_9400019
see sequences with inserts
-----D-V-N-----FN-S-P-----RG-K-----V-----A-EA-KX-UN-----G-----A-E-----D-----
A_U06_95052
-----SD-----FTR-L-----I-----I-----T-----SC-EA-YH-----HWE-T-----T-----V-----
A_U06_95058
-----SD-----N-TP-E-----I-----I-----T-----SC-E-----V-----H-----M-TGA-----I-----
A_U06_95177
-----N-T-E-----I-----H-----AK-A-----G-E-----R-E-----TKP-----
A_U06_95217
-----SD-----TT-P-T-----I-----I-----D-WGK-E-----G-VE-----QHN-----D-----F-----
A_U06_95303
-----N-L-FN-S-E-----I-----M-----D-N-N-----K-KA-X-VE-----G-E-----
A_U06_95321
-----SD-----TTN-P-----I-----I-----SV-----Y-----SK-VA-----V-----KEHW-----N-A-T-----
A_U06_95329
-----D-N-FTR-R-----I-----H-----AD-DEA-----IR-----Q-E-----V-----R-----
A_U06_95337
-----N-T-D-----I-----OI-K-----T-P-----R-----F-----T-MG-----K-VN-----G-H-G-----MS-----
A_U06_95347
-----N-FTR-H-I-----I-R-----K-----E-N-----D-----E-I-----TH-V-----D-----T-R-----

```

V3 LOOP

```

A_CONSENSUS_97  VVIRSENITDPAKTIIVQLVSPVGIN CTAPMNT.FKSV.RIGPQQAFY.ATG.DIIQ.DI.RQANC NWSRTENMT.LQOVATQLL.RKYFEN.NKTE.THIIFANE.SGGOLEITVTSF.NCGGSEFTYCNM
ALL_CONSENSUS3  II-----N-S-E-----R-H-----N-----I-----KK-----EH-----P-----T
A_UG6.95350      -----S-L-N-R-----R-----N-----Y-----KK-----VQ-----TK-G-----P.I-----R-----T
A_UG6.95361      -----SD-----DKY-R-----I-----R-----B-----N-----H-Y-----E-KQSA-E-----K-VK-----LMS-----MS-----
A_UG6.95363      -----N-----PK-----P-----S-H-----B-----N-----H-Y-----T-N-A-----E-K-----TH-GE-----
A_UG6.95375      -----S-----TK-R-----I-----T-----TS.K-----M-----R-Y-----T-MSEA-KD-----VQ-----DM-----NT.A-----
A_UG6.95420      -----SD-I-----N-E-----P-I-----H-----T-----R-Y-----KSS-EE-----X-----G-HMN.TT-----NS-----R.R-----
A_UG6.95421      -----S-V-----FN-S-T-----G-----I-H-----N-----SA-K-----VD-----HWV-----T-----NS-----
A_UG6.95434      -----O-N-----F-E-----T-I-H-----N-----E-D-----D-----N-G-----H-----P-----
A_UG6.95456      -----N-----NQS-I-----H-H-----R-----E-N-A-----RK-VE-----TH-G-----T-----D-----
A_UG6.95517      -----SD-----PTQ-I-----I-----I-----H-----GS-R-----VN-----HME-A-----T-NS-----V-----
A_UG6.95539      -----NL-----PTK-R-----I-----IRN-Q-----T-----K-----N-----A-B-----N-SR-----Y-G-----T-----F-----
A_UG6.95563      -----FMQS-E-----I-----H-----D-G-----N-----N-TR-----HR-VL-----G-----S-NG-----
A_2R.232      Seq sequences with inserts
A_2R1.6557.sh    -----S-----D-S-----S-R-Q-----R-----R-----T-N-K-----D-XA-----E-N-----T-----
A_2R1.6559      -----D-----SQQ-H-----V-----RD.R-----K-Y-----E-N-AV-----R-R-----GE-G-----N-TR-----A-----
A_2R1.6563      -----F-D-G-----N-S-----I-----RGI.HM-----IL-----S-----TIN-K-----R-----E-TT-----T-----H-----
A_2R1.6569      -----A-----KQ-R-----I-----RGI.HM-----IL-----S-----KKD-BA-----HK-----NH-N-----NS-----
A_2R1.6571.sh    -----A-D-N-----FN-S-Q-----I-----V-----TN-----I-----A-YA-----K-VX-----H-P-----
A_2R1.6589      -----T-----P-----H-----N-----A-SK-----D-----TH-N-----MRP-----
A_2R1.6593      -----F-----L-----DS-----I-----X-----N-----AA-----K-K-----G-----R-----NS-----
A_2R1.6595.sh    -----N-S-S-----I-----V-----E-----E-----HK-VE-----GT-R-----T-----T-S-----
A_2R1.6597      -----P-M-Q-----T-----T-----R-----A-----BK-VE-----ATL-A-----T-----VV-TS-V-----
A_2R1.6599      -----F-----HKT-Q-----T-----R-I-H-A-----A-----R-----KKD-D-----Y-----GEQ-G-----RN-----V-----T-----
A_2R3.K114      -----R-----F-D-T-N-----AK-T-----I-----R-----R-----K-----L-----Q-----E-----A-MK-----V-----T-----
A_2R3.L414      IA-----S-----F-H-E-----I-----S-----RGI.ML-----N-A-----G-R-E-----RK-V-----N-GE-----V-TKP-----V-----T-----
A_2R3.P104      -----DK-I-----S-----RGI.ML-----N-A-----G-R-E-----RK-V-----N-GE-----V-TKP-----V-----T-----
A_2R3.P261      -----V-----D-----I-----G-I.ML-----T-C-----E-----G-----BA-----K-VG-----H-GE-----V-T-P-----V-----T-----

```

	AAA	A	AA	AAA	V3 LOOP	S	AAA	AA	A	AAA	AA	A	A	A
B_CONSENSUS_97	VVIRSENPTDMARTIIVQNMSEVEIN	CYRENDQWFRKXI	HIGSGNGAFYYTIG	ELIG	DILRQAMC	NLSRAHNMNT	LRQIVKKLLERQFQFN	NMT	T	IVFMQSISSGQPEIVMHSF	NCGG	EFFYCN	T	
ALL_CONSENSUS	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR.21201	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.06	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.13	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.21	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.22	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.23	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.24	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.39	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.33	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.34	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.35	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AR1.56	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU1.NRC	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU1.NRC2	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU1.NRC3	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU2.C1BOG	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU3.10631	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU3.10891	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU3.9411	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU3.EG3	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU3.JH951	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU3.L1FT1	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_AU3.TR1	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_EE.SIN194	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR.002can	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR.B2	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.10553	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.1056	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.10565	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.10575	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.1058	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.1059	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.10593	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.7930	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.7932	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.7934	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.7936	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.7940	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.7942	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.7946	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.8615	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.862	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.8623	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.8625	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.8633	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR1.8635	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR003	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR004	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR014	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR017	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR018	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR019	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR020	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR021	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR023	I	I	I	I	I	I	I	I	I	I	I	I	I	I
B_BR2.W2BR024	I	I	I	I	I	I	I	I	I	I	I	I	I	I

V3 LOOP

A_CONSENSUS_97 VVASENPTONAKTIIVQNESVEIN CTAPNNNTTRKSI.RIGPGRAFIYITG.EIIG.DILQDANC NLSRAQSDMT.LAQIVKILLERQPGN.NKT..T.IVFNQSISSGQPEIVMHSF.NOGG.RFFYCNT
 ALI_CONSENSUS_97 II--I-N-----R-----Q-----A-----D-----I-TE-----Q-VA-----H-----I-AP-----L--TT-----
 B_BR2.W2BR026S-T-----P-----T-----A-----I-MT--BS--E-----N-Q-----T-----I-SX--RT--G--A--K--L-200..II..Q.V-----
 B_BR2.W2BR028L-M-----T-I-----M-H--T--A-----I-Q-----G-----
 B_BR2.W2BR030P-Q-----S-----G-----A-----D-----
 B_BR3.HR317N-----V-----M-----V-----L-----A-----
 B_BR3.HR327.shR-----GG--INKE.N-----
 B_BR3.HR3477I-N-V-----V-----R-----M-L-----A-----D-----Y-----
 B_BR3.HR3625T-----NF--AR--OPT-----
 B_BR3.HR3626N-----IA-----R-----S-----A-----D-----
 B_BR3.HR3636 G-----N-----I-----Q-----V-----P-A--S--HF.A-----KVM-B-----A-----B-----Q-----
 B_BR3.HR370YVK-S-----S-----L-W-----A-----D-----Y-----
 B_BR3.HR3101 II--Q-IG-----I-H-----N-----A-----D-----I-Q-----B-AI--G--K-----K--A-T-----
 B_BR3.HR3104I-H-----G-----T-----TLF.A-A-----I-A--TK-AM--K-----S-----K-----ERG-----
 B_BR3.HR3204 See sequences with inserts
 B_BR3.HR3205 -I-----I-----V-----S-----A-----G-A-D-----T-----N-----
 B_BR3.HR3228A2 Y-----V-----M-----KT-A-----N-N-----R-----X-----E-----AIR--EM-----NISYN-----
 B_BR3.R312L-N-----K-V-----S-----M-----I-----A-----K-E-----E-A-----X-----K-QPW.QDET-----
 B_BR3.R314V-----N-----I-----H-----T-----V-----L-----N-----I-QT--Q-----M-AA-----X-----K-QPW.QDET-----
 B_BR3.R319 G-----D-----N-----I-----H-----T-----QY-T-G-----A-----V-----I-KE-----AT-----K-----
 B_BR3.R340I-H-----I-----V-----S-----M-----N-----V-----SD-K-----EK-AT-----H-K--B--P-----MT-----
 B_BR3.R354I-----V-----Q-----L-----T-----A-----S-----Q-----L-----T-----A-----S-----
 B_BR3.R359N-----G-----L-N-----A-----A-----T-MKT--EK--E-----A-----V-----T-----
 B_BR3.R362V-----N-----I-----S-----L-N-----A-----V-----B-----T-----AI-----K-----E-A-N-----
 B_BR3.R364N-----L-N-----Q-----D-----E-----EK-----AI-----VI-----G-RNOM-----
 B_BR3.R3379D-----I-----V-----G-----D-----D-----D-----R-AI-----N.GTB-----I-----E-----
 B_BR3.R3478N-----X-----Q-----M-N-----H--M-----M-----GIB-----T-AE-----I-T-P-----
 B_BR3.R3483I-----M-----V-----G-----M-L-----A-----G-V-----I-----K-----E-A-----R-M-----
 B_BR3.R3484L-----V-----S-----G-----M-----A-----T-----K-Y-----I-T-E-----R-AI-----E-----I-----N-----T-----D-----
 B_BR3.R3623V-----I-----T-----S-----R-----L-G-----G-----N-----E-F-----T-MT--E-----AE-----A-E-N-----
 B_BR3.R3626N-----I-----Q-----V-----P-A--S--HF.A-----I-----KVM-B-----A-----B-----Q-----
 B_BR3.R3636 G-----N-----I-----Q-----V-----P-A--S--HF.A-----I-----KVM-B-----A-----B-----Q-----
 B_BR3.R3102 I-V-----M-----I-----A-----V-----V-----V-----I-----A-----I-ETR-K-----EK-A-----K-VV-----X--S-Q-A-----
 B_BR3.SPB1 I-----N-----H-----V-----S-----M-----N-----A-----D-----R-----A-----S-----R-----R-----
 B_BR3.SPB2 I-----K-----T-----I-----T-----M-----N-----A-----D-----I-----
 B_BR3.SPB3 I-X-----S-----T-K-D-----I-----G-----M-N-----T-----A-----R-----KVA--R-----ER-AI-----M--NT--M--D.KN-----D-----
 B_BR3.SPB4 -Q-----N-----TP--V-----AK-S-----V-----KR--D-----I-----TE-----H-K-----I-C-----R-----
 B_BR4.B2167A See sequences with inserts
 B_BR4.B2200AM-----Q-----V-----S-----L-P-----L-A-----I-MTE--K--N-AI-----G-----
 B_BR5.S04L-M-V-----M-K-H-----G-----G-----L-K-----A-----I-----T-----Q-VAI-----
 B_BR5.S05I-N-----V-----K-----F-T-----R-----T-M-----LP-----G-----ED-----I-----G-----R-----A-R-----
 B_BR5.S06I-S-----T-----E-----G-----M-N-----T-----A-----Q-----I-EN-S-----I-----G-----N.KT-----TI-P-----
 B_BR5.S10V-T-----Q-----S-----M-----A-----A-----SV-----R-----R-----H-----I-----
 B_BR5.S11M-----T-----V-----T-----M-V-----L-A-----D-----XVQ-EK-----AI-----A-M-----
 B_BR5.S13M-----I-----K-----V-----M-----N-----M-----T-LD-E-----VA-----M-----
 B_BR5.S13T-----I-K-E-----G-----M-N-----T-----A-----V-Q-----L-----AI-----B-----I-----
 B_BR5.S14N-----T-----G-----M-N-----T-----A-----T-----I-S-----TK-H-----H-----
 B_BR5.S15SM--K-----T-----G-----M-N-----V-----A-----I-K-----E-----H-AI-----K-----H-----E-----I-----
 B_BR5.P3I-N-----T-V-----S-----M-L-----V-----K-----I-T-----EK--G-----
 B_BR6.P4I-----N-----T-----G-----M-N-----L-V-----Y-----T-----S-----AI-----K-----P-----T-----
 B_BR6.P6D-----M-----V-----S-KK--RR-----I-Q-----V-----M-----V-----T-----
 B_BR6.P7M-----M-----K-----A-----S-----K-----V-F-----L-A-----R-----XTE--ER--E-V-I-----S-----M-T-----T-----
 B_BR6.P9I-----M-----KE-----G-----V-----R-X-----I-----L-----D-----T-----V-----S-----D-KP-----L-----T-----ME-----
 B_CAL.B11 II--A-----K-----K-----Q-----Q-----V-----I-OTB-----B-----S-----K-----V-----R-----T-----
 B_CAL.B20 I-----A-----B-----K-----V-----S-----Q-----K-----A-----Q-T-----FIN-----R-V-G-----K-----KI-----T-----
 B_CAL.C1013BM-----K-P-E-----I-----N-----K.TL-----V-----IN-TO-----K-----E-----I-RN-----T-----
 B_CAL.C103 I-----D-----N-----IV-----A-----D-----M-----Y-----IIN-TO-----K-----G-----I-----
 B_CAL.C103 -I-----D-----N-----T-K-----P-----I-----T-----E-----K-----
 B_CAL.C104 I-----D-----SN-----MDE-N-T-----L-T-----RG-----V-----V-----A-D-K-----N-----I-----VE--K-----E-----AT--QQ-----

	***	***	V3 LOOP	***	***	***	***											
B_CONSENSUS_97	VVIRGENFYINARTIIIVQLMESVEIN	CTRPNNQVTRRSI	HIGPGRGFYTTG	ELIG	DILQAMC	MLSRADMMAT	LRQIVKLLERQFGN	NKT	T	IVFQGISGGDFIWSHF	NCQG	EFFYCHT						
ALL_CONSENSUS	II	I	N	R	Q	A	D			I	TE	Q	VA	H	I	AF	L	TT
B_CAL.C103																		
B_CAL.C106		A	S	D	K	E												
B_CAL.C107			N															
B_CAL.C109		S			T													
B_CAL.C110	I	D	I	T	M	K	X											
B_CAL.C111			N			V												
B_CAL.C112		A		TR	H													
B_CAL.C113					H													
B_CAL.C114	I		N	R	X													
B_CAL.C115				Y														
B_CAL.C116	I		N		I													
B_CAL.C117			N		N	I												
B_CAL.C118			N		V													
B_CAL.C119			N		K	Q	V											
B_CAL.C120		A		N	RI	A												
B_CAL.C121																		
B_CAL.C122																		
B_CAL.C123																		
B_CAL.C124																		
B_CAL.C125																		
B_CAL.C126																		
B_CAL.C127																		
B_CAL.C128																		
B_CAL.C129																		
B_CAL.C130																		
B_CAL.C131																		
B_CAL.C132																		
B_CAL.C133																		
B_CAL.C134																		
B_CAL.C135																		
B_CAL.C136																		
B_CAL.C137																		
B_CAL.C138																		
B_CAL.C139																		
B_CAL.C140																		
B_CAL.C141																		
B_CAL.C142																		
B_CAL.C143																		
B_CAL.C144																		
B_CAL.C145																		
B_CAL.C146																		
B_CAL.C147																		
B_CAL.C148																		
B_CAL.C149																		
B_CAL.C150																		
B_CAL.C151																		
B_CAL.C152																		
B_CAL.C153																		
B_CAL.C154																		
B_CAL.C155																		
B_CAL.C156																		
B_CAL.C157																		
B_CAL.C158																		
B_CAL.C159																		
B_CAL.C160																		
B_CAL.C161																		
B_CAL.C162																		
B_CAL.C163																		
B_CAL.C164																		
B_CAL.C165																		
B_CAL.C166																		
B_CAL.C167																		
B_CAL.C168																		
B_CAL.C169																		
B_CAL.C170																		
B_CAL.C171																		
B_CAL.C172																		

V3 LOOP

B_CONSENSUS_97
ALL_CONSENSUS97
B_CNL.073
B_CNL.074
B_CNL.080
B_CNL.141
B_CNL.142
B_CNL.144
B_CNL.147
B_CNL.149
B_CNL.150
B_CNL.151
B_CNL.157
B_CNL.159
B_CNL.161
B_CNL.162
B_CNL.163
B_CNL.166
B_CNL.167
B_CNL.169
B_CNL.170
B_CNL.171
B_CNL.173
B_CNL.174
B_CNL.192
B_CNL.193
B_CNL.195
B_CNL.196
B_CNL.199
B_CNL.200
B_CNL.202
B_CNL.206
B_CU.950041
B_CYL.HO042
B_CYL.HO11
B_CYL.HO21
B_CYL.HO25
B_CYL.HO27
B_CYL.HO28
B_CYL.HO39
B_CYL.HO40
B_CYL.HO45
B_CYL.HO48
B_CYL.HO294
B_CYL.HO433
B_CYL.HO464
B_CYL.HO503
B_CZ.BTSPH
B_DE.D3
B_DE.HAN
B_DE.aerococci
B_DE1.A
B_DE1.C
B_DE1.D
B_DE1.E
B_DE1.F
B_DE1.I
B_DE1.J
B_DE2.3493
B_DE2.3497
B_DE2.3498

VVIRSDPTNARTIIVQLNESVEIN CTRFHWVTRKSI HIGPGGAFYITG EITG DILQGANL HLRADNNT LKQIVKLLNEQFGN MKT T IVFQSISSGGDETVNHSF MCGG EFFYGMT

II--I-N-----R-Q--A-D-----I-TE--Q-VA--H--I-AP--L--TT--
S--V-----X--C--YL--N--Q--N--STQ--TE--
S--RV-----L--Q--W--Q--ST--TE--H--
N-S-N-P--A--Y--CH--RV--R--I-Q-Y--E--E--G--S--H--
S--RV-----L--Q--W--Q--STQ--R--TE--
S--RV-----D--I--L--Q--WH--Q--STE-H--R--TE--
S--V-----PL-S--W--Q--STQ--TA--
S--V-----PL--Q--W--Q--ST--AS--
S--N-RV-----L--Q--W--Q--ST--T--R--
S--RV-----PL--Q--W--Q--R-STQ--TE--
S--RV-----PL-L-K--W--A--Q--ST-S--AB--
S--T-I-----L--Q--W--H--STQ--TE--
S--RV-----L--Q--W--Q--H-ST--Q--AE--G--V--
S--TRV-----SL--Q--W--Q--STQ--K--TE--
Q--M--V-----SL--WK--Q--S-Q-T--AE--H--V--E--
R--S--QV--K--A--FL--W--A--Q--ST--H--VTE--
H--H-----VL--H--Q--K--Q-Q-Q--E--I--V--T--
S--T-V-----L--Q--W--Q--STQ-T--T--Q--T--
S--V-----RV--PL--W--Q--ST--N--K--
S--N-V-----PL--Q--W--Q--ST--T--
S--RV-----SL--Q--W--Q--STQ-K--TE--
S--V-----PL--Q--W--Q--T-STQ--R--TE--
S--RV-----I--L--Q--W--M-V--STQ--T--K--
S--V-----RV--PL--W--Q--ST--N--K--
S--RV-----L--Q--H--Q--STQ--TE--T--
S--RV-----L--Q--N--Q--ST--R--TE--
D--N-----T--YL--N--Q--R--Q--AB--
S--RV-----PL--K--W--Q--ST--T--A--
S--N-V-----T--PL--H--Q--STQ--R--TG--
S--V-----L--H--Q--ST--R--T--
S--RV-----PL--K--N--Q--ST--T--
S--N-----S--A--E--K--N--I--D--R--T--T--S--
SN-----T--S--A--D--X--I--P--
SN-----KDA--Y--Q--R--RZ--L--V--K--N--X--T--S--G--V--D--A--
NY-----V--K--N--E--A--A--SKA--R--AL--K--E--S--
N-----K--P--Q--X--H--R--K--K--K--S--X--TD--V--I--K--P--
I--N-----I--M-----AIE--V--A--G--V--RRL--I--P--Y--I--P--
N-----K--I--I--F--R--
N-----TI--F--A--D--X--RT--Y--AT--YV--S--H--
I--A--SN-V--K--A--R--TH--VM--T--R--I--T--NGT--H--YAA--I--
N-----T--K--T--Q--T--A--V--I--
N-----H--H--B--G--E--Q--V--K--I--A--N--E--G--K--I--OP--
N-----V--I--M-----A--Q--R--SKA--RK--I--K--S--S--
See sequences with inserts
I--S--L--T--I--FM--A--DV--I--GB--RK-AA--K--
I--S--X--R--R--M--A--D--QD--D--T--K--
N-----K--K--H--A--O--I--K--K--
D-----K--V--B--R--R--AR--K--K--I--O--DS--R--R--T--
See sequences with inserts
N-----I--V--R--N--A--N--L--T--E--R--VAR--
I--KRP--I--A--R--V--K--
I--K-P--KKK--RR--A--N--V--K--Y--
I--X--X--P--
T--N--X--V--G--X--
E--T--X--V--S--Q--Q--T--
N-----KA-XX--Q--A--
N-----H--KA-XX--P--EK--V--
Q-----I--A--A--I--K--EK--I--R--
D--SI--T--P--V--Y--XG--P--X--D--I--E--K--XV--E--AX--K--A--
N-----V--V--N--A--D--Q--A--R--I--XD--E--Q--

V3 LOOP

B_CONSENSUS_97 VVIRSENFTDAAKTIIVQLMEGVETIN CTRPQNTTRKSI.NIGPQMGAFYVTPG.BIIQ.DILQAMNC NLGRANKYMT.LKQIVKKLLERQFON.NKT..T.IVPQSI5GGDPEIVWHSP.MCOG.BPFFYMT
 ALL_CONSENSUS -I--I--N-----D-----R--Q--A--D-----I--TE--Q--VA--H--I--AP--L--TT--
 B_DS2.3499 -M-X--N-VRN-----G-----X--S-G--R-----A-E-A-T-----I-----Q-----G--V-----L--TT--
 B_DS.03 -----L-----KDP-Q-----R--T-----P-----Y-----
 B_DS.06 -I--D-----V-----P-----A--D-----
 B_DS.07 -----D-S-T-----K-I-----P-----A--D-----
 B_DS.09 -----T-----P-----A--D-----
 B_DS.10 I--Q--T-----AIK-----P-----A--D-----
 B_DS.13 I--R--T-----S-V-----M-----
 B_DS.14 -----A--N--H-----R--K-V-----A--D-----
 B_DS.15 -----D--N-V-----K-P-----P-----A--D-----
 B_DS.17 -----D--N-V-----K-P-Q-----R--K-X-----D--N-----
 B_DS.18 --V--I--T-----R-AIQ-----T-----A--D-----
 B_DS.20 --V--A--S--T-----K-P-I-----S-Q-----T--A--N-----
 B_DS.26 -----NHV-----KDP-A-----G-----A--D-----
 B_DS.106 I-----I-----H--I-----I-----G-----M-----T-----N-----
 B_DS.108 I-----I-----H--IQ-----G-----M-----T-----N-----
 B_DS.017 -I-----N-----RV-S-----WH-----Q-----K-V-----
 B_DS.III -----V-----T-T-----G--P-----A-A-S--V--T-----
 B_DS.IV I-----N-----G-----L-----
 B_DS.IX -----I--T-----TI-----Y--RV-S-----R--A--B-----
 B_DS.P12 -----D--S-----T-----G-----L--G-----A--D-----
 B_DS.P13 II-----S-----QTI--T-----M--H--R-----L--S-----A--D-----
 B_DS.P14 -----V-----V-----YKV--RR-----A--D--IT-----
 B_DS.P17 -----P--R-----I-----RG--N-----A--D--N--E-----
 B_DS.P18 I-----L--N-----P--Q--K--I-----Q-----X--A--R-----S-----
 B_DS.P19 -----D--N-----K--P-----I-----G-----V--A-----
 B_DS.P20 -----D--V-----T--I-----S-----FT-----M-----TVH--AI--R-----S-----
 B_DS.P21 -----SN--V-----TIQ-----I-----S-----
 B_DS.P6 I-----D-----V-----I-----S-----
 B_DS.P9 -----K--I--P-----I-----I-----H--A-----
 B_DS.R1 -----N-----K--IK-----RGV--R--G-----N-----
 B_DS.S1 -----D--N-----K--AI-----P-----I--D-----L-----
 B_DS.S61 See sequences with inserts
 B_DS.THF13A I-----A--I-----RG-----G--D--T--M-----
 B_DS.THF17 I-----N-----H--I-----S-----R-----
 B_DS.V -----V-----N-----K-----G-----A--D-----
 B_DS.VI -----V-----N-----K-----G--AL--G--V--A-----
 B_DS.VIII I-----N-----IA-----G--L--G--P--A--M-----
 B_DS.X -----N-----K--P--X-----G-----A--D-----
 B_FM.CA -----I--N--V-----K-----RR--T-----VY-----A--X-----
 B_FM.CB -----KI--M--S-----T-----G-----A--D--N-----
 B_FR.CC I-----D-----Q--V-----D-----A--D-----
 B_FR.CO -----AM-----K--E--K-----P--K-----A--D-----
 B_FR.J91 -----D--S-----T-----P-----A-----
 B_FM.LAI See sequences with inserts
 B_FR2.FRMP6 E...PKASH--G--H-----A--T-----TD--R--P-----A--D--N--L-----
 B_FR2.FRMP7 -----N-----T-----P-----I--K--H--D--R-----Y-----
 B_FR2.FRMP27 IM--A--S-----K--P-----I-----R--S-----V-----Q-----Y-----
 B_FR2.FRMP29 I-----I--T-----K--PI-----I-----R--S-----V-----A--D-----Y-----
 B_FR2.FRMP32 -----I--K--I-----T-----R--P-----S-----G--H--M-----A--D-----
 B_FR2.FRMP36 -----N-----T-----P-----S-----G--H--M-----A--D-----
 B_FR2.FRMP45 #...A--L-----I-----P--Q-----Q--PM-----A--D-----K-----
 B_FR2.FRMP47 -----AK--ISN--Q--H-----T-----L--S-----A--D-----
 B_FR2.FRMP51 -----SN-----T--P--Q-----R--PM-----V-----R-----
 B_FR2.FRMP52 -----I-----I-----P--Q-----Q-----C--G--V--L--K-----
 B_FR2.FRMP53 -----I-----N-----IV-----S--Q--T-----E--A--N-----Y-----
 B_FR2.FRMP54 #...S--I-----P--Q-----C--M-----H--A--D--V--N-----R-----
 B_FR2.FRMP61 -----N-----T-----P-----S-----M--M-----A--D-----
 B_FR2.FRMP77 T-----A-----I-----P--Q-----G--P-----A--D-----
 B_FR3.FRMP78 See sequences with inserts
 BI--TQ--TR--SQLT--T--DKC--N--LI--H-----
 I--Q--A--K--MKV--T-----E-----I--K-----
 I--EQ--L--D-----K-----I--H-----
 I--Q--Q--Q--A-----V-----P--X-----
 S--I--K--R--T--Q--D-----G-----IX-----
 I--EIA--K--QEVG--N--H--F-----S--APX-----
 A--TIQ--D--QE--AI--G-----T-----TH-----
 I--STQ--KR--LT--E-----K-----I--M--V-----
 T--S--G--R--E-----YSL--K-----
 T--NKT--EQ--RE--AI--Q--K-----X-----
 T--D--A--T-----N--T--P-----
 T--GR--C--EC--R--MAI--K--CB--T-----P--X-----
 D-----B--Q--AV--EA--H--RT-----TP-----
 T--NDE--DG--RKVAT--G--E-----H--N-----

		V3 LOOP			
B_CONSENSUS_97	VVIRSEDTTQNTIIVQAESEVHN	CTHMAWTRRESI.HIGPGNGAFYTTG.EIIG.DILQANC	MLSRKMMHT.LKQIVKKLREBQFGN.NKT..T.IVPKQSTGQZUPEIVNHISF.NCGG.EKFFVCNT		
ALL_CONSENSUS	II-----I-N-----	-----R-----Q-----A-----D-----	-----I-----TE-----Q-----VA-----H-----	-----I-----AP-----L-----TT-----	-----
B_FR1.FRMP93	-----N-----	-----Q-----	-----N-----R-----I-----	-----N-----	-----G-----
B_FR1.FRMP122	-----S-----I-----	-----R-----S-----L-----YDRV-----M-----R-----K-----Y-----	-----I-----EE-----K-----S-----Q-----K-----	-----I-----K-----	-----
B_FR2.FRMP159	-----T-----T-----K-----	-----G-----	-----D-----	-----T-----	-----E-----A-----LR-----
B_FR2.FRMP291	-----N-----K-----Q-----	-----R-----S-----	-----RD-----D-----G-----V-----	-----	-----X-----
B_FR2.FRMP329	-----V-----N-----I-----KT-----V-----	-----L-----	-----I-----SU-----T-----QRVAA-----EG-----	-----	-----H-----X-----
B_GA.OYI	-----I-----S-----N-----I-----K-----	-----NR-----S-----H-----K-----Q-----	-----T-----EK-----S-----AT-----	-----K-----R-----	-----A-----DR-----
B_GB.AIT	-----D-----I-----K-----	-----E-----TL-----A-----E-----Y-----	-----TQ-----RD-----I-----	-----K-----	-----BLC-----
B_GB.CAN	-----N-----K-----F-----	-----L-----A-----TV-----A-----D-----R-----	-----ST-----	-----T-----K-----	-----I-----
B_GB.Man	-----L-----D-----H-----	-----S-----S-----V-----R-----H-----V-----R-----AVT-----	-----I-----K-----EK-----	-----E-----K-----	-----I-----T-----S-----
B_GB.V12	-----M-----	-----SK-----I-----B-----S-----	-----T-----MGT-----D-----L-----A-----	-----	-----C-----KP-----
B_GB.V74	-----N-----V-----	-----RG-----	-----I-----T-----D-----	-----T-----	-----S-----
B_GB.V77	-----S-----	-----S-----	-----I-----D-----	-----S-----	-----E-----
B_GB.V82	-----KDP-----K-----	-----R-----Y-----V-----B-----Q-----N-----	-----I-----S-----	-----I-----B-----	-----R-----
B_GB.V9	-----K-----	-----G-----P-----I-----A-----S-----Q-----X-----	-----I-----K-----E-----R-----T-----R-----K-----	-----I-----	-----
B_GB1.CPML	See sequences with inserts	-----I-----	-----I-----EE-----R-----QRVA-----	-----K-----E-----	-----A-----T-----P-----A-----
B_GB2.CPML2	-----V-----Y-----K-----	-----I-----	-----I-----EE-----R-----QRVA-----	-----K-----E-----	-----A-----T-----P-----A-----
B_GB3.4995	See sequences with inserts	-----I-----	-----I-----EE-----R-----QRVA-----	-----K-----E-----	-----A-----T-----P-----A-----
B_GB3.CPML7	See sequences with inserts	-----I-----	-----I-----EE-----R-----QRVA-----	-----K-----E-----	-----A-----T-----P-----A-----
B_GB4.AC01	-----N-----T-----S-----	-----GV-----L-----S-----W-----A-----Q-----	-----I-----V-----	-----EE-----I-----K-----	-----KP-----T-----A-----F-----ES-----
B_GB4.JB14	-----D-----S-----N-----K-----	-----I-----G-----PH-----	-----I-----Q-----	-----I-----Q-----	-----M-----S-----
B_GB4.M2347015	I-----SN-----TKP-----V-----	-----Y-----A-----Q-----	-----I-----KTD-----	-----E-----E-----	-----A-----B-----
B_GB4.M2886401	-----N-----K-----	-----K-----RR-----F-----V-----	-----I-----	-----H-----V-----N-----	-----R-----KN-----T-----M-----R-----
B_GB4.M3015617	-----A-----I-----V-----	-----R-----SH-----VY-----	-----I-----ED-----X-----	-----V-----	-----T-----
B_GB4.M73767719	-----Q-----S-----K-----Q-----	-----S-----E-----LF-----D-----N-----	-----TQ-----	-----AI-----S-----R-----	-----L-----L-----T-----S-----
B_GB4.M73768520	-----L-----N-----K-----A-----	-----S-----P-----H-----A-----	-----TIN-----	-----Q-----G-----V-----	-----H-----S-----
B_GB4.M829	-----N-----I-----H-----	-----G-----R-----T-----H-----A-----R-----R-----	-----I-----Q-----	-----AT-----E-----	-----V-----L-----I-----
B_GB5.4654	See sequences with inserts	-----X-----	-----I-----V-----	-----I-----K-----X-----	-----A-----R-----KP-----
B_GB5.4657	See sequences with inserts	-----G-----SXG-----	-----S-----D-----	-----R-----AX-----	-----A-----
B_GB5.4658	-----S-----Y-----T-----	-----A-----G-----K-----X-----X-----	-----D-----D-----Q-----Y-----K-----K-----	-----K-----	-----
B_GB5.4659	-----A-----I-----Q-----	-----T-----T-----D-----	-----D-----S-----	-----GR-----E-----K-----E-----D-----	-----K-----
B_GB5.4660	I-----X-----SN-----K-----P-----Q-----	-----X-----Q-----TH-----Q-----Y-----	-----I-----V-----	-----GR-----E-----	-----R-----
B_GB5.4661	-----V-----N-----I-----T-----Q-----	-----B-----S-----V-----G-----N-----	-----I-----Q-----	-----G-----K-----	-----H-----
B_GB5.4662	I-----I-----N-----S-----L-----H-----	-----G-----V-----A-----DY-----	-----I-----Q-----	-----T-----X-----	-----
B_GB5.4664	-----D-----LN-----T-----R-----N-----T-----	-----V-----N-----D-----	-----I-----KH-----D-----QR-----I-----Q-----K-----	-----I-----XP-----	-----X-----
B_GB5.4665	-----D-----N-----V-----H-----K-----D-----	-----I-----P-----F-----D-----G-----	-----I-----KRG-----	-----Q-----VA-----ST-----B-----	-----
B_GB5.4666	-----V-----A-----Q-----	-----TK-----VY-----A-----Q-----K-----	-----I-----Q-----	-----Q-----AH-----ST-----	-----
B_GH.Q46	-----I-----	-----N-----W-----A-----H-----	-----IV-----D-----R-----T-----	-----S-----L-----V-----T-----	-----
B_HT1.R7	-----I-----	-----P-----W-----A-----D-----	-----Q-----B-----KA-----L-----T-----	-----S-----L-----	-----
B_HT1.D1HA65	-----I-----	-----P-----W-----A-----D-----	-----I-----N-----TD-----NR-----AT-----	-----KY-----K-----	-----KTH-----A-----T-----
B_HT1.D1HA652	-----I-----	-----P-----W-----A-----D-----	-----I-----N-----TD-----NR-----AT-----	-----KY-----K-----	-----KTH-----A-----T-----
B_HT1.D2HA596	-----I-----	-----P-----W-----A-----D-----	-----I-----N-----TD-----NR-----AT-----	-----KY-----K-----	-----KTH-----A-----T-----
B_HT1.D2HA593	-----I-----	-----P-----W-----A-----D-----	-----I-----N-----TD-----NR-----AT-----	-----KY-----K-----	-----KTH-----A-----T-----
B_HT1.D2HA594	-----I-----	-----P-----W-----A-----D-----	-----I-----N-----TD-----NR-----AT-----	-----KY-----K-----	-----KTH-----A-----T-----
B_HT1.D2HA596	-----I-----	-----P-----W-----A-----D-----	-----I-----N-----TD-----NR-----AT-----	-----KY-----K-----	-----KTH-----A-----T-----
B_HT1.D2HA599	See sequences with inserts	-----I-----	-----I-----N-----TD-----NR-----AT-----	-----KY-----K-----	-----KTH-----A-----T-----
B_HT2.H13954	-----G-----K-----TQ-----H-----Q-----I-----	-----S-----R-----S-----R-----A-----K-----N-----	-----K-----S-----T-----A-----R-----T-----G-----LRR-----T-----	-----A-----QPP-----	-----
B_HT2.H13958	-----N-----I-----T-----	-----R-----S-----S-----R-----A-----K-----N-----	-----I-----T-----S-----	-----K-----A-----	-----QP-----
B_HT2.H13960	-----N-----I-----K-----V-----	-----R-----S-----S-----R-----A-----K-----N-----	-----STG-----	-----G-----E-----	-----R-----P-----L-----
B_HT2.H13962	-----N-----I-----K-----V-----	-----R-----S-----S-----R-----A-----K-----N-----	-----STQ-----	-----G-----E-----	-----R-----P-----L-----
B_HT2.H13968	See sequences with inserts	-----I-----	-----I-----	-----N-----	-----KP-----A-----T-----Q-----
B_HT2.H5986	-----K-----	-----S-----PM-----K-----D-----Y-----	-----I-----G-----	-----R-----	-----T-----I-----S-----T-----
B_HT2.H5988	-----G-----H-----	-----T-----F-----D-----	-----I-----Q-----	-----H-----E-----	-----D-----
B_HT2.H5990	-----N-----Q-----	-----M-----A-----N-----B-----	-----I-----T-----E-----R-----A-----D-----	-----T-----	-----
B_HT2.H5992	-----T-----	-----KR-----M-----D-----AY-----Q-----	-----V-----I-----E-----R-----	-----R-----	-----T-----
B_HT2.H5994	-----S-----KDP-----Q-----	-----M-----K-----V-----	-----E-----A-----T-----T-----	-----A-----KN-----	-----T-----
B_HT2.H5996	-----N-----I-----	-----G-----G-----A-----G-----	-----I-----	-----NK-----	-----TI-----D-----KP-----A-----T-----Q-----

B_CONSENSUS_97 VVIRGENTONAKTIIVQNESEVIN CTAPSPNTRKSI.NIGPGRGAFYTTG.EIIG.DILROANC MLSRAKAMIT.LKQIVKALLREDFON.MKT..T.IVPNQISICOOB#STVMBSP.NCOG.BFFYON?

ALL_CONSENSUS II--I-N-----R--Q--A--D-----I-TE--Q-VA-----N-----I-AP-----L--TT-----

B_HT2.H5998 See sequences with inserts -----S-----D-----I-T-D--V-D--K-----N-----Y-----

B_HT2.H6000 KKP-Q-----D-----I-T--R-V-D--K--K-----N-----Y-----

B_HT2.H6002 K-P-Q-----D-----K-D-KL--HKV-----G--R--H--Q--N-----I-

B_HT2.H6004 L--E-T--N--V--D--FM--K--A--D--N--K-D-KL--HKV-----G--R--H--Q--N-----I-

B_HT2.H6006 See sequences with inserts -----S-----D-----I-ETD--IK--T--Q--I-S--S--T--R--

B_HT2.H6008 See sequences with inserts -----S-----D-----M-TQ-DG--T--R--I-S--T-----

B_HT2.H6010 S--L--H--Q-H--I-----P-----A--D--K--I-ETD--IK--T--Q--I-S--S--T--R--

B_HT2.H6012 I--A--V--D-----P-----A--D--Y--M-TQ-DG--T--R--I-S--T-----

B_HT2.H6014 N--I--V--D-----D-----Q--N--R-E-A--GK--AE-----

B_HT2.H6016 N--H--Q--N--S--A--D--N--T-M-T--S-----L--N--

B_HT2.H6018 P--V--P-----P-----D-----I-EE--K--QK--T--G--R--D-----K--

B_HT2.H6020 G--QW--KAP--R-----A--D-----I-UQ--S--Q-LAA--G-----

B_HT2.H6022 SHT--S--G--G-----S-I-V--L-T-I--N-T-----

B_HT2.H6024 G-L-G--T--S--S--A--D-----I--Q--GR-AR--V--R--N-----

B_ID1.1701 V--S-----KVP-Q--S-----A--D--R--TE--KKV-D--G-----

B_ID1.IN93304 I--R--T--SI--IR-G--V--Q--L--I-XSD--D--AD--A-----L-----

B_ID1.IN93305 TVP-Q-----NRR--S-----V.AAR.A-V--N--TE--KV-E--GSE-X--D--A--N-----

B_ID1.IN93307 I--N--K--Q-----NRR--S-----V.AAR.A-V--N--S--N-----I-----

B_ID1.IN945128 S--S--M--I--G--TL--VV--S--D-QK--R-AD--IK--B--K-P-----

B_ID1.IN94570 D--N--P--H--V--R--P-----G--VF--A--R--T-RTE--E-R--D-----T-----

B_ID1.IN94581 S--T--TV--Q--S--Q--Y--H--A--B--R--TE--QKV--S--QKE-X--D--H-----D--

B_ID1.IN94581 A--T-----S--GV--A--E--A--TQ--IR-----I-----

B_ID1.IN94581 A--T-----A--Q-----I--D--R--SI--S-----S-----

B_ID1.20016 See sequences with inserts -----G--R-----F.A--D-T-----T--E--I--Q-----

B_ID1.20023 See sequences with inserts -----G--R-----F.A--D-T-----T--E--I--Q-----

B_ID1.30005 See sequences with inserts -----G--R-----F.A--D-T-----T--E--I--Q-----

B_ID1.30008 See sequences with inserts -----G--R-----F.A--D-T-----T--E--I--Q-----

B_IT.FD I--A--S-----N-R--S-----V.AAR.T--V-----I--N-----AI--K-E-----I--N-----

B_IT.F042 N--B-----S-----N--K--T--D-----S--K-----AI--E-----R-----

B_IT.S41a A--S--T--S--N--K--A--S-----R--Q-----Q-----I--E-----R--AG--E-----G-----

B_IT.S41a2 I--S--T--K--A--S-----R--S--A--D--T-----I--E-----R--AD--G--V-----H--G-----

B_IT1.115 P--T-----A-----R--S--A--D--N-----I--E-----R--AD--G--V-----H--G-----

B_IT1.130 K--M--RS--Q--Q-----R--K-----N-----I--E-----R--AD--G--V-----H--G-----

B_IT1.136 N--T-----P-----P-----D-----I--E-----R--AD--G--V-----H--G-----

B_IT1.145 K--H--HES--A-----Q-----P-----D-----I--E-----R--AD--G--V-----H--G-----

B_IT1.190 KKKK--T-----T-----Q-----V--TV--VW--D-----I--E-----R--AD--G--V-----H--G-----

B_IT1.193 SN--V--K--I-----G-----V--TV--VW--D-----I--E-----R--AD--G--V-----H--G-----

B_IT1.196 I--K--T--V--K--I-----G-----V--TV--VW--D-----I--E-----R--AD--G--V-----H--G-----

B_IT1.199 A--L--I--NH--KV--G--XG-----F--T--K--Y--V--XK--DX--X--AI--G--K-----K-----

B_IT1.204 A--S-----T-----S-----E-----N-----EG--K--I-----N--T-----

B_IT2.13A I--N--N--A-----Q-----R--N-----A-----I--K-----D-----I-----H-----

B_IT2.178 T--X--K-----G-----G-----I--E-----G-----I-----I-----

B_IT2.328 N--N--A--A-----Q-----R--N-----D-----I--E-----G-----I-----I-----

B_IT3.A07 A--S--N--V-----K-----G--N--G--L--A-----K-----DVMGT-----AI-----R--TN-----V-----

B_IT3.AZ97 A--S--N--VIT-----K-----N-----W-----N-----Y-----BRSDK--E--I--G--KL--IE-----R--A-----

B_IT3.B047 K--S--V-----K-----N-----G-----Q-----K-----I--S-----T-----K--LVR-----K-----ER-----

B_IT3.B372 M--I-----S-----AE-----S-----T--ED--R--GHV--P--K-----T--RMH-----

B_IT3.BV76 N--T--S--H--R--P--Q-----S-----N--G-----A--D-----I-----D-----I-----K-----

B_IT3.CN26 SN--H--XV--V-----S-----G--R-----TV--AE--K-----I--TQ--E--B--N-----P-----SR-----T-----

B_IT3.DA123 I--H--I-----H-----T-----A--Q-----I--A--K-----N--AI-----A-----I-----

B_IT3.DXV105 A--H--I-----H-----T-----A--Q-----I--A--K-----N--AI-----A-----I-----

B_IT3.MF86 N--I-----K-----I-----P-----A-----S-----D-----ER--I-----I-----

B_IT3.MR39 A--I-----A-----I-----T-----V-----DIGI--G-----G-----V-----V-----S--P-----

B_IT3.PF38 A--I-----A-----I-----T-----V-----DIGI--G-----G-----V-----V-----S--P-----

B_IT3.TN39 V--T--K-----I--K-----RV--TN--VY-----I--E--K-----E--AM--K--E-----N-----DS-----

B_UP.ETP I--N-----IV-----I-----N-----T--S--R--AIE--K--M-----I--EM-----T-----T-----

B_UP.GDMA K--P--V-----SKT--RR-----R--Q--A--L-----IN--R--A-----Q--K--V-----R-----

V3 LOOP												
B_CONSENSUS_97	VVIRSEPTDIAKTIIVQNESVLD	CTPDDNTARKSY	NIGGEGAFYYTTC	ETIO	DLISQAMC	NLSRAKQNT	LQIVKLLERQFON	NET	T	IVFNSISGQGPETVWHSF	WCGG	EFFVQNT
ALL_CONSENSUS	II---I-M-----	---R---Q---A---D---	---R---Q---A---D---	---	---	---I---E---Q---VA---	---R---	---	---	---I---AP---	---	---
B_JP.JNH1H	---S---K-V-----	---FL---Q---W---S---Q---	---	---	---	---I---ST---	---TE---	---	---	---	---	---
B_JP.JP23A	---L---L---H---T---	---STWK---	RM.TL---H.V---	S---D-T---	---	---K-Y---	---	---	---	---	---	---
B_JP.KM03	---D---S-----Q-I---	---G---R-----V---AAE.K---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat19	---F---I-M---S-----Q---	---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat2	---K-T-Q-H-----	---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat20	---AM-----	---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat26	---M-----FA-V---	---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat47	---T---Q---I---	---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat63	---V-D-L-----K---	---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat65	---M---I---K---	---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat8	---RT-Q-H-----	---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat9	---D---S-----H---	---	---	---	---	---	---	---	---	---	---	---
B_JP1.Pat15	---I---R---A---	---	---	---	---	---	---	---	---	---	---	---
B_KP.Kr111	---K-SH-VD-----KD-V---	---	---	---	---	---	---	---	---	---	---	---
B_LT.LIT11A	---V---S---IQ---	---	---	---	---	---	---	---	---	---	---	---
B_LT.LIT12A	---N-----S---	---	---	---	---	---	---	---	---	---	---	---
B_LT.LIT18A	---V---S---IQ---	---	---	---	---	---	---	---	---	---	---	---
B_LT.LIT21A	---M-----	---	---	---	---	---	---	---	---	---	---	---
B_LT.LIT22A	---I-M---RV-----T-Q---	---	---	---	---	---	---	---	---	---	---	---
B_M01.M00003	---N---RV-----	---	---	---	---	---	---	---	---	---	---	---
B_M01.M00003	---RV-----	---	---	---	---	---	---	---	---	---	---	---
B_M02.1740	---RV-----D---	---	---	---	---	---	---	---	---	---	---	---
B_M02.1755	---RV-----	---	---	---	---	---	---	---	---	---	---	---
B_M02.1763	---RV-----	---	---	---	---	---	---	---	---	---	---	---
B_M02.1782	---N---RV-----	---	---	---	---	---	---	---	---	---	---	---
B_M02.M1739	---RV-----	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9214087	---V-I---A---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9214089	---V---A---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9214093	---V---A---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9214096	---V---S-A---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9315157	---V-T---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9315158	---V-Z---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9315160	---V-T---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9315168	---V-T---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9315171	---V---A---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9315173	---V---T---	---	---	---	---	---	---	---	---	---	---	---
B_MY1.9315174	---M-V-----V---	---	---	---	---	---	---	---	---	---	---	---
B_ML.114C	---K-T-Q-H-----	---	---	---	---	---	---	---	---	---	---	---
B_ML.168	---M---I---V-D---	---	---	---	---	---	---	---	---	---	---	---
B_ML.X1com	See sequences with inserts											
B_ML.h159	---N-----IA---	---	---	---	---	---	---	---	---	---	---	---
B_ML.Wolfscon	---D---I-----Y---	---	---	---	---	---	---	---	---	---	---	---
B_ML1.A1	---D-I-----K-A-Q---	---	---	---	---	---	---	---	---	---	---	---
B_ML1.A2	---M---M---K-P-V---	---	---	---	---	---	---	---	---	---	---	---
B_ML1.A3	---S-----	---	---	---	---	---	---	---	---	---	---	---
B_ML1.A4	---K-----	---	---	---	---	---	---	---	---	---	---	---
B_ML1.A5	---N---V-----	---	---	---	---	---	---	---	---	---	---	---
B_ML1.A7	---M---T---	---	---	---	---	---	---	---	---	---	---	---
B_ML1.A9	---	---	---	---	---	---	---	---	---	---	---	---
B_ML1.A10	---	---	---	---	---	---	---	---	---	---	---	---
B_ML1.A13	---I-A-M-----K---	---	---	---	---	---	---	---	---	---	---	---
B_ML2.Wolfsch495	---D---I-----KKA-Q-I---	---	---	---	---	---	---	---	---	---	---	---
B_ML2.Wolfsch	---T---K-----	---	---	---	---	---	---	---	---	---	---	---
B_ML3.NET1	---M---K-P-R---	---	---	---	---	---	---	---	---	---	---	---
B_ML3.NET2	---D-S-----KDP-N-T---	---	---	---	---	---	---	---	---	---	---	---
B_ML3.NET3	---L---V-----	---	---	---	---	---	---	---	---	---	---	---
B_ML3.NET4	---SN-----K-P-K---	---	---	---	---	---	---	---	---	---	---	---
B_ML3.NET5	---M---S-----	---	---	---	---	---	---	---	---	---	---	---

V3 LOOP

```

B_CONSENSUS_97 VVIRSEMTDIAKTIIVQLNESVELN CTPFMDTRKSEI.HIOPGRGAPVYVITG.EIIG.DILNQANC MLRRAMMWT.LEQIVKLLERDQGN.NKT..T.IYHQSISSGDDPTVMHSF.HCGG.EFFYCHT
B_NL4.863_84 II---I-N-----R---Q---A---D---N-----I--TE---Q-VA---H---I-AP---L--TT---
B_NL4.861_84 ---D---I---K-P-----Q-----D-----SS-YD---R-VA-----
B_NL4.8A206_84 ---D---K-P-V-----G-----D---E-V-X---B---
B_NL4.8A65_85 ---D---KAP-V-----A-----TO-PP---Q-AI---K---E---
B_NL4.8N136_85 ---D---KAP-V-----A-----K-E---E---V-----F---
B_NL4.8N132_84 ---V-A-----G-----RM.TL---K.V-----I-K-R---E---AT---V-----
B_NL4.8N142_84 ---N-I-----I-----E-D-A---E-DG---S-----
B_NL4.8N152_84 ---D-N-N---K-----S-----T-D-I---B-A---M---R---H---
B_NL4.8N16_84 ---D-S---A-----G-----I-E---M-Q---I---E---R---KR---
B_NL4.8N28_85 ---D---R---A-V-----D-----I-K---D---Q---Q---V-----
B_NL4.8N90_85 I---N---I-----I-----INST-D---T---R-----
B_NL4.8N93_85 ---N---V-N-----I---E---D---E-Y-Y---K-----I-P---
B_NL4.8N1091_90 M---N-----IR-D---I---R---P-L-S---D---Y---TIM---D---G---D---P---
B_NL4.8N112_90 ---N---V-----M-K-T---D---I-K-E---R---T---K-----R---
B_NL4.8N135_90 ---S-----XA-----P-----S---I---D---E---AT---R---I---P-A-N---
B_NL4.8N153_82 ---A---N---I---K-A-R-----P-----A-D---S---I---D---E---AT---R---I---P-A-N---
B_NL4.8N164_90 ---D-N---KTP-M-T---I-Q---P-----A-D---S---I---D---E---AT---R---I---P-A-N---
B_NL4.8N169_85 ---D-N---G-----V-A-Q-----G-B---I.ANT.R-----I-GE---T---R---
B_NL4.8N17_82 ---D-I---NSG---V-A-Q-----P-----K-A---D---I---VD-ED---AK---R---
B_NL4.8N18_85 ---D-N---K-A-Q-----G---K---A-E-A---K-T---V---A---Q-L-T---R---
B_NL4.8N226_90 ---D-N---I---K-A-----G---X-----Q---E-AR---E---I---
B_NL4.8N23_82 ---F-D-IS---K-A-----N-----GL---I-T---D---AT---K-A---K-E---M---
B_NL4.8N29_85 ---A-L---K-A-----G---DG---R---R---I---E-D-G---E-AR---R---I---
B_NL4.8N39_82 ---D-N---K-----D---G---R---R---I---E-D-G---E-AR---R---I---
B_NL4.8N35_80 ---D-N---K-----D---G---R---R---I---E-D-G---E-AR---R---I---
B_NL4.8N36_82 ---D-N---K-----D---G---R---R---I---E-D-G---E-AR---R---I---
B_NL4.8N38_85 ---D-N---K-----D---G---R---R---I---E-D-G---E-AR---R---I---
B_NL4.8N3_90 ---S---Q-----RG---Q---N---I-T---D---AT---K-A---K-E---M---
B_NL4.8N50_83 I---N---T---K-D---IS---G---A---D---N---IT---D---AT---K-A---K-E---M---
B_NL4.8N57_85 ---D---K-----IS---G---A---D---N---IT---D---AT---K-A---K-E---M---
B_NL4.8N66_82 ---L---N-V---K-P-----L-X-V-----I-E-D---E-AR---R---I---
B_NL4.8N90_85 ---D---I---F-----P-----I---Q---Q---E---R---D---I-S---
B_NL4.8N96_88 ---A-L---I---Q-----K---ED---Q---T---G---E---I-K---
B_NL4.8N99_85 ---D---H---R-----A---D---I---N---H---M---R---L-Q-V---A---
B_NL4.8T34_81 ---D---TE---K-----N---A---M-V-T---L-Q-V---A---
B_NL4.8T48_90 ---D---D---S---I---R---R---E---S-V---H---
B_NL4.8T020_86 ---A---I---T-----R---P---A---N---I-G---E---RLAR---R---Q-P---
B_NL4.8T025_89 ---A---T-I---Q-----N---R---S---D---D---I---K---I---
B_NL4.8T035_87 ---D---V-----P---P---D---S---D---D---I---K---I---
B_NL4.8T057_87 ---D---T-----R---P---D---S---D---D---I---K---I---
B_NL4.8T083_87 ---D-N---G-V-----I-G---A---D---I---Q---Q---AN---V---
B_NL4.8T164_90 ---D---S-GT---T-----A---NT---A---D---Q---D---I---Q---Q---AN---V---
B_NL4.8T117_90 ---A---T-S---K-----S---G-Q---Q---T---D---D---I---Q---Q---AN---V---
B_NL4.8T138_89 ---D-N---I-M---K-----G---Q---A---Q-T-N---I---E---R---H---G---K---P---
B_NL4.8T141_91 ---A---T-S---K-----A---T---K---I---D---D---T---Q---K---D---V---R---
B_NL4.8T21_90 ---D---I---AP-----D---I---F---B---D---R---I---G---K---R---
B_NL4.8T57_91 ---D---N---N-----T-H-N---V---T---RG---V---N---IT-KB-B---T-V-N---K---T---
B_NL4.8T007_86 ---D---N---N-----T-H-N---V---T---RG---V---N---IT-KB-B---T-V-N---K---T---
B_NL4.8T008_88 ---D---N---N-----T-H-N---V---T---RG---V---N---IT-KB-B---T-V-N---K---T---
B_NL4.8T048_89 ---D---N---N-----T-H-N---V---T---RG---V---N---IT-KB-B---T-V-N---K---T---
B_NL4.8T057_91 ---D---N---N-----T-H-N---V---T---RG---V---N---IT-KB-B---T-V-N---K---T---

```

V3 LOOP									
B_CONSENSUS_97	VVIBSENTIRAKTIVOLAESVEIN	CTFPMNNTANKSI	HIGPGHAFYTTG	KIIG	DILQAMC	NLSRAKMMT	LKQIVKELLERBOFGN	NKT	T
ALL_CONSENSUS	II--I-N	---R--D	A--D	---	---	I--TE--Q-VA	---H	---	I-AP--L--T
B_ML4.I1064_90	---A--S--I	---T	---	---	---	I--DD--RR--I	---	---	---
B_ML4.I1072_90	---M--S--T--G	---V	---	---	---	I--O--Y	---	---	A-SP
B_ML4.I1077_91	---A--M--I	---P	---	---	---	I--D--EM--I	---	---	A--
B_ML4.I1084_90	---M--I	---	---	---	---	I--D--RV--V	---	---	---
B_ML4.I131_90	---A--M--I	---V	---	---	---	I--Q--I	---	---	---
B_ML4.I1404_88	---A--S--I	---V	---	---	---	I--D--R--I	---	---	---
B_ML4.I1408_86	---D--I	---Q	---	---	---	I--EM--E--K--I	---	---	K--
B_ML4.I15020_89	---TA--	---	---	---	---	I--D--DA--V--I	---	---	D--
B_ML4.I15032_87	---L--M--A	---	---	---	---	I--T--B--I	---	---	I-KH
B_ML4.I15037_89	---N--S--T--S	---V	---	---	---	I--T--A--R--I	---	---	K--R
B_ML4.I15038_89	---M--I	---	---	---	---	I--D--K--I	---	---	A-SP
B_ML4.I17028_91	---A--T--K--I	---	---	---	---	I--D--BA--V--H	---	---	E--
B_ML4.I170_88	---A--EN--V	---	---	---	---	I--D--R--I	---	---	D--H
B_ML4.I190_89	---D--N--I--M	---	---	---	---	I--D--R--I	---	---	E--
B_ML4.K999_86	---N--RDA--V	---	---	---	---	I--B--B--E--K--	---	---	---
B_ML5.pt10con	---N--K--N	---	---	---	---	I--ST--S--V--D	---	---	H--
B_ML5.pt12con	See sequences with inserts	---	---	---	---	---	---	---	---
B_ML5.pt13con	---D--N--I--N--Q	---	---	---	---	I--T--D--K--I	---	---	K--T
B_ML5.pt14con	II--I--N--K--I	---	---	---	---	K--TE--R--T	---	---	K--
B_ML5.pt15con	---I--T--H--K	---	---	---	---	I--G--E--K--K	---	---	T--
B_ML5.pt16con	---N--K--K	---	---	---	---	I--G--E--K--K	---	---	K--
B_ML5.pt17con	---N--K--IV	---	---	---	---	I--V--D--T--D	---	---	---
B_ML5.pt19con	---N--I	---	---	---	---	I--KB--D--I	---	---	R--
B_ML5.pt2con	---D--N	---	---	---	---	I--DY--Q--I	---	---	K--
B_ML5.pt3con	---N--R--P--K--H	---	---	---	---	I--KT--K--E	---	---	E--
B_ML5.pt3con	---I--Q--N--IQ	---	---	---	---	I--Q--T--K--Q	---	---	---
B_ML5.pt4con	---D--S--K--T	---	---	---	---	I--D--M--K--K	---	---	---
B_ML5.pt5con	---A--N--I--K--T	---	---	---	---	I--T--Q--A--K--	---	---	---
B_ML5.pt6con	---I--I--N--A	---	---	---	---	I--KE--E--Q--I	---	---	K--
B_ML5.pt7con	See sequences with inserts	---	---	---	---	---	---	---	---
B_ML5.pt8con	---S--N--K--H	---	---	---	---	---	---	---	---
B_ML5.pt9con	---D--N--V	---	---	---	---	---	---	---	---
B_ML5.16	---A--SN--K--A	---	---	---	---	---	---	---	---
B_ML5.320	---D--S--T--V	---	---	---	---	---	---	---	---
B_ML7.ME1058con	---A--R--V	---	---	---	---	---	---	---	---
B_ML7.ME537con	---D--TI	---	---	---	---	---	---	---	---
B_ML8.672	---I--L--A	---	---	---	---	---	---	---	---
B_ML9.H239	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H1054	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H1081	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H1082	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H1110	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H1135	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H1140	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H1145	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H1159	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H179	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H411	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H412	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H424	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H536	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H569	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H57	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H571	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H709	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML9.H716	---I--N--H--I	---	---	---	---	---	---	---	---
B_ML10.BR8914	---I--N--H--I	---	---	---	---	---	---	---	---

V3 LOOP

B_COMBINSUB_97	VVRISENTDNARTIIVQIMSEIVEIN	CTYPRNMHTRRRSI_HIGPGRGAFYTTG_EIIG_DILQAMC	MUSNAKMENT_LQIVVOLLEREPQFN_HRT_T_IVVMQSLGGQDPEIVHHSF_NCOG_RFFYCHT
ALL_CONSENSUS	11--E-N-----	---R---Q---A---D---D---	7--TE---Q-VA---H---
B_NIL10_QF921953	---N-----T-----	---V-P-----B---	---T-AA---MKT-S---K---
B_NIL10_GG93013	L-D-----X-----	---S---P-----D---	1-GT---DA---AQ-----Y---
B_NIL10_MAR136	---K---T-----KP---	---N-----D---	1--E-D---AC-----V---
B_NIL10_SRS9038	3--A-LSN-----K-P-Q---	---R-P---K-V---D---	---T-B---G-----
B_NIL10_SRS9115	---K-----S-----V---	---R-PN---V-A-D---	1-T---A---I---G---
B_NIL10_SRS9233	1--K-I---K-FI---	---RM---V---Q---	---IB-D---AD---KV---
B_NIL10_SRS925752	---N---V-----V---	---N---V---Q---	1-E---V-T---K-K---
B_NIL10_SRS9269	---D-SN---N---T---	---N---V---Q---	---IN-T---E-AR---K-N_KT---
B_NIL10_SRS93047	---V---TG-N---V---	---R-F---A-R-A---D---	1-G-Q---H-HR---
B_NIL10_UHS93272	---V---TG-N---V---	---R-F---A-R-A---D---	---Q-K---S-AH---K---
B_NIL10_UHS94030	---V---TG-N---V---	---R-F---A-R-A---D---	---E-B---V-T---A-R---R---
B_NIL10_UHS94038	---V---TG-N---V---	---R-F---A-R-A---D---	---E-BK---E-I---G---
B_NIL11_gatPecm	---V---TG-N---V---	---R-F---A-R-A---D---	1-TE---E-T---K---
B_NIL11_patnNcm	---V---TG-N---V---	---R-F---A-R-A---D---	---R-AG---K---
B_NIL13_000H	---D---I---K-P---	---X---N---	---T-D---AR---
B_NIL13_0157	---D---I---K-P---	---X---N---	---Y-ST---DX-R-R---
B_NIL13_0211	---D---I---K-P---	---X---N---	---ST-D---
B_NIL13_0230	---D---I---K-P---	---X---N---	---K---A-E-A---U-E---
B_NIL13_0337	---D---I---K-P---	---X---N---	1--N-T-A-Q---E-K---
B_NIL13_0583	---D---I---K-P---	---X---N---	---SL-D---R-V---
B_NIL13_0617	---D---I---K-P---	---X---N---	---T-D---E---
B_NIL13_0617	---D---I---K-P---	---X---N---	---DN-D---R-I---K-T---
B_NIL13_0617	---D---I---K-P---	---X---N---	---TG---T---G---
B_NIL13_0617	---D---I---K-P---	---X---N---	---GIN-M---T---G---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-K---Q-T---K-K---
B_NIL13_0617	---D---I---K-P---	---X---N---	---ST---R-AI---K---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-E-S---K-AI---V-T---
B_NIL13_0617	---D---I---K-P---	---X---N---	---I-V---Q-A---K-E-K---
B_NIL13_0617	---D---I---K-P---	---X---N---	---Y---AI---K---
B_NIL13_0617	---D---I---K-P---	---X---N---	---ER-D---I---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-MKT---Q-R-V-E---K-E---
B_NIL13_0617	---D---I---K-P---	---X---N---	---ED---K-AI---
B_NIL13_0617	---D---I---K-P---	---X---N---	---Q-E---R-I---Q-E---
B_NIL13_0617	---D---I---K-P---	---X---N---	---I-M-T---R-AI---
B_NIL13_0617	---D---I---K-P---	---X---N---	---GT-D---R-I---Q-V---
B_NIL13_0617	---D---I---K-P---	---X---N---	---Y---AI---K---
B_NIL13_0617	---D---I---K-P---	---X---N---	---I-Q---I-R---K---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-N-T---AI---G-A---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-T---S-AI---K---
B_NIL13_0617	---D---I---K-P---	---X---N---	---TC---T---X-AI---R-N---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-E---T---E---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-GT---NL---
B_NIL13_0617	---D---I---K-P---	---X---N---	---Q-K---Q-T---P---
B_NIL13_0617	---D---I---K-P---	---X---N---	---ST---T---K---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-R-T---R-AI---G-V---
B_NIL13_0617	---D---I---K-P---	---X---N---	---D---S-T---A-N---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-M-T---AI---E-A-R---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-N-T---AI---G-K---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-N-T---EK-I---K-T---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-A-E---Q-T---D-S---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-E---AI---P---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-S-K---G-A-T---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-E---AI---Q-K---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-E---AI---E---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-E---S-H-I---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-S---GN-AI---
B_NIL13_0617	---D---I---K-P---	---X---N---	1-A-E---L---

V3 LOOP

B_CONSENSUS_97 VVIRSEMTFNTAHTIVQLNESVEIN CTRFQNTARKSI.NIGPGGAGFYVYTTG.EITG.DILQDANC NIGRAADMT.LKQIVKELLERDQFQIN.MCTF..T.IVFNQSIGGGDREIVNNSP.WGGG.EFFYCHT
 ALI_CONSENSUS98 IT--I--N-----H-----Q-----A-----D-----Q-----V-----Y-----I-----G-----EK-----I-----R-----P-----F-----
 B_P01.D2PR733 -----H-----I-----K-----T-----A-----Q-----K-----Y-----V-----TQ-----Q-----ERVAE-----KEM-----P-----A-----A-----KL-----
 B_PV.12837 -----IS-----N-----D-----Q-----T-----Y-----Y-----KV.SL-----VN-----A-----Q-----K-----Y-----V-----TQ-----Q-----ERVAE-----KEM-----P-----A-----A-----KL-----
 B_PV.12838 -----T-----RTIT-----T-----V-----G-----R-----R-----TV-----R-----V-----T-----D-----G-----TS-----S-----
 B_PV.12839 -----S-----I-----T-----Q-----L-----M-----T-----F-----S-----K-----N-----ER-----K-----Y-----T-----K-----
 B_PV.1284 -----IS-----LF-----K-----L-----T-----F-----R-----A-----R-----K-----I-----ERQ.SDW-----ERVAE-----GSMVP-----T-----
 B_PV.12842 -----S-----T-----N-----KAA-----V-----G-----K-----Y-----D-----I-----S-----D-----VE-----K-----E-----
 B_PV.12845 -----K-----N-----K-----A-----V-----M-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_PV.12847 -----S-----A-----K-----M-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_PV.12849 -----IS-----H-----K-----TR-----QM-----VL-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_PV.1285 -----N-----V-----N-----KI.MRR-----A-----H-----R-----N-----V-----N-----Y-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_PV.1286 -----N-----KBPV-----A-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU1.RU33A -----L-----N-----V-----K-----M-----Q-----A-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU1.RU34A -----I-----D-----V-----IQ-----V-----G-----T-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU2.10120 -----SS-----T-----X-----Q-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU2.154 -----V-----X-----Q-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU2.190 -----A-----S-----T-----K-----G-----R-----K-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU2.20637 -----T-----K-----T-----S-----A-----P-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU2.3336 -----IQ-----A-----S-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU2.3818 -----V-----V-----P-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU2.4035 -----V-----V-----P-----A-----D-----N-----I-----S-----RA-----RRVAI-----Q-----P-----M-----
 B_RU2.4113 See sequences with inserts -----A-----G-----T-----A-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU2.42615B -----Y-----T-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU2.4439 -----YLTTSRS-----TK-----IQ-----V-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU2.4456 -----K-----V-----V-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU2.4714 -----V-----V-----V-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU2.5014 -----A-----S-----T-----K-----R-----G-----R-----K-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU2.5852 -----I-----V-----IQ-----G-----G-----G-----Q-----I-----LQ-----Q-----A-----T-----K-----
 B_RU2.999 -----G-----DO-----SS-----S-----IQ-----RG-----G-----G-----Q-----I-----LQ-----Q-----A-----T-----K-----
 B_RU3.RU10 -----L-----V-----TK-----IQ-----G-----G-----G-----Q-----I-----LQ-----Q-----A-----T-----K-----
 B_RU3.RU11 -----L-----V-----TK-----IQ-----G-----G-----G-----Q-----I-----LQ-----Q-----A-----T-----K-----
 B_RU3.RU12 -----N-----T-----K-----S-----K-----P-----L-----D-----T-----K-----TI-----T-----IK-----E-----K-----
 B_RU3.RU13 -----N-----T-----K-----S-----K-----P-----L-----D-----T-----K-----TI-----T-----IK-----E-----K-----
 B_RU3.RU14 -----V-----K-----P-----Q-----G-----R-----N-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU3.RU17 -----N-----L-----K-----S-----R-----G-----P-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU3.RU18 -----V-----K-----IQ-----S-----Q-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU3.RU27 -----L-----N-----V-----IQ-----S-----R-----RV-----S-----I-----A-----QUT-----EQ-----K-----E-----T-----K-----
 B_RU3.RU28 -----N-----V-----TIQ-----S-----R-----RV-----S-----I-----A-----QUT-----EQ-----K-----E-----T-----K-----
 B_RU3.RU3 -----V-----IQ-----G-----P-----G-----P-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU3.RU5 -----D-----K-----P-----V-----G-----G-----R-----N-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU3.RU6 -----S-----L-----K-----IQ-----I-----R-----N-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_RU3.RU9 -----V-----K-----IQ-----I-----R-----N-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE.p0112113 -----V-----H-----K-----V-----G-----N-----R-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE1.1032 -----A-----SS-----I-----K-----IQ-----I-----G-----T-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE1.1433 -----D-----I-----N-----T-----K-----S-----R-----N-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE1.1866 -----D-----I-----N-----T-----K-----S-----R-----N-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE1.2815 -----D-----I-----N-----T-----K-----S-----R-----N-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE1.295 -----D-----I-----N-----T-----K-----S-----R-----N-----A-----D-----M-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE1.93 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE1.930 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE2.R1 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE2.R2 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE2.R3 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE2.R4 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE2.R5 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE3.AJ2195 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE3.BJ562 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE3.CJ569 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----
 B_SE3.CJ1074 -----I-----T-----K-----P-----A-----A-----A-----N-----Y-----I-----TQ-----S-----SHVAA-----K-----K-----P-----T-----V-----P-----

	***	***	V3 LOOP	***	***	***	***
B_CONSENSUS_97	VVIRSENPTDPAKTIIVQLNESVEIM	CYRPMNNTTRKSEI	HIQPGRCAPTYTTPG	ELIG	DILNOAHNC	NLSRAQMMNT	LEQIVKKLIEREQCN
ALL_CONSENSUS	II	I	M	A	D	I	TE
B_SK3_EJ2090	---	---	---	---	---	---	---
B_SK1_BT811	---	---	---	---	---	---	---
B_SK1_BT912	---	---	---	---	---	---	---
B_SK1_BT815	---	---	---	---	---	---	---
B_SK1_BT818	---	---	---	---	---	---	---
B_SK1_BT820	---	---	---	---	---	---	---
B_SK1_BT822	---	---	---	---	---	---	---
B_SK1_BT823	---	---	---	---	---	---	---
B_SK1_BT828	---	---	---	---	---	---	---
B_SK1_BT829	---	---	---	---	---	---	---
B_TH_93TH067	---	---	---	---	---	---	---
B_TH_97238	---	---	---	---	---	---	---
B_TH_98174	---	---	---	---	---	---	---
B_TH1_7787	---	---	---	---	---	---	---
B_TH1_864	---	---	---	---	---	---	---
B_TH1_8643	---	---	---	---	---	---	---
B_TH1_8645	---	---	---	---	---	---	---
B_TH1_8647	---	---	---	---	---	---	---
B_TH1_8649	---	---	---	---	---	---	---
B_TH1_865	---	---	---	---	---	---	---
B_TH1_8653	---	---	---	---	---	---	---
B_TH1_8655	---	---	---	---	---	---	---
B_TH1_8659	---	---	---	---	---	---	---
B_TH2_CN237	---	---	---	---	---	---	---
B_TH2_78132	---	---	---	---	---	---	---
B_TH3_W2TW014	---	---	---	---	---	---	---
B_TH3_W2TW026	---	---	---	---	---	---	---
B_TH4_N752	---	---	---	---	---	---	---
B_TH4_N753	---	---	---	---	---	---	---
B_TH4_N754	---	---	---	---	---	---	---
B_TH4_N755	---	---	---	---	---	---	---
B_TH4_N756	---	---	---	---	---	---	---
B_TH4_N757	---	---	---	---	---	---	---
B_TH4_N758	---	---	---	---	---	---	---
B_TH4_N759	---	---	---	---	---	---	---
B_TH4_N760	---	---	---	---	---	---	---
B_TH4_N761	---	---	---	---	---	---	---
B_TH4_N762	---	---	---	---	---	---	---
B_TH4_N763	---	---	---	---	---	---	---
B_TH5_018	---	---	---	---	---	---	---
B_TH5_1948	---	---	---	---	---	---	---
B_TH5_2619	---	---	---	---	---	---	---
B_TH6_0258	---	---	---	---	---	---	---
B_TH6_0286	---	---	---	---	---	---	---
B_TH6_1611	---	---	---	---	---	---	---
B_TH6_1615	---	---	---	---	---	---	---
B_TH7_82	---	---	---	---	---	---	---
B_TH7_85	---	---	---	---	---	---	---
B_TH7_86	---	---	---	---	---	---	---
B_TH7_87	---	---	---	---	---	---	---
B_TH7_89	---	---	---	---	---	---	---
B_TH7_90	---	---	---	---	---	---	---
B_TH7_91	---	---	---	---	---	---	---
B_TH7_92	---	---	---	---	---	---	---
B_TH7_93	---	---	---	---	---	---	---
B_TH7_94	---	---	---	---	---	---	---
B_TH7_96	---	---	---	---	---	---	---
B_TH7_99	---	---	---	---	---	---	---
B_TH7_101	---	---	---	---	---	---	---

DEC 97

[illegible]

V3 LOOP												
B_CONSENSUS_97	WVIREZNTAKTIIVQINSEVIN	CTPPTNTARKSI	HIGPGRCATYTTG	ETIG	DILQAHG	MISRAKNDPT	LKQIKKLLRETOPH	NKT	T	IVFMQGISQGDPRIVHNSF	NCQG	EPFYCHT
ALL_CONSENSUS	II	I	N									
B_TN1.236	-P-L-S	KMP-K	-G-	-R-	M-L-	-Q-	-D-	-H-	-I-	AP	-L-	TT
B_TN1.237	-D-S-T	-K-R-				-D-						
B_TN1.257	-K-M	DT-K				-V-M-S	TL	A	A	N		
B_TN1.261	-K-V-N	K-I				TK	IRH	T	AR	-K-		
B_TN1.358		V-N				I	IR	TL	R	V		
B_TN1.366		I-M	K-T									
B_TN1.382			RKP									
B_TN1.384		A	K-T-N-T									
B_TN1.386			I									
B_TN1.391			K-T-H									
B_TN1.394		L-N-V	-Q-V-									
B_TN1.646			K-T									
B_TN1.84	See sequences with inserts											
B_TN1.90	See sequences with inserts											
B_TN1.91		V	-K-T-I									
B_US.ACP		D	I	K-R								
B_US.ADP39		A-S-T	K-PI									
B_US.ADA		S	M	K								
B_US.ALA												
B_US.BAL		A	A	V								
B_US.BCSG3		A	A	K-A								
B_US.BRVA		N	V									
B_US.BMB			T									
B_US.CDC42		M	V									
B_US.DIAK		D	N	P	KD-Q							
B_US.FASH		I	A	SN	A-V							
B_US.HDNR		K	S		V							
B_US.J71		M	H	K-LA								
B_US.JN		V	D	M	KT-V							
B_US.JRC5P		D			X							
B_US.MA		E	R	I	K-P							
B_US.MD					Q							
B_US.NTSCG		N	V									
B_US.P896		I			V							
B_US.R3S	See sequences with inserts											
B_US.SBA		D	M	K-A								
B_US.SB8		M		T								
B_US.SBC		D	M		Q							
B_US.SC		L	N	K-A	V							
B_US.SP128		D	N		K-V							
B_US.SP162		D	N		K							
B_US.SP2		D	N		A							
B_US.SP33		D	N	L	V							
B_US.TN100			V		TQ							
B_US.TN1000			I									
B_US.TN1002		I	N	I								
B_US.TN1003		D	I	T	K-P							
B_US.TN1005			N	I								
B_US.TN1006		D	A	S	M	KDP						
B_US.TN1007				G	K-V							
B_US.TN1009			N		IV							
B_US.WEAU		I		N	V	I						
B_US.WM			N		K-P-Q							
B_US.WKJ22		I			X							
B_US.WR27	See sequences with inserts											
B_US.YU		I		N	V							
B_US.YVINAMcon		I	SK	SN	V	M	T	EH				
B_US1.HC01			A									
B_US1.HC05			N	I	H	KT-N-T						
B_US1.HC06			A		T	I						

V3 LOOP

```

B_CONSENSUS_97 VVIRSEPTONARTITVQNESVEIN CTFPNDTRRSI HIGPGRGATYTTG EILQ DILQANC NLGRAGKNT LKQIVKGLLRBQFGN MNT . T EVFNQISGGDPEIVMHSF NCGG EPHYCHT
ALL_CONSENSUS6 II---I-N-----Q-V-----R---Q---A---D-----I-TE---Q-YA---H---I-AP---L-TT-----QY---
B_US4.ZhuPcr ---N-----Q-V-----R---Q---A---D-----CIQ---E-R---K---I-T---TT-----TT-----
B_US4.ZhuPcrV L---N-----N-----D---I-----TN---K-V---V---D-----Q---T-N-T---E-AI---KL---KM---T---P---D---
B_US5.pt5b1 ...-L-N-----T-T---S---R---A---V---B-Q---R---I-K-R---R-VAI---KL---KM---T---P---D---
B_US5.pt5b1 ...-K-A-N-T---E-E---N---A---D-----I-CT---M-I---QF---T---F---
B_US6.I01 ---I-E-T---K---G---L---G---A---D---SS---I-V-E---KH---A---
B_US6.I02 ---D---N---I---V---T---A---D---I-TQ---V-T---Q---A---
B_US6.I03 ---N-----R---M---A---I---I-CT---V---A---
B_US6.I04 I---A-S---I---H---A---RG---M---A---I---E-R---EK---I---R---
B_US6.I05 ---D-S---KDP-N-T---I---P---M---D---M---I-A---Q-G---K---
B_US6.I06 -I---D-S---K-I---P---M---D---M---I-G-E-A---I---K---
B_US6.I07 -V---N---L---R-A---L---S-V---A---Y---T-NGS---D---T---K---A-K---
B_US7.306 ---L-N---K-P---K---K---D---N---I---D---R---I---
B_US7.565 ---D-S---K-I---I---X---A---D---M---I-X---D---Q---V---
B_US7.657 ---D-S---K-I---I---X---A---D---M---I-X---D---Q---V---
B_US8.R1 ...-D-L---I---V---A---D---M---I-X---D---Q---V---
B_US8.R2 ---D-N---K---M---E---R---E---A---E---GE-A---E---R-E---
B_US9.S1 ---D-N---R---V---M---W---A---D---I-CT---N-KA---E---K---
B_US9.S2 ---D-N---R---V---M---W---A---D---I-CT---N-KA---E---K---
B_US9.S3 ---D-N---R---V---M---W---A---D---I-CT---N-KA---E---K---
B_US9.S4 ---I---R---K---I---G---FM---MP---A---D---K---I-ES---K---G---Y---K---K---
B_US10.HA145con .....E-H-V---K-P---G---FM---MP---A---D---K---I-E-V---G---E---
B_US10.SFBDon See sequences with inserts
B_US10.SFPBcon See sequences with inserts
B_US11.074 I---I-N---K-P---M---N---K---IMST---E---I---L---S---
B_US11.107 ---N---I---T---RG---M---A---K---I-CT---Q---L---T---T---
B_US11.306 See sequences with inserts
B_US11.333 ---N---I---V---V---TN---VN---V---K---TT---R-V-E---R-R---A---H---L---T---
B_US11.349 ---D---V---V---TN---VN---V---K---TT---R-V-E---R-R---A---H---L---T---
B_US11.419 ---I---A-I---L---SV---K-V---I-G---A---S---D-K---K---T---A---
B_US11.44 ---I---A-I---L---SV---K-V---I-G---A---S---D-K---K---T---A---
B_US11.550 ---N---I---P---P-A---D---I-N---T---M---KD---R---X---T---M---
B_US11.552 ---D-S---TP---I---P---P-A---D---I-N---T---M---KD---R---X---T---M---
B_US11.725 ---D-S---TP---I---P---P-A---D---I-N---T---M---KD---R---X---T---M---
B_US11.UK ---N---T---G---M---W---A---D---I---Q---A---
B_US12.CHB2A ---N---T---G---M---W---A---D---I---Q---A---
B_US12.CHIMON ---N---T---G---M---W---A---D---I---Q---A---
B_US12.LC1A ---A---A---K---S---KA---M---A---A---Y---TR-CTR---EKV-E-I---
B_US12.LC2A ---D-S---K-R---E---M---K---Q---D---TT---D---G---
B_US12.LC3A ---D-S---K-R---E---M---K---Q---D---TT---D---G---
B_US12.LC4A ---D-S---K-R---E---M---K---Q---D---TT---D---G---
B_US13.141con -I---S---I---S---Q---L---I---E-Q-T---I-KE---K---R---P---V-V---I-S---Y---
B_US13.144con ---D-S---V---K-A---S---Q---L---I---E-Q-T---I-E---G---T---
B_US13.145con ---S---N---T---I-G---R---LQ---SS-G---R---TIMPO---LV-D---N-K---R---A-T-FPP---V-Y---
B_US14.1 ---V-A-H---I---Q---S-I---T---I---A---G---R-Y---I-T---T---Q---K---K---
B_US14.2 ---V-V---R---R-A---S---S---T---I---A---G---R-Y---I-T---T---Q---K---K---
B_US14.3 ---V-V---R---R-A---S---S---T---I---A---G---R-Y---I-T---T---Q---K---K---
B_US14.4 I-L---S---I---P---A---D---VI-K-N-T---B---E---I---S---P---
B_US15.A ---N---S---I---P---A---D---I-E---M---R---I---E---
B_US15.B ---N---RI---H---NN---D---D---I-E---M---R---I---E---
B_US15.C I-V-L---I---H---M---S---PN---D---R---D---S---D---H-AI---K-E---
B_US15.D ---S-A-L---K-D-VN---S---PN---D---R---D---S---D---H-AI---K-E---
B_US15.E I-A-SE-D---R-T-N---I---R---P---A---D---E---R-VAI---K-E---I-K---
B_US15.F ---A-L---I---H---M---S---PN---D---R---D---S---D---H-AI---K-E---I-K---
B_US16.NVRT1 ---A---I---H---M---S---PN---D---R---D---S---D---H-AI---K-E---I-K---
B_US16.NVRT3 ---A---I---H---M---S---PN---D---R---D---S---D---H-AI---K-E---I-K---
B_US17.CB1 ---A-L---V---K-Y---G---M---A---V---I-CF---S---T---E-Y---I---
B_US17.CB2 ---N-V-N---K-Y---G---M---A---V---I-CF---S---T---E-Y---I---
B_US17.CB3 I---N-V-N---K-Y---G---M---A---V---I-CF---S---T---E-Y---I---
B_US17.CB4 ---D-S---KDP---I-G---A---Q-V---I-T---K---H-AT---K---I---S---Q-L-S---

```

V3 Region Alignment

B_CONSENSUS_97 VVIRSENYCAKTIIVQLNESVBN CTRFNNTARKSI.HIQFGAGFYTTO.BIG.DILQAQHC NLSRAKQNT.LKQIVKLLKCEQFON.NKT.T.IVFNDISGGDPEIVMHSF.NCGG.EFFYCHT
 ALL_CONSENSUS II--I-N-----R-----Q-----A-----D-----I-TE-----Q-VA-----H-----I-AP-----L-TT-----
 B_US28.3799a -----N-Q-----K-A-----H-----R-----N-----D-T-----I-----E-A-----I-----G-----R-T-----I-----
 B_US28.861a -----D-----K-A-----Q-----T-----G-----K-----I-----S-----Q-----N-----
 B_US28.863 -----I-K-----T-----H-----A-----S-----R-----T-----A-E-----I-----O-----B-----N-----
 B_US28.868 -----K-----K-----S-----R-----T-----A-E-----I-----ST-----S-----R-----K-----E-----
 B_US28.86aa -----D-SS-BMI-----KT-----A-E-----F-Y-----I-----TA-----D-----YK-----I-----R-----
 B_US28.86Ma -----A-L-----I-----R-----K-----M-----A-----P-----I-----DO-----BK-AI-----S-Y-----
 B_US28.873a -----V-----S-----G-----N-----T-----D-----R-----K-----S-Y-----
 B_US28.87aa See sequences with inserts -----R-----A-----I-KED-DK-----AR-----N-P-----K-----
 B_US28.87Ma -----D-L-----M-----A-----H-----I-----KED-DK-----AR-----N-P-----K-----
 B_US28.880a -----I-M-----A-V-----
 B_US28.896a See sequences with inserts -----K-----B-----K-----H-----M-----A-----I-----T-----E-----DK-----I-----
 B_US28.89Ma -----YN-PK-----I-----K-----B-----K-----H-----M-----A-----I-----T-----E-----DK-----I-----
 B_US29.ME1 II--D-M-T-----A-----R-V-R-----G-----MP-R-----D-----TO-----N-AQ-----T-----A-K-----
 B_US30.A9307 -----N-----N-----V-----D-----D-----I-KT-----S-----Q-V-----E-----Y-----
 B_US30.B9309 -----N-----N-----K-A-----R-----SL-----VY-----D-----K-----Q-----S-----ST-----R-----E-----Y-----
 B_US30.C9105 -----N-----I-----D-G-----G-----G-----A-----I-K-----R-----B-----ST-----R-----E-----Y-----
 B_US30.D9110 -----IS-L-N-----T-K-----G-----G-----S-----A-----N-----STQ-----E-----K-----B-KP-----Y-----
 B_US30.E9108 -----D-----T-----K-P-V-----G-----G-----S-----A-----I-G-----N-----E-----T-KM-----Y-----
 B_US30.F9106 -----F-----K-----P-----SRL-S-----S-----R-T-----I-S-----G-----A-----Y-----
 B_US30.G9004 -----P-----P-----V-----B-B-----R-----T-----A-----D-----I-EG-----Q-----A-----K-----Y-----
 B_US30.H9308 -----P-S-----I-----K-I-----S-G-----SM-----VI-----Q-----I-----R-----B-----VAB-----YK-----I-SPP-----Y-----
 B_US30.I9108 -----D-----L-----I-----V-----G-----Q-----A-----A-----S-M-T-----L-----GK-----E-----A-----Y-----
 B_US30.J9110 -----D-I-----T-----K-P-N-T-----V-----G-----N-----A-----U-----IT-Q-----Q-----T-----KDK-----Y-----
 B_US31.9309073 -----S-----Q-----Q-----R-----N-----L-----A-----I-STQ-----K-----A-----N-----A-H-----
 B_US31.9206727 I--K-S-----I-----K-----V-----V-----S-----W-----A-----U-----LVQ-----R-VAI-----K-----RN-----T-----L-----
 B_US31.9108056 -----L-----T-----T-----G-----A-----A-----I-----X-A-----X-----G-----A-----L-----KS-----A-KP-QE-----
 B_UT.093 -----KKV-----IK-----H-D-----RL-S-----ASR-G-----N-----I-B-A-----Q-AI-----QF-D-----T-----V-----
 B_UT.1193 -----I-----N-V-----V-----G-----G-----M-----K-----F-----I-X-A-----Q-AI-----QF-D-----T-----V-----
 B_UT.1699 -----I-----I-N-----I-----T-K-----R-----M-R-----V-----A-B-G-K-----T-----N-E-----MR-AI-----I-----
 B_UT.270 -----IK-----L-----T-----T-K-----I-----L-----D-----Y-----I-KT-----EK-----R-VAT-----SGH-D-----V-----N-----P-----
 B_UT.352 -----A-S-----T-V-----K-----R-----LT-----A-----D-----N-----I-B-E-----K-----B-----G-----Z-----A-----
 B_UT.376 -----V-D-----M-----R-----V-----G-----P-----R-----A-----T-D-V-----I-D-----G-----Y-----T-----I-----R-----R-----
 B_UT.406 -----N-----N-----TRA-Q-----P-----M-----D-----I-E-----A-----N-----RTIM-----DI-----KHS-----H-----
 B_UT.672 -----I-----I-----P-----AP-----M-----I-K-A-----R-----Q-AI-----L-----A-----HP-QE-----
 B_UT.726 II--N-----D-V-----M-----H-----A-----Y-----DI-T-----K-----M-----A-----T-----
 B_VE.VB1 -----V-S-----I-----Q-----G-----Y-----I-----RT-----I-K-----K-----VA-----R-----L-----B-----R-----
 B_VE.VB2 -----I-----L-----T-----I-----G-----G-----V-----A-----G-----I-V-----G-----A-----
 B_VE.VB3 -----Q-----N-----I-----R-----M-----P-----L-----M-----N-----I-GT-----V-----
 B_VE.VB4 -----N-----N-----V-----I-----G-----Y-----E-----K-----N-----K-Y-----NKTY-----EK-----N-----V-----I-----P-----S-----
 B_VE.VB5 -----I-----N-----Q-I-----R-----7M-----VL-----Q-----Q-----Y-----D-----RKVAI-----E-----I-----XP-----T-----
 B_VE.VB6 -----N-----N-----X-A-----BD-----L-L-----R-----D-----T-----I-T-----V-----BR-----VFN-----S-----
 B_VE.VB7 -----A-S-----R-----A-R-----S-----R-----K-----R-----TIMK-----E-----GK-----T-----R-----
 B_VE.VB8 -----N-----I-----P-----A-----V-----I-----EX-----X-----
 B_VN1.NCM9 -----D-L-----M-----I-K-VY-----R-----I-----K-----T-----GN-----T-----I-B-----E-----B-----I-----L-----YT-----S-----
 B_ZA.2A504 -----D-N-----N-----G-----Q-----TV-----A-----K-----N-----I-ST-----R-----G-----R-----AN-----R-----L-----
 B_ZA.2A509 -----D-----N-----G-----Q-----TV-----A-----K-----N-----I-B-----K-----E-----V-----RK-----N-----K-----T-----S-----
 B_ZA.2A510 See sequences with inserts -----K-----K-----K-----K-----R-----ET-----N-----E-----V-----S-----
 B_ZA.2A512 -----K-----V-----K-----A-----R-----Y-----RD-----Y-----M-----Y-----ST-----R-----VA-----Y-----I-----T-----Y-----
 B_ZA.2A513 I--D-S-----H-----Q-----D-----S-----S-----A-----D-----I-TQ-----N-----AI-----R-----N-----TH-----T-----S-----
 B_ZA.2A524 -----D-S-----H-----Q-----D-----S-----S-----A-----D-----I-TQ-----N-----AI-----R-----N-----TH-----T-----S-----

C_CONSENSUS_97
 ALL_CONSENSUS
 C_BR.W2BR025
 C_BR1.HSP203
 C_BR2.91BR015
 C_BR1.91BR001
 C_BR1.91BR003
 C_BR1.91BR004
 C_BR1.91BR005
 C_BR1.91BR006
 C_BR1.91BR007
 C_BR1.91BR008
 C_BR1.91BR009
 C_BR1.91BR010
 C_BR1.91BR011
 C_BR1.91BR012
 C_BR1.91BR013
 C_BR1.91BR014
 C_BR1.91BR015
 C_BR1.91BR016
 C_BR1.91BR017
 C_BR1.91BR018
 C_BR1.91BR019
 C_BR1.91BR020
 C_BR1.91BR021
 C_BR1.91BR022
 C_BR1.91BR023
 C_BR1.91BR024
 C_BR1.91BR025
 C_BR1.91BR026
 C_BR1.91BR027
 C_BR1.91BR028
 C_BR1.91BR029
 C_BR1.91BR030
 C_BR1.91BR031
 C_BR1.91BR032
 C_BR1.91BR033
 C_BR1.91BR034
 C_BR1.91BR035
 C_BR1.91BR036
 C_BR1.91BR037
 C_BR1.91BR038
 C_BR1.91BR039
 C_BR1.91BR040
 C_BR1.91BR041
 C_BR1.91BR042
 C_BR1.91BR043
 C_BR1.91BR044
 C_BR1.91BR045
 C_BR1.91BR046
 C_BR1.91BR047
 C_BR1.91BR048
 C_BR1.91BR049
 C_BR1.91BR050
 C_BR1.91BR051
 C_BR1.91BR052
 C_BR1.91BR053
 C_BR1.91BR054
 C_BR1.91BR055
 C_BR1.91BR056
 C_BR1.91BR057
 C_BR1.91BR058
 C_BR1.91BR059
 C_BR1.91BR060
 C_BR1.91BR061
 C_BR1.91BR062
 C_BR1.91BR063
 C_BR1.91BR064
 C_BR1.91BR065
 C_BR1.91BR066
 C_BR1.91BR067
 C_BR1.91BR068
 C_BR1.91BR069
 C_BR1.91BR070
 C_BR1.91BR071
 C_BR1.91BR072
 C_BR1.91BR073
 C_BR1.91BR074
 C_BR1.91BR075
 C_BR1.91BR076
 C_BR1.91BR077
 C_BR1.91BR078
 C_BR1.91BR079
 C_BR1.91BR080
 C_BR1.91BR081
 C_BR1.91BR082
 C_BR1.91BR083
 C_BR1.91BR084
 C_BR1.91BR085
 C_BR1.91BR086
 C_BR1.91BR087
 C_BR1.91BR088
 C_BR1.91BR089
 C_BR1.91BR090
 C_BR1.91BR091
 C_BR1.91BR092
 C_BR1.91BR093
 C_BR1.91BR094
 C_BR1.91BR095
 C_BR1.91BR096
 C_BR1.91BR097
 C_BR1.91BR098
 C_BR1.91BR099
 C_BR1.91BR100

C_CONSENSUS_97
ALL_CMCN36
C_INQ.CMCN36
C_INQ.CMCN37
C_INS.GP1
C_INS.GP2
C_INS.GP3
C_INS.GP4
C_INS.GP5
C_INS.GP6
C_INS.GP7
C_INS.GP8
C_ME.EAL13
C_MM.S9750
C_ML1.12129
C_ML1.12203
C_ML1.12205
C_ML1.12209
C_ML1.12213
C_ML1.12215
C_ML1.12225
C_ML1.12229
C_ML1.12233
C_ML1.1227
C_ML1.6508
C_ML1.6512
C_ML1.6518
C_ML2.D3MA955
C_ML2.D3MA960
C_MY.9214082
C_MY.9214083
C_NL1.NL34024
C_NL1.IM93050
C_NOL.V3N13
C_NOL.V3N14
C_NOL.V3N17
C_NOL.V3N19
C_RU.RUS1A
C_RU.RUS2A
C_RU.RUS13A
C_RU.RUS20A
C_RU.YANA
C_RML.1134cons
C_RML.566cons
C_SW.SEC64
C_SO.1574
C_SO.SPH45
C_TW1.252
C_UG.45
C_UG.UG258
C_UG1.00513
C_ZA.HOP
C_ZA.ZA514
C_ZA.ZA517
C_ZA2.BooyD
C_ZA2.Dlu
C_ZA2.GOH
C_ZA2.ZAH18
C_ZA2.ZAH20
C_ZW.2647

V3 LOOP

[illegible]

III-126
DEC 97

V3 Region Alignments

III-128
DEC 97

[illegible]

```

      ***          * 2* ***          ***          ***          *
      B_CONSENSUS_97  IIIRSEDLTHAKTIIVNKKSVEN CTRPSNMT,RTSI, TIGPGQ,VFY,RTG,DIIQ,DI,KKAYC KINOTNNSEV,LRQVTERLL,KEMFNG,NKT, .T.IIFQPPVSGGGLSITNNF,NCRG,SPPYCMT
      ALL_CONSENSUS  -----I-----Q-----S-----N-----K-----R-----A-----A-----Q-----H-----N-----E-----MT-----Q-----AK-----R-----A-----S-----Q-----T-----S-----G-----
      E_TH9.179      -----V-----V-----N-----N-----N-----Q-----F-----T-----A-----L-----T-----
      E_TH10.92TH001 -----I-----H-----N-----N-----N-----Y-----R-----K-----I-----R-----T-----
      E_TH10.92TH012 -----I-----H-----N-----N-----N-----Y-----R-----K-----I-----R-----T-----
      E_TH1.286      X-----L-----A-----B-----MK-----N-----G-----Q-----I-----LL-----R-----AVV-----K-----IQ-----GN-----RA-----K-----I-----R-----Q-----Q-----KZ-----
      E_TH1.346      V-----L-----A-----B-----MK-----N-----G-----Q-----I-----LL-----R-----AVV-----K-----IQ-----GN-----RA-----K-----I-----R-----Q-----Q-----KZ-----
      E_TH1.396      V-----L-----A-----B-----MK-----N-----G-----Q-----I-----LL-----R-----AVV-----K-----IQ-----GN-----RA-----K-----I-----R-----Q-----Q-----KZ-----
      E_UK1.11643      -----I-----F-----R-----H-----K-----S-----T-----E-----KA-----A-----D-----N-----S-----L-----
      E_US_POC39506      -----B-----T-----V-----R-----H-----K-----S-----T-----E-----KA-----A-----D-----N-----S-----L-----
      E_UY.UR2      -----I-----M-----K-----Y-----K-----A-----G-----L-----T-----
      E_UY.UR4      -----I-----M-----K-----Y-----K-----A-----G-----L-----T-----
      E_UY.UR6      -----I-----M-----K-----Y-----K-----A-----G-----L-----T-----
      E_UY.UR7      -----I-----M-----K-----Y-----K-----A-----G-----L-----T-----
      E_VN1.BX6      -----X-----M-----P-----T-----K-----I-----T-----TT-----
      E_VN1.CT5      -----X-----M-----P-----T-----K-----I-----T-----TT-----
      E_VN1.DN3      -----X-----M-----P-----T-----K-----I-----T-----TT-----
      E_VN3.VN1      -----L-----A-----K-----S-----P-----
      E_VN2.VN2      -----L-----A-----K-----S-----P-----
      E_VN2.VN3      -----L-----A-----K-----S-----P-----
      E_VN2.VN4      -----L-----A-----K-----S-----P-----

```

	AAA	AA	AAA	V3 LOOP	AAA	AA	AAA	*	*		
P_CONSENSUS_97	QIIRSNISDWTETIIVHINETAQIN	CTRRPMTYTPKSIHILGPGQ	AFYSATG	DIDG	DIGRHAHC	NYSQT	YNTALQVR	AKLTKSHFPD	NET	TNIKFNSS	SGEDLEITHNSPCNCRGTSEFFYCN
ALL_CONSENSUS	I---E---TN-A---Q---S-E-	RI---	---	---	---	---	---	---	---	---	---
P_AR1_20016	V---X---X-A-I---S---X---	QX---R---	X---	X---	---	---	---	---	---	---	---
P_AR1_21280	I---A---A---S-X-X---	XX---R-X---	X---	E-X---	---	---	---	---	---	---	---
P_AR2_15	IV---A---A---S-T---	R---X---	Q---	Q---	---	---	---	---	---	---	---
P_AR2_16	IR---A---A---S-L---	PI---R---	T---	E---	---	---	---	---	---	---	---
P_AR2_18	I---A---A---S---	QI---	T---	---	---	---	---	---	---	---	---
P_BR_7944	---A---A---S-T---	Q---R---	T---	E---	---	---	---	---	---	---	---
P_BR_RJ103	---A---A---S-D---	Y---P---	R---	E---	---	---	---	---	---	---	---
P_BR1_82126	---A---A---S---	YF---	R---	E---	---	---	---	---	---	---	---
P_BR1_82162	IV---A-I---S---	---R---	L---	K---	---	---	---	---	---	---	---
P_BR1_82163	---A---A---S---	G---	I---	R---	---	---	---	---	---	---	---
P_BR2_HSP209	---LRE---A---S---	---	I---	R---	---	---	---	---	---	---	---
P_BR2_HSP229	I---A---A---S---	S---	---	T---	E---	---	---	---	---	---	---
P_BR2_HSP238	I---A---A---S---	---	I---	R---	---	---	---	---	---	---	---
P_BR2_HSP255	I---A---A---S---	---	I---	R---	---	---	---	---	---	---	---
P_BR2_HSP255A	I---A---A---S---	---	I---	R---	---	---	---	---	---	---	---
P_BR4_938802017	IV---A-I---S---	---	R---	S---	R-V-TA-E---	---	---	---	---	---	---
P_CN_GA4	---E-RK-I-I-Q-RS-E-	---	RI---	V---	---	---	---	---	---	---	---
P_CN_CAI6	---B-T---N-Q-RSIE-	---	SKTI-PR-	RI-L-R-V---	VM---	---	---	---	---	---	---
P_CN_CAZ0	---E---A---Q-RS-E-	---	I---	R---	---	---	---	---	---	---	---
P_CN_CML148	I---B-T---N-Q-ID---	---	RI---	V---	E---	K---	DV---	---	---	---	---
P_CN_CMR271	I---K-T---Q-RS-E-	---	---	---	---	---	---	---	---	---	---
P_CN_CMR275	I---F-TN-A---Q-RSIE-	---	---	---	---	---	---	---	---	---	---
P_CN_CMR304	See sequences with inserts	---	---	---	---	---	---	---	---	---	---
P_CY_HO441	V---E-T---S-N-Q-VS-P-	---	F---	---	RI---	T---	F---	---	---	---	---
P_FRI_FRMFB4	I---E---A---S---	---	G---	---	---	---	---	---	---	---	---
P_GA_VI354	---E-TS-N-Q-E---	---	G---	RI---	B-VI---	S-A-T---	---	---	---	---	---
P_GAI_VI354	I---E-TS-N-Q-E---	---	G---	RI---	R-VI---	S-A-T---	---	---	---	---	---
P_ML1_BR94009	I---PM-A---S-T---	---	---	---	---	---	---	---	---	---	---
P_ML1_ZR89081	I---A---A---S---	---	---	---	---	---	---	---	---	---	---
P_ML1_ZR93069	I---S-Q-N-Q-U-E-	---	D---	---	Q---	I-R-V---	---	---	---	---	---
P_R01_14018	---	---	---	---	---	---	---	---	---	---	---
P_R01_14020	---	---	---	---	---	---	---	---	---	---	---
P_R01_14024	---	---	---	---	---	---	---	---	---	---	---
P_R01_14027	---	---	---	---	---	---	---	---	---	---	---
P_R01_14028	---	---	---	---	---	---	---	---	---	---	---
P_R01_14034	---	---	---	---	---	---	---	---	---	---	---
P_R01_14036	---	---	---	---	---	---	---	---	---	---	---
P_R01_14041	---	---	---	---	---	---	---	---	---	---	---
P_R01_14046	---	---	---	---	---	---	---	---	---	---	---
P_R02_2	---	---	---	---	---	---	---	---	---	---	---
P_R02_3	---	---	---	---	---	---	---	---	---	---	---
P_R02_4	---	---	---	---	---	---	---	---	---	---	---
P_R02_5	---	---	---	---	---	---	---	---	---	---	---
P_R02_6	---	---	---	---	---	---	---	---	---	---	---
P_R02_7	---	---	---	---	---	---	---	---	---	---	---
P_R02_8	---	---	---	---	---	---	---	---	---	---	---
P_R02_9	---	---	---	---	---	---	---	---	---	---	---
P_R02_16	---	---	---	---	---	---	---	---	---	---	---
P_R02_17	See sequences with inserts	---	---	---	---	---	---	---	---	---	---
P_R02_18	---	---	---	---	---	---	---	---	---	---	---
P_R02_19	---	---	---	---	---	---	---	---	---	---	---
P_R02_20	---	---	---	---	---	---	---	---	---	---	---
P_R02_21	---	---	---	---	---	---	---	---	---	---	---
P_R02_22	---	---	---	---	---	---	---	---	---	---	---
P_R03_RM53002	---	---	---	---	---	---	---	---	---	---	---
P_R03_RM5301	---	---	---	---	---	---	---	---	---	---	---
P_R03_RM53012	---	---	---	---	---	---	---	---	---	---	---

[illegible]

H_CONSENSUS_57 VITRORHINOTRINIVOLKTPVETIN CTRPENTY AKSI WQPGQ AFY ATG QIIG DI AQMC NISRTMND LHEVUTQ...RHHFN NK...T.IIPKQ SQQDLBITNSP NCRG EPFICF
 ALL_CONSENSUS I--E--W-A-T--NES-----K-----A-----G-----A-----G-----
 R_CF_9BCT056 I--T--T-----T-----T-----L--R-----K--Q-----GI-L--N...S--N--M--VR-----N
 H_CM_C13 See sequences with inserts
 H_RU1_MLV10 See E-N-M-N-I-LA-----I--N-----EGL--Y-IA-----T-I-----G-----
 R_ZR_VT557 -----T-----S-P-----S-----K-Y--T-E--KR-----Q--E-----Q-----E--M--N-T-----

I_CONSENSUS_9? VVRSQNTNTNTHTITVLAKAVLIN CYPNGPTF.RKSV.HISGQ2.TVY.ATG.EKIO.DE.NQNHQ NISQNDWHTD.LKVISKEE..KRLPF.NK...Y.IKAPP.
 ALL_CONSENSUS II--S--N-A-T---NBS-S- ---N---I,R---AF---D--- ---RTS-N..QVAKK..RKH---I---S.SGDLSTTHSR.NGGS.IFFIOM
 I_CINCon

```

          AAA          AA   AAA          AAA   AA   A          AAA   AA   A          *
D_CONSENSUS_96  IRINGRNTSDGGKHIVTLNLTNTMT  CERPGNQTVQ.EIKI..GPMA..WYSHGLA..AGNNNSRA.AYC  TYNATDNEK.AL.KQTAE...RY..LSLV...NN...TSNVT.TMF.NRSSGGDAEVTHLMFCHGEFFYCHT
Q_CM.AMF70      --N-A-D-LEQ-----L-----QDIDQ..NR-----IG..GTA--S-----K--Q--I-----
Q_CM.NV5180     --T-A-----TP-----I-E-LAE--D-YT-----R--R-7-KRSTSPR-V-----K-K-V-N..Q--I-----N-----Q-----DSIN.MT--H-----L-----
Q_ES1.1158      V-I-R-----L-----S--ENVVGE..H-----A-----GKEN-T-----M-ST-E-T
Q_FR.CFBCP01    --KT-B-----VG-I--H--LS--H-----LS--H-----KSEIK-V-----E-T-H-I
Q_FR.CFBCP02    --N-A-----Q-----T-----Q-H-----R-----SE--R-----E-DI-K-H
Q_FR.CFBCP03    --A--TNT-N-----T--N-----RGIK--G-----S-V-OS--DL-GN-----H-M-E-I
Q_FR.CFBCP05    --R-T-MT-----TS-----M-X-RMK--R-AT--..LR--V-AAK.TESQNTG-I-----X-X-X-V
Q_FR.CFBCP07    --L-A-----Q-----X-T-X--H--LK-----K--X-----K-T-N
Q_FR.CFBCP08    --RMA-----RMA-----DGMATQ--MS-----L-K..SMTVR--S--H-----R--T--GGI-----PTE-I--..K-N-----I-R-----
Q_FR.DUR       --RNSG--L-----V--MS-----QIE-RE-K-A--T-F-----R--T--GGI-----PTE-I--..K-N-----I-R-----
Q_FR.VAU       --T-----S--LI--TN-TA--T-----I--K-MA-----A-S-NIK-D-T-----N-S--N-----MIT-----BY..MGQV--RFGNH--E--NFF-----
Q_GAI.VI586    --T-----S-G-----K--H-----E-K-T--R-----R-----MT-----K-V--TF-Q-MD--P-----

```

U. CONSENSUS_36
U_AR.20021
U_BK1.KJ10
U_BU1.91M009
U_CFI.4040
U_CFI.4056
U_CFI.406
U_CFI.4087
U_RE.K124
U_RE.KN976
U_ZL.RN9402H
U_ZL.K14803
U_GG1.9200035-2
U_GG1.9200008-9
U_ZH.ZM184
U_ZH.ZM1
U_ZH.ZM3

A Subtype

At this time there are viral sequences from 320 HIV-1 infected individuals associated with HIV-1 subtype A. The A subtype consensus sequence (A_CONSENSUS_97) generated from these sequences was based on the most common amino acid found in each position in an alignment of these sequences. All of these sequences have been published and/or have been made available for printing in the database by their authors.

- 1) **BJ1.ID#:** These 18 sequences are from female prostitutes, born in either Ghana or Togo, who live in Benin. 15 are from directly sequenced PCR products, derived via RT-PCR from patient serum RNA. 3 (233, 251 and 253) are from cloned PCR products, also by RT-PCR from serum RNA. Another subtype A sequence (366, U61870) is not included here, because it was nearly identical (254 of 255 bases identical) to the SF170 (M66533) sequence from Rwanda isolated in 1988, and thus it likely represents a lab artifact. [Heyndrickx et al.(1996)]. Accession numbers U61854–U61869, U61871 and U61873. Two subtype G isolates were also found in this study.
- 2) **BY.BLR10A:** This sequence is from Byelorussia. [Lukashov et al.(1995)]. Accession number L38411.
- 3) **CA.HWCL1:** This sequence is the first published sequence of subtype A HIV-1 in Canada. The patient had moved from Uganda in 1983, and was diagnosed as HIV+ in 1989. Viral RNA was recovered from archived, stored patient serum by binding viral particles to CD4-coated wells of an ELISA plate. The RT-PCR amplification product was cloned and 10 clones were sequenced. It is not clear whether this sequence is from a single clone, or a consensus of all 10. [Montpetit(1995)]. Accession number U34049.
- 4) **CF1.ID#:** These thirteen sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. The sequences are consensus sequences from cloned PCR products. DNA was isolated from co-cultured PBMCs. [Murphy et al.(1993)] Accession numbers L11457–L11458, L11461–L11463, L11469–L11471, L11474–L11475, L11477–L11479, L11484–L11496, L11498, L11518 and L11523–L11524. Some of them were sequenced again by Schmitt et al. unpublished: U43275, CF1.4023; U43109, CF1.4054; U43139, CF1.11423; U43136, CF1.1286.
- 5) **CF2.GAN and SAS:** These sequences were kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg, France. They are part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. Accession numbers U43111, U43171, M81063, M81064.
- 6) **CI1.ID#:** These six sequences are from six different AIDS patients suffering from pulmonary tuberculosis at the Pneumology Hospital of Cocody, Abidjan, Ivory Coast. In this study sequences came from coculture (on donor PBMCs) supernatant viral RNA, proviral DNA from each patients' PBMCs after coculture with donor PBMCs, and proviral DNA directly from uncultured PBMCs. PCR or RT-PCR was used to amplify the env V3 region, and 4–7 cloned PCR products were sequenced. A total of 66 sequences from the six patients were published. All 66 were subtype A, and inpatient sequences were more similar than interpatient sequences [Audoly et al.(1996)]. Accession numbers U59559–U59624.
- 7) **CI2.CI-ID#:** These sequences are from 11 of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-14 and CI-20 were symptomatic, and the others were asymptomatic. CI-14, CI-45 and CI-47 were serologically dually reactive for HIV-1 and HIV-2. The C2V3 region is part of a 900 bp sequence. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3–4 clones were sequenced, and the consensus of those clones is presented here. [Janssens et al.(1994a)]. Accession numbers X72024–X72027, X72030–X72039, X72043–X72056, and X72059–X72065.
- 8) **CI3.ID#:** These 21 sequences are from Abidjan, Ivory Coast. No other information is yet available. Ellenberger D.L. unpublished 1997. Subtypes A, D and G were found for Ivory Coast patients in this set. Ugandan sequences of subtype A were also part of this set. Accession numbers AF000449, AF000450, AF000452, AF000453, AF000454, AF000455, AF000456, AF000457,

- AF000459, AF000462, AF000464, AF000465, AF000466, AF000467, AF000470, AF000471, AF000472, AF000473, AF000474, AF000475, AF000493
- 9) **CM1.CA-ID#**: These sequences are 11 of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic (CA7, CA11, CA15, CA17, CA18, and CA21) and symptomatic (CA1, CA2, CA6, CA19 and CA22) individuals. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. [Nkangasong et al.(1994)]. The other six sequences were subtypes B, E, F and H. Accession numbers for the subtype A sequences: X80438, X80440, X80442, X80444–X80447, X80449, X80450, X80453, X80454.
 - 10) **CM2.CMR61**: CMR61 is one of two A group sequences from a 23 year old female commercial sex worker from Cameroon who was found to be triple-infected with subtypes A and D, as well as O group HIV-1 [Takehisa et al.(1997b)]. The other subtype A sequence from the same patient is labelled as CMR709 and is from a separate blood sample. Accession numbers U58148, U58150.
 - 11) **CM3.ID#**: These 15 sequences are all from Cameroon. Takehisa et al unpublished 1997. Accession numbers U69992–U69996, U70000, U70002–U70008, U70010 and U70014 U28719 (father); U28677 (father's partner); U28665 (mother)
 - 12) **CY.HOcon**: This is a consensus sequence from four individuals who were epidemiologically linked to one another: a father, a mother, their child and a woman who was a heterosexual partner of the father. These samples, like others in this study (see also subtypes B, C, F and I) were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. PBMC DNA was PCR amplified and cloned. Individual clones were sequenced. The father (patient HO34) was 36 years old and asymptomatic. He was known to have been seropositive for at least 2 years and had a CD4 count of 410. The mother (patient HO17) was 35, had been infected for at least two years, had a CD4 count of 2 and died in May 1994. The child (patient HO49) was 2 years old, had been infected since birth, and was asymptomatic, with a CD4 count of 1,211. The partner of the father (patient HO42) was 45 years old, had been infected for at least 4 years, was symptomatic with CD4 count of 80, and died in August 1994. Because of the close epidemiological linkage, a consensus of all 5 sequences is presented here. [Kostrikis et al.(1995)]. Accession numbers U28683 (child); U28674, U28719 (father); U28677 (father's partner); U28665 (mother).
 - 13) **DJ.DJ-ID#**: These three sequences from Djibouti were from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. DJ258 is NSI and uses the CCR5 coreceptor [Trkola et al.(1996b)]. Accession numbers L22939, L22941, and L23064.
 - 14) **ET.TP95001**: This study describes the distribution of HIV-1 subtypes in Ethiopia. HIV-1 RNA was collected from sera (from a majority of asymptomatic individuals) and RT-PCR amplified and sequenced directly. One subtype A and numerous subtype C sequences were found [Abebe et al.(1997)]. Accession number U88756.
 - 15) **FR1.ID#**: These 6 sequences are from members of the French military who are believed to have been infected while deployed outside of France (Djibouti, Guyana, and Central African Republic). Other sequences from this study were subtypes B, C, E, F, and unclassified. [Lasky et al.(1997)]. Accession numbers for subtype A were U58800–U58806 and U58778.
 - 16) **GA.VI191**: This sequence from Gabon was from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession number L22952. X90924, VI685
 - 17) **GA1.ID#**: These 6 sequences are from Gabon. Two (G41, G135) are from 1988–1989 samples from patients with AIDS living in Franceville Gabon. VI685 is from a 1992 sample of an AIDS patient from the Libreville General Hospital. VI1076 is from a 1993 sample of an AIDS patient from the Libreville General Hospital. LBV2310 and LBV23 are from 1988 samples from asymptomatic individuals sampled from the general population. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Delaporte et al.(1996)]. Accession numbers X90914, LBV23; X90915, G135; X90917, G41; X90918, LBV2310; X90921, VI1076; X90924, VI685. See also subtypes C, D, F, G and O sequences from this same study.

Sequence Descriptions

- 18) **GHLD687**: A single sequence from an individual from Ghana. Virus was cocultured on PBMCs and the env gene was PCR amplified. Provided by Georg-Speyer Haus, Frankfurt, Germany (Dr. Ursula Dietrich). [Dietrich et al.(1993)]. Accession numbers L07652, X68407.
- 19) **GH2.ID#**: These 9 sequences are from Ghana. Subtypes D and an A/G recombinant were also detected in this study [Takehisa et al.(1997a)]. Pol gene sequence is available for GH9 (U67040). Accession numbers for subtype A are U67051–U67054, U67057–U67059.
- 20) **GH3.ID#**: These 13 sequences are from Ghana. Subtypes D and G were also detected in this study [Ishikawa et al.(1996)]. Two additional subtype A sequences are not included here (GH14 U67689 and GH17 U67690) because they were greater than 98% identical to sequences reported by Takehisa et al [Takehisa et al.(1997a)] and may be from the same patients, although labelled differently in the two studies (GH14 here 99% identical to GH15 by Takehisa; GH17 here 98.2% identical to GH16 by Takehisa). Accession numbers for subtype A are U67685–U67700.
- 21) **IN1.ID#**: These two entries are from 1992 dried blood spot samples from Vellore near Madras, in Tamil Nadu state in southern India. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Samples came from previously identified HIV seropositive homosexual men. Other samples from the same region were all sub (see C_IN4.#). [Cassol et al.(1996)]. Accession numbers U53286 and U53291.
- 22) **IN2.GT6**: This sequence is from India. Seven other sequences from this publication were subtype C (see C_IN5.ID#). [Tsuchie et al.(1993)]. Accession number D13425.
- 23) **KE.K89**: This sequence is named "KENYA" in the GenBank entry, but is identified as K89 in the original manuscript. It is a Kenyan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession number L22943.
- 24) **KE.KEN-ID#**: These 19 sequences were derived from patients who were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced. [Janssens et al.(1994c)]. Accession numbers for the entire set of 23 patients surveyed in this study: U12984–U13006.
- 25) **KE1.ID#**: These 6 sequences were derived from patients who were part of a May–June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2–V3 region was amplified by RT-PCR. The PCR product was cloned and 20 clones from each patient were sequenced. Two other patients from this study had viral subtypes C and D. [Zachar et al.(1996a)]. Accession numbers U32658, U33763, U33764, U33766, U33767 and U34905.
- 26) **NL1.ID#**: These 13 sequences are from recent immigrants to The Netherlands from various countries. The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. [Lukashov et al.(1996)]. Accession numbers L76877, GH902304; L76903, GH9401488; L76883, GH915200; L76913, GH9501283; L76893, GH929927; L76896, ZR9303337; L76870, RW890388; L76875, RW893568; L76905, TZ9401664; L76881, KE913514; L76891, UG928308; L76887, AO924646; L76889, TZ925825
- 27) **RU1.ID#**: These 2 sequences are from Russia. Bobkov et al. 1996 Unpublished. Accession numbers: U33098, IVA6; U33097, GANI.
- 28) **RU2.ID#**: These two sequences are from patients living in Kaliningrad, Russia, [Leitner et al.(1996b)]. RU21 is a 40 year old male heterosexual, believed to have become infected in the Ivory Coast in 1988. RU24 is a 44 year old heterosexual male, believed to have become infected in Zaire. Sequences were determined by direct sequencing of PCR product from uncultured PBMC proviral DNA. Although RU24 is listed in the publication as unclassified subtype, this sequence clusters in phylogenetic analyses with SAS (U43171) and BJ193 (U61859) which are classified as subtype A. Because of the A-outlier nature of the RU24 env sequence, the authors also sequenced the gag P17 region of this sample, and found that it fell with subtype A (sequence not submitted to the databases). RU21 is a more typical subtype A. Accession numbers U69656 and U69658.

- 29) **RU3.ID#:** These 12 highly related sequences are from a set of 41 sequences from 12 IV drug users in Russia. Uncultured PBMC DNA was PCR amplified and cloned and individual clones sequenced. [Bobkov et al.(1997b)]. Although the sequences are similar enough to each other to suggest direct epidemiological linkage, no such linkage, other than the fact that they are all IV drug users, is indicated. Gag p17-24 and Env V4-V5 were also sequenced. Accession numbers U93611, U93612, U93614, U93616, U93618, U93620, U93622, U93624, U93627, U93629, U93631, U93633, U93635, U93636, U93637, U93638, U93642, U93643, U93645, U93646, U93647, U93648, U93650, U93651, U93655, U93656, U93657, U93658, U93659, U93660, U93661, U93665, U93666, U93668, U93671, U93673, U93675, U93676, U93677, U93678, U93679
- 30) **RW.564C:** This sequence represents 10 identical sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. Patient pair 564 was from Rwanda. A sample was collected from the infant at 30 months of age. Samples were also collected from the mother at 12 and 30 months after the birth. Mother sequences are not included in this consensus. [Mulder-Kampinga et al.(1993)], [Mulder-Kampinga et al.(1995)], [Lepage et al.(1996)], and [Kampinga et al.(1997)]. The child 564 Env sequence is from the entry with Accession number Z47881. Mother 564 sequences are in entries with Accession numbers Z47882-Z47902. Mother sequences are not included in this alignment. Gag gene sequences from mother/child pairs are also available in entries with accession numbers Z47903-Z47911, Z47912-Z47928, Z47929-Z47935 and Z47936-Z47950. The second mother/child pair was from the Netherlands, see G_NL.127C. The third mother/infant pair in this study was from the Netherlands, see B_NL.114C.
- 31) **RW.SF1703:** This sequence is from Rwandan isolate sf170, a biologically active clone reported to be macrophage-tropic. [Cheng-Mayer et al.(1988)]. See also U61870, which is not reported to be from SF170, but is greater than 99% identical to it. From this same isolate; 537 bases of the 5' LTR are in M66534, 619 bases of nef are in M81729 and 508 bases of tat and rev are in M66535. For env, the Accession number is M66533.
- 32) **RW1.W2RW-ID#:** These are six of seven sequences from asymptomatic individuals from Rwanda sampled in 1992. The seventh sequence (92RW009) has now been sequenced over the complete genome and found to be A/C recombinant. Thus it has been moved to the U_RW section. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. Accession numbers U08630-U08640, U08645, U08647-U08665, U08763-U08766, and U08793-U08794. One of these sequences may be mislabelled in the database; U08634 is labelled as patient 016, but is identical to sequences from patient 008. Sequences from samples 92RW020 and 92RW021 were greater than 98% each other, so only 92RW021 is presented here.
- 33) **RW2-ID#:** Nine consensus sequences from Rwanda. Saah, A. Unpublished 1994. Accession numbers U23216-U23373.
- 34) **RW3.PV#:** Bex, F. et al. Unpublished. These five sequences are from Rwandese patients with AIDS. The sequencing was done by the EEC Centralized Facility for HIV genome characterization, Georg-Speyer-Haus, Frankfurt, Germany. The complete envelope gene for PVP1 is available from a clone obtained after short-term co-culture on donor PBMCs. Two other shorter sequences of PVP1 env direct from patient PBMCs are also available. Accession numbers L07082-L07091. L07088 and L07089 were withdrawn from the databases by the authors, who felt they may represent PCR artifacts.
- 35) **RW4.ID#:** These 8 consensus sequences and 2 individual sequences, are from 7 infants and 3 mothers in a study of mother-infant transmission in Rwanda [Kampinga et al.(1997)], [Lepage et al.(1996)], [Mulder-Kampinga et al.(1995)] and [Simanon et al.(1994)]. Mother 566 was apparently infected with both subtype A and subtype C HIV-1, but the env regions from the child clones were all subtype A. See also C_RW1.ID#. Accession numbers for child 566 are Z76160-Z76161, Z76167-Z76168, Z76169-Z76176, Z76233-Z76248, Z76262-Z76273 and Z76717-Z76724; mother 730, Z76353-Z76362; child 564, Z76074-Z76083; mother 226, Z76046; child 538, Z76134-Z76143, Z76373-Z76382, Z76393-Z76412; child 074, Z75958; child 082, Z75998-Z76009, Z76650; child 081, Z75959-Z75968; child 618, Z76198-Z76207; mother 439, Z76048-Z76057, Z76064-Z76068 and Z76070.

Sequence Descriptions

- 36) **SE.H4:** This sequence is from an infant, born in Sweden to a woman who was believed to have been infected sexually in Uganda. The infant was born between June 1989 and August 1992 [Contag et al.(1997)]. Accession numbers U56274-U56283 and U56328.
- 37) **TZ.TAN-ID#:** These four sequences are from a set of 14 Tanzanian samples from symptomatic individuals, using serum samples taken in 1988 to generate PCR clones from viral RNA for sequencing. [Zwart et al.(1993)]. Accession numbers L01313, L01315-L01316, L01335, L01337-L01339.
- 38) **TZ2.ID#:** These two sequences were from patients at a clinic in Dar es Salaam, Tanzania. The individuals from which the virus was cultured showed clinical signs of AIDS, and the year of viral isolation was 1988. Viral cDNA was PCR amplified from donor PBMC, and one cloned PCR product per donor was sequenced. [Siwka et al.(1994)]. Accession numbers U12408, U12409.
- 39) **TZ3.ID#:** These 4 sequences are from the Mara region of rural northwest Tanzania. [Robbins et al.(1996)]. Subtype D was also found in this study. Accession numbers U61875-U61878.
- 40) **UG.1033:** This sequence is a consensus sequence of blood and CSF samples taken from a Ugandan patient 1033, CDC class IV-A. [Keys et al.(1993)]. Accession numbers Z23177, Z23182-Z23184, and Z23220-Z23223.
- 41) **UG.92UG037:** This sequence is from a complete genome PCR amplified from proviral DNA. The patient was a 31 year old asymptomatic female from Entebbe, Uganda. [Gao et al.(1996b)]. Accession number U51190. See also U09124, U09127.
- 42) **UG.964:** A single sequence used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. The sequence was derived from PCR amplified DNA from peripheral blood leukocytes. The patient was an asymptomatic individual from Uganda. [Pestano et al.(1995)]. Accession number U11599. See also B_US17.ID#, C_UG1.45, and D_UG7.ID#.
- 43) **UG.U455:** This sequence is from the 1985 Ugandan isolate U455; the complete genomic sequence is available. [Oram et al.(1990)]. The env ORF in this sequence is interrupted by an in-frame stop codon beyond the COOH end of the V5 region. Accession number M62320.
- 44) **UG.UG06:** This sequence is from blood collected from the Mulago Teaching Hospital in Kampala, Uganda. Viral RNA was harvested after 10-14 days of coculture with donor PBMCs and reverse-transcribed with AMV-RT. The env V3 region was pCR amplified and cloned. This sequence is from an individual clone. [Atkin et al.(1993)], [Pestano et al.(1993)]. Accession number M98503.
- 45) **UG1.W2UG-ID#:** Three sequences from asymptomatic individuals from Uganda in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. Accession numbers U08666-U08669, U08767-U08770, U08788-U08792, U08795 and U09125.
- 46) **UG2.ID#:** These 11 sequences are part of a set of sequences derived from 22 Ugandans who were attending an AIDS clinic, sampled in 1990. Consensus, PCR-clones, peripheral blood DNA. [Albert et al.(1992)]. Accession numbers M98902-M98905, M98908-M98910, M98914-M98917, M98919, M98924-M98928, M98938-M98941, M98946-M98966, and M98976-M98978.
- 47) **UG4.UG-ID#:** Two Ugandan sequences from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989-1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession numbers L22957 and L22951.
- 48) **UG5.ID#:** These 8 sequences are from Gulu, northern Uganda. They are from direct sequence of PCR product amplified from uncultured PBMCs. Blood samples were drawn from 217 pregnant women attending a clinic in Gulu, northern Uganda. Ages ranged from 17 to 37 years. The 29 seropositive women (13.4% of the 217 tested) were all asymptomatic. [Buonaguro et al.(1995)]. Accession numbers U44878-U44880, U44882, U44883 and U44885-U44887. Two subtype D sequences were also found in this study (see D_UG8).

- 49) **UG6.ID#:** These 21 sequences are from Abidjan, Ivory Coast. No other information is yet available. Ellenberger D.L. unpublished 1997. Subtypes A, D and G were found for Ivory Coast patients in this set. Ugandan sequences of subtype A were also part of this set. Accession numbers AF000496, AF000497, AF000498, AF000499, AF000500, AF000501, AF000502, AF000503, AF000504, AF000505, AF000506, AF000507, AF000508, AF000509, AF000510, AF000511, AF000512, AF000513, AF000514, AF000515, AF000516
- 50) **ZR1.ID#:** These ten sequences are part of a set of 14 A and D sequences from women from Zaire; 8 were healthy, 4 showed minor signs of illness, and 2 had AIDS. PCR-direct, peripheral blood DNA. [Potts et al.(1993a)]. Accession numbers L19624-L19626, L19628-L19630, L19632-L19634, and L19636.
- 51) **ZR2.ID#:** These sequences are from Zaire. They are unpublished, by M. Reitz et al. Accession numbers U43097-U43100.

B Subtype

At this time we have included viral sequences from 1078 HIV-1 infected individuals associated with HIV-1 subtype B. The B subtype consensus sequence (B_CONSENSUS_97) generated from these sequences was based on the most common amino acid found in each position in an alignment of these sequences. Please note that none of the studies which have published sequences of only the V3 loop sequences are included here, as the DNA sequences were deemed too short for phylogenetic analyses. (For example, LaRosa G, et al., *Science* 249:932-935 (1990) and Fouchier RAM, et al., *J. Virol.* 66:3183-3187 (1992).)

- 1) **AR.21281:** This sequence is from direct sequencing of PCR product from uncultured PBMCs, from a 1993 sample from Buenos Aires, Argentina. The patient had AIDS and reported promiscuous heterosexual risk behavior. Two other samples taken from unrelated patients in 1993 were subtypes F and one was found to be a subtype B/F recombinant. [Marquina et al.(1996)]. Accession number U68525.
- 2) **AR1.ID#:** These 11 sequences are from Rosario, Argentina. A total of 24 patients from different risk groups visiting a clinic in Rosario were included in this study. Of the 14 sequences determined, 11 were found to belong to subtype B and 3 were found to belong to subtype F. DNA was extracted from whole blood and PCR amplified. PCR products were directly sequenced. Subtypes of all 24 patients were tested by HMA. [Campodonico et al.(1996)]. Accession numbers U37030, U37031 and U37034-U37042.
- 3) **AU1.ID#:** Put forth as evidence that coinfection by multiple HIV-1 strains can occur in vivo, these three consensus (MRC1, MRC2, and MRC3) come from an Australian homosexual male who had been infected by more than one sexual partner, and harbored three distinct strains of HIV-1 B. The authors also found recombinant sequences, not included here. The sequences were PCR amplified from plasma RNA and PBMC DNA. [Zhu et al.(1995)]. Accession numbers U16372-U16388.
- 4) **AU2.C18CG:** This sequence is one of a group relating to an Australian blood donor infected with HIV-1 and six Australian recipients, all of whom remain symptom free with normal CD4 counts 10 to 14 years after infection. [Deacon et al.(1995)]. Samples from only the donor, D36 (U37271), and two patients, C18 (U37267, U37270) and C98 (U37268, U37269), appear to have been sequenced. These sequences have deletions in the *nef* gene and in the region of overlap of *nef* and the U3 region of LTR. The authors point to the importance of NEF or the U3 region of LTR in determining the pathogenicity of HIV-1 and suggest this strain of HIV-1 as a possible basis for a live attenuated vaccine. The complete genome of the virus from recipient C18 is in the entry with accession number U37270.
- 5) **AU3.ID#:** In this analysis of transmitting and non-transmitting mothers and the infants to whom the virus was transmitted, the authors claim that differences in variability of the env V3 loop crown octapeptide (HIGPGRF in this clone) were significantly correlated with vertical transmission. Sequences were determined by PCR directly from uncultured PBMCs and cloned PCR products. Only seven of the sequences are presented here: from infants, and nontransmitting mothers from which at least 250 bp was sequenced. 1089, 1063, 961 and JW are the infants. Accession numbers for transmitting mothers and their infants are: U66627-U66633, U66638-U66645 and U66650-U66653. Accession numbers for nontransmitting or indeterminate mothers are: U66625, U66626, U66634-U66637, U66646-U66649 and U66654-U66660. Database entries for many of these sequences contain "ann" in error, the paper shows these positions as deletions, not unknown sequence. *Nef* sequences from the LW-JW mother-infant pair of long-term survivors were published in [Wang et al.(1997b)], [Wang et al.(1997a)], [Saksena et al.(1997)] with accession numbers U73339-U73369.
- 6) **BE.SIMI84** One of two cloned env sequences from a patient with AIDS from Belgium. A vaccinia construct that expresses this gene was created to vaccinate the patient's non-infected brother with the goal of immune therapy by adoptive transfer of lymphocytes. [Bex et al.(1994)]. Accession number L07421.
- 7) **BR.002con** This consensus sequence is from entries with accession numbers L35489-L35493. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. Ranjbar, S. et al. Unpublished (1994).
- 8) **BR.BZ** This sequence is from an individual in a Brazilian HIV cohort study. PBMC DNA was PCR amplified in two sequential rounds, and six cloned PCR products were sequenced on both strands. A single clone containing an uninterrupted envelope open reading frame was reported. [da Costa et al.(1995)]. Accession number U28336.

- 9) **BR1.ID#:** These 21 sequences represent the B env subtype sequences found among 22 Brazilian out-patients with varying degrees of disease progression. They are consensus sequences from cloned PCR products. PCR was performed on PBMC DNA. [Potts et al.(1993b)]. Accession numbers for 21 of the 56 clones from which consensus sequences were calculated: L19225-L19237, L19240-L19246 and L20963.
- 10) **BR2.W2BR-ID#:** These 13 sequences are from individuals from Brazil. They are consensus sequences from cloned PCR products. Some of the clones were from cell-culture DNA, and some from cell-culture supernatant RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. [De Wolf et al.(1994)], [Osmanov et al.(1994)], [Gao et al.(1994a)]. Accession numbers U08670-U08714, U08771-U08778, U08780-U08782, U08792, U08796, U08797-U08800, .
- 11) **BR3.RJ- or SP-ID#:** These 32 sequences came from the Brazilian cities Rio de Janeiro and Sao Paulo. The sequences that are very short, containing V3 loop fragments insufficient for phylogenetic analysis, are not included here (5 of the 26). The full set included 20 viral sequences of the B subtype, an F subtype and a B-F recombinant (see subtypes F and U). The year of isolation for the sequences ranged from 1990-1992 for Rio de Janeiro, and 1992 for Sao Paulo. The only two with CD4+ cells < 200 were RJ636 and RJ27. The CDC clinical class ranged from II-IV. DNA extracted from PBMCs of HIV infected individuals was amplified, and the PCR product was directly sequenced. [Morgado et al.(1994)] [Sabino et al.(1994c)] and [Sabino et al.(1996)]. More recent unpublished sequences from the same isolates are also included here. Accession numbers U00400-U00401, U00403, U00405, U00407-U00414, U00416-U00418, U00421, U00424-U00425, U00427, U08975, U31586, U31587 and U31589-U31591.
- 12) **BR4.BZ-ID#:** These 2 sequences are from seropositive Brazilian patients. Virus was cultured on donor PBMCs and proviral DNA was harvested from positive cultures. PCR was used to generate sequencing templates. [Louwagie et al.(1994)]. Accession numbers L22087 and L22088. The gag gene sequences from these same isolates are also available in L11752 and L11754. See also F_BR2.BZ-ID#.
- 13) **BR5.ID#:** These 10 sequences are from entries with accession numbers L19328-L19337. Banda, C. I. Unpublished (1993).
- 14) **BR6.ID#:** These six sequences are from patients living in cities in Brazil (P3, Sao Paulo; P4, P6 and P7, Bahia; P8, Parana; P9, Rio de Janeiro) and sampled between 1987 and 1989. Sequences were determined from directly sequenced PCR products (except P6 which was a cloned PCR product), after coculture of patient PBMCs with donor PBMCs. [Couto-Fernandez et al.(1994)]. Accession numbers X78512-X78517.
- 15) **CAL.ID#:** These 19 sequences are from 19 different Vancouver, Canada homosexuals represented in 36 database entries by N. Michael et al. Accession numbers U52888-U52906, U53103-U53119. Patients A and B (U52888-U52904) were described in [Michael et al.(1997)] as a normal-progressor rapid-progressor pair, with patient A being the rapid progressor. Only patient A sequence (U52889) is presented here, as patient B sequences were closely related. The C patients were local controls, also homosexuals from Vancouver, who are not well described in the publication. Patient A sequences from U52901 and U52902 are greater than 99% identical to JR-CSF over their entire length, and U52904 has some regions of identity to JR-CSF. Another linked pair in this set included patients C108 and C118, of which only C108 is presented here. Patient A died of AIDS-related *P. carinii* pneumonia on October 6, 1995, almost exactly 10 months after experiencing symptoms of primary HIV infection. Patient A was homozygous for an HLA-DR1 allele previously found to be associated with rapid progression. Patient A never seroconverted for HIV, despite high viral load and the ability to seroconvert for hepatitis B after vaccination in 1990.
- 16) **CH1.ID#:** These 19 sequences came from 24 individuals living in Geneva, Switzerland who were recently infected at the time of blood drawing. Samples were collected between January 1988 and September 1993. Sequences were determined directly from PCR products of uncultured PBMC DNA or serum cDNA. All subjects were asymptomatic, 19 subjects had p24 antigen levels ranging from 5 to 6,357 pg/ml and 5 subjects had no detectable p24 antigen. Two subjects were epidemiologically linked (K11 and K16) so only one of those two is presented here. Two other individuals showed identical DNA sequences over the entire V3 region (K53 and K77) so only one of them is presented here. Three other individuals (K13, K42 and P4) had sequences nearly identical to the LAI (IIIB) laboratory strain of HIV. Although the authors

Sequence Descriptions

- are convinced that these are not contaminants, and that a IIB-like strain of HIV is circulating in Geneva, they are not included in this alignment. [Antonoli et al.(1995)]. Accession numbers U10957–U10980.
- 17) **CI.CI-22:** A single B subtype sequence from a set of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-22 was symptomatic. The C2V3 region is part of a 900 bp fragment that was sequenced for each individual. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3 clones were sequenced, and the consensus of those clones is presented here. [Janssens et al.(1994a)]. Accession numbers X72040–X72042.
 - 18) **CM.CA5:** A single B subtype sequence from a set of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals, specifically, patient CA-5 was asymptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. [Nkengasong et al.(1994)]. Accession number X80452.
 - 19) **CN1.ID#:** These 54 sequences are from the Yunnan Province, China. No information is yet available. Unpublished 1995, Y. Shao and H. Wolf. Several of these sequences (U20009, U20012, U20013, U20018, U20023, U20024) are greater than 96% identical to the SF2 strain of HIV-1 (see B_US.SF2). Accession numbers U20001–U20054.
 - 20) **CN.1798:** This sequence is from a dried blood spot collected in 1992 from the spouse of an IV drug user in China. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. Accession number U53316.
 - 21) **CU.94CU053:** This sequence is from a bisexual male, most probably infected via heterosexual contact in 1992, in Cuba. Virus was isolated in 1994, 2 years after seroconversion, by cocultivation of patient PBMCs with donor PBMCs. This isolate exhibits a rapid/high, syncytium-inducing phenotype. [Gomez et al.(1996)]. Accession number U48855.
 - 22) **CY1.HO#:** These sequences are from samples, like others in this study (see also subtypes A, C, F and I), which were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. PBMC DNA was PCR amplified and cloned. Individual clones were sequenced. Patient 04 was a 24 year old man, asymptomatic with a CD4 count of 516, who had lived abroad, and had been seropositive for at least 4 years. Patient 11 was a 29 year old homosexual man, symptomatic with a CD4 count of 277, and seropositive for at least 7 years. Patient 21 was a 38 year old bisexual man, asymptomatic with a CD4 count of 743, who was infected in Greece and seropositive for at least 4 years. Patient 25 was a 34 year old heterosexual man, asymptomatic with a CD4 count of 650, who had lived in the U.S. and was seropositive for at least 6 years. Patient 27 was a 20 year old homosexual man, asymptomatic with a CD4 count of 709, infected in Cyprus and seropositive for at least 2 years. Patient 28 was a 39 year old homosexual woman, asymptomatic with a CD4 count of 430, infected in Cyprus and seropositive for at least 1 year. Patient 29 was a 49 year old heterosexual man, asymptomatic with a CD4 count of 420, infected in Cyprus and seropositive for at least 2 years. Patient 39 was a 37 year old heterosexual man, asymptomatic with a CD4 count of 470, had lived in Greece and was seropositive for at least 5 years. Patient 40 was a 29 year old heterosexual woman, symptomatic with a CD4 count of 92, who had lived in the U.S. and was seropositive for at least 6 years. Patient 43 was a 36 year old homosexual man, symptomatic with a CD4 count of 396, who had lived in the U.K. and Greece and had been seropositive for at least 6 years. Patient 45 was a 32 year old heterosexual man, asymptomatic with a CD4 count of 453, who was infected in Cyprus and seropositive for at least 1 year. Patient 46 was a 32 year old heterosexual man, whose partner had died of AIDS, asymptomatic with a CD4 count of 107, had lived in the U.K. and was seropositive for at least 4 years. Patient 48 was a 22 year old hemophiliac man, asymptomatic with a CD4 count of 276, who had been seropositive for at least 11 years. Patient 50 was 32 year old homosexual man, whose partner had AIDS, asymptomatic with a CD4 count of 315, who had been seropositive for at least 1 year. Accession numbers U28663, U28666 (04); U28664 (11); U28667–U28671 (21, 25, 27, 28, 29); U28675, U28676 (39, 40); U28678 (43); U28680–U28682 (45, 46, 48); U28684 (50).
 - 23) **CZ.BTSPR:** This sequence is from a 32 year old homosexual man from Prague, Czech Republic. It was included as part of a study of HIV-1 sequences from the neighboring Slovak Republic. DNA from cocultured PBMCs was PCR amplified and the PCR product cloned. Eight clones were sequenced from each patient, although only one sequence is presented in the publication and in the database. It is not stated whether each single sequence is a consensus of the eight clones or a single clone. [Zachar et al.(1996b)]. Accession number U53195.

- 24) **DE.D31:** This sequence is from isolate D31. [Kreutz et al.(1992)]. It has never been well described, it is only shown as HIV1-D31 in figure 3 of the paper. The complete genome has been sequenced. Accession number U43096.
- 25) **DE.HAN:** This sequence is from an infectious clone from the German isolate DE.HAN-2. [Saueremann et al.(1990)]. Isolate HAN2 was isolated from a 39 year old homosexual German patient with AIDS related complex in 1986. This patient died from complications of AIDS in 1987. HAN2 was highly cytopathic in T cell line MT-2 it was able to productively infect MT-4, H9 or Jurkat cell lines. Genomic DNA from infected MT-2 cells was used to prepare a lambda phage genomic library. Two full-length clones, HAN2/2 and HAN2/3 were purified. HAN2/3 was used for DNA sequencing. Accession number U43141.
- 26) **DE.Serocons:** This sequence is a consensus sequence from 7 hemophilia patients who all received the same lot of beta-propiolactone and UV-light inactivated clotting factor in Bonn or Goettingen, Germany, from November 1989 to March 1990. The virus and the patients have been extensively studied over time, since initial seroconversion. The sequences which were combined to create the consensus were from proviral DNA from cultured PBMCs, PCR amplified and cloned. [Kasper et al.(1994)]. Accession numbers S76444 and S76446.
- 27) **DE1.ID#:** These 7 sequences are from a Neurology thesis by I. Weber of Rostock, Germany. They are from blood and CSF specimens from 6 patients. For each of the patients, phylogenetic analysis showed that blood and CSF sequences were more similar to each other than to other database sequences. For one patient, one of the two blood sequences was different enough from another blood and CSF pair to be included here. Accession numbers Z78482, Z78485-Z78493.
- 28) **DE2.ID#:** These 4 sequences are from a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and 6 Scottish intravenous drug users. The sequences were used in a study of HIV-1 vpr, vpu, and env V3 regions and how they vary between risk groups. [Kuiken et al.(1996b)]. Accession numbers Z68529, Z68530, Z68537 and Z68538.
- 29) **ES.ID#:** These 36 sequences are from 41 patients sampled in Madrid, Spain between 1985 and 1991. Proviral DNA was extracted from uncultured patient PBMCs and the C2V3 region was PCR amplified. The PCR products were directly sequenced. Two of the sequences reported in this set (D22-28 and D22-48) were 99.5% identical to the LAI (IIB) lab strain of HIV-1 and are not included here. 3 other groups of sequences had members that were greater than 98% identical to each other (R1, R2 and R3; THF13-2, THF12-24; S1, S4) and only one of each of them is presented here. [Quinones-Mateu et al.(1996)]. Accession numbers U40533-U40552 and U45286-U45307.
- 30) **ES.S61:** This sequence is from a 1989 blood sample from a 4 year-old boy from Madrid, Spain with CDC stage P2CD2 disease. Virus was cocultured on PBMCs and MT-2 cells prior to sequencing. The SI/NSI phenotype of this isolate on MT-2 cells was traced to a single amino acid change in the V3 loop [Sanchez-Palomino et al.(1993)] and [Olivares et al.(1997)]. Accession numbers L04604, L04605, L04606 and L05659.
- 31) **FR.J91:** This sequence is from one of the JBB clones from the French patient Bru. [Wain-Hobson et al.(1991)] and [Guo et al.(1991)]. Accession numbers X57449-X57459 and X57461.
- 32) **FR.LAI:** This sequence is from the French isolate LAI (formerly BRU) which is also referred to as IIB. [Wain-Hobson et al.(1985)]. Also see: [Alizon et al.(1986)], [Lukashov & Goudsmit(1995)] and [Wain-Hobson et al.(1991)]. Accession numbers K02013, L23090-L23103, X01762, L48380-L48399, M64178-M64223, M64406-M64415 and M64768-M64775. Other sequences which are of this type include: PV22, K02083; MFA, M33943 [Stevenson et al.(1990)]; F12, Z11530; BH8, K02011; BH10, M15654; TH4, L31963; MCK1, D86068; PM213, D86069; F12CG, Z11530; and HXB, K03455, M38432, M64775 and M14100. The variation of the IIB isolate in culture was studied by [Lockey et al.(1996)], Accession numbers U54647, U54649, U54651, U54653, U54655, U54657, U54659, U54665, U54667, U54681, U54683, U54685 and U54689. The variation of IIB/LAI in 9 years of infection in a chimpanzee has been studied by Fultz et al. unpublished, Accession numbers U56866-U56883 and U56888-U56899. The IIB/LAI isolate of HIV-1 has also been extensively studied in cases such as the infected lab worker. See for example [Jr et al.(1994)], [Pincus et al.(1994)] U12030-U12055. The tropism of isolates from the lab worker for primary PBMCs and failure to grow in T-cell lines was localized to the V3-loop by Lishan Su et al. [Su et al.(1997)]. Recombinant virus pNL4-3, with envelope

Sequence Descriptions

- from LAI(BRU) and gag-pol from NY5 has also been studied: [Adachi et al.(1986)] Accession number M19921, [Duensing et al.(1995)] Accession number L42371 and [Salminen et al.(1995)] Accession number U26942. Other database entries with IIB-LAI sequences can be found in the patented sequences section of GenBank, in the cloning vector section, and in the primate section (for example U19867, A00647, A04321 and M18404).
- 33) **FR1.ID#**: These 4 sequences are from Toulouse, France. In this study, 4 mother-infant pairs were followed during pregnancy and after birth. The inter- and intra-patient sequence similarities of this set of 308 sequences has been controversial, because some infant sequences were identical to sequences from other mothers. For purposes of this V3 section, only one sequence from each of the 4 infants is presented here. [Briant et al.(1995)], [Korber et al.(1995)] and [Jr et al.(1996)]. Database entries for all 308 sequences are found with accession numbers U24717-U24999 and U25001-U25025.
 - 34) **FR2.ID#**: These 20 sequences are from members of the French military who are for the most part believed to have been infected while deployed outside of France (Chad, Djibouti, Gabon, Guyana, Mayotte and Cameroon). An additional sequence (FRMP040, U58787) was listed as subtype C in the paper, but clustered with subtype B in analysis done at the HIV Database and it has been listed as subtype unclassified until more information is available. Other sequences from this study were subtypes A, C, E, F, and unclassified. [Lasky et al.(1997)]. Accession numbers for subtype B were U58808-U58826 and U58777.
 - 35) **GA.OYI**: This sequence is from the Gabonese isolate OYI (designated elsewhere as isolate 397), isolated from a healthy HIV-1 infected individual. GA.OYI appears to have been the first viral sequence from Africa that phylogenetically clustered with North American viruses. [Huet et al.(1989)]. Accession number M26727.
 - 36) **GB.AIT**: This sequence is from an individual at the time of seroconversion. Proviral DNA was extracted from PBMCs from a patient who was viraemic and had an equivocal HIV-1 antibody status, and the env V3 region was PCR amplified. The PCR products were cloned and the DNA sequence determined for 15 clones. These data showed that the V3 region contained only limited sequence heterogeneity with a major variant (shown here) accounting for 66% of the protein quasispecies present. [Ait-Khaled & Emery(1993)]. Accession number S69598.
 - 37) **GB.CAMI**: This sequence is from the British isolate CAM1. McIntosh A, and Karpas A, Thesis (1991), Cambridge University, England. Accession numbers D10112, D00917.
 - 38) **GB.Man** This sequence was PCR amplified from the 1959 "Manchester sailor" kidney tissue. The sequence of the complete genome is available and clusters with subtype B contemporary HIV-1 sequences. [Zhu & Ho(1995)]. Accession number U23487.
 - 39) **GB.V-ID#**: A set of six sequences from a study of hemophiliacs from Scotland who were originally thought to have been infected by the same batch of factor VIII. (ScV12 is a sequence from a hemophiliac from the U.S., included as a control). All are consensus sequences of multiple direct PCR sequences obtained from limiting dilution of PBMCs. The Scottish hemophiliacs were infected in 1984 and the PBMCs were obtained for analysis in 1989. Although the samples were potentially related, they were deemed sufficiently divergent in this region for inclusion in this set. [Simmonds et al.(1990)], and [Balfe et al.(1990)]. Accession numbers M61327-M61346 and M61391-M61407. Database entries with accession numbers M84240-M84317 are more sequences from patient 82, taken over the period from 1984-1991. [Holmes et al.(1992)]. Database entries with accession numbers L13488-L13497 are also from these patients [Zhang et al.(1993)], as are U58393-U58465 [Cleland et al.(1996)], [Leigh Brown & Cleland(1996)], [Ludlam et al.(1985)].
 - 40) **GB1.CPHL1**: This is a consensus from the British isolate 93-08020, clones 1, 4, 7, 18, 19 and 43. It was referred to as 93-08020 in [Arnold et al.(1995c)] and was isolated from the patient referred to as CPHL1 in [Arnold et al.(1995a)]. CPHL1 is a surgeon and CPHL2 was a patient of his in 1986, approximately 7 years prior to sampling for this study. Because sequences from CPHL1 and CPHL2 are no more similar to each other than to sequences from the general population, transmission cannot be concluded, and both sequences are included in this alignment. Accession numbers U21100 (clone 1) and U23112-U23116 (clones 18, 19, 4, 43 and 7 respectively).
 - 41) **GB2.CPHL2**: This is a consensus from the British isolate 93-17305, clones 3, 11, 18 and 25 [Arnold et al.(1995c)]. It was isolated from the patient referred to as CPHL2 in [Arnold et al.(1995a)]. Accession numbers U23117-U23120 (clones 11, 18, 25 and 3 respectively).

- 42) **GB3.ID#:** The CPHL7 sequence is a sequence from the British isolate 94-24612, clone 13 [Arnold et al.(1995c)]. It was isolated from the patient referred to as CPHL7 in [Arnold et al.(1995a)]. Accession number U23126. U23127 is a second clone from this same isolate. Sequences from three other patients epidemiologically linked to CPHL7 (CPHL6, accession numbers U23130-U23132; CPHL8, U23128-U23129; CPHL9, U23133-U23135) are not included in this alignment. The 4995 sequence is from entries with accession numbers U23136-U23138.
- 43) **GB4.ID#:** These eight sequences are from British isolates from St. Bartholomew's Hospital, London (M23470, M26864, M30156, M37677 and M37658) and Hammersmith Hospital, London (AC, JB and WB). Sequences were determined from cloned PCR products from PCR amplified DNA from either cocultured (M23470 and M26864) or uncultured (M30156, M37677, AC, JB and WB) patient PBMC proviral DNA. [Douglas et al.(1996)]. Complete envelope gp160 sequences were determined for at least one clone from each patient. Ugandan samples also sequenced in this report were subtypes D or D/A recombinant. Accession numbers for London samples were U36859-U36864, U36869, U36870, U36872-U36880 and U36882.
- 44) **GB5.ID#:** These 11 sequences are from Scotland. They are part of a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and six Scottish intravenous drug users, from which regions of vpr, vpu and env were sequenced. The authors found consistent differences in the sequences between the homosexuals and IV drug users. Only 34 of the 47 patients' sequences are reported in the publication. [Kuiken et al.(1996b)]. See also B_NL12 and B_DE2 sets.
- 45) **GM.GM6:** A sequence from Gambia, as yet unpublished. Bobkov et al. 1996 unpublished. Accession number U33101. See also Gambian sequences of subtypes C and I.
- 46) **HT.RF:** This sequence is from the full-length lambda clone HAT-3, from Haitian isolate RF. RF is from a 28 year old Haitian male who had moved to the United States at age 25, in 1980. He had no history of IV drug use, homosexuality or blood transfusions. In October 1983 he had 20 lb weight loss, giardia with diarrhea, thrush, and diffuse lymphadenopathy. His CD4/CD8 ratio was 0.08. He died in December, 1983. Primary culture from a November 1983 blood sample was co-cultured on HUT-78 cells. [Jr et al.(1992)] [Starcich et al.(1986)] [Popovic et al.(1984)]. Accession numbers M17451 and M12508.
- 47) **HT1.D-ID#:** These seven sequences are from Haitians, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratories of Dr. Beatrice Hahn at the University of Alabama, and Dr. Marcia Kalish at the Centers for Disease Control, Atlanta, GA. Except for D2HA590, the full gp160 was sequenced from clones derived from expanded culture stocks. D2HA590 is a direct sequence from PCR amplified DNA from expanded culture. The sequence ID numbers are abbreviated, for example D2HA590 can be read as DAIDS sequence (D), isolated in 1992 (2), Haitian (HA), patient 301590 (590). Full length env for some of these have been expressed [Gao et al.(1996a)]. Accession numbers: U08441-U08447, U04900. Both U08441 and U08442 are sequences from patient HT1.D1HA651 and are identical over the region of interest. Accession numbers for additional clones derived from these patients: U04901-U04906.
- 48) **HT2.H-ID#:** These 25 sequences are from Haitians. All sequences were PCR amplified from the infected individuals PBMCs, and this set includes direct sequences of PCR amplification products, consensus sequences of multiple clones of PCR products plus one direct sequence, and single clones of PCR products. Full length env for some of these have been expressed [Gao et al.(1996a)]. These sequences were provided by the Centers for Disease Control, Atlanta, GA USA (Dr. Chin-Yih Ou), and John Hopkins University School of Hygiene and Public Health, Baltimore, MD USA (Dr. Neal Halsey), and the Centers for Development and Health, Complexe Medico Social de la Cite Soleil, Port-au-Prince, Haiti (Dr. Reginald Boulos). Accession numbers L07145-L07161, L07163-L07165, L07167-L07207, L07209-L07239, L07241-L07246, U08441-U08447.
- 49) **ID.1701:** This sequence is from a dried blood spot collected in 1992 from a male homosexual patient in Indonesia. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. Accession number U53317.
- 50) **ID1.ID#:** These 7 sequences are from a set of 14 sequences from Indonesia [Porter et al.(1997)]. Accession numbers U68195-U68201. Subtype E was also identified in 7 other samples from Indonesia in this study. PCR products were directly sequenced from either uncultured PBMC DNA or cocultured PBMC DNA.

Sequence Descriptions

- 51) **IN.IND9:** This sequence was from a heterosexually infected patient from New Delhi, India. DNA was isolated from cocultured PBMCs after one week of culture. PCR product was cloned and a single clone was sequenced. [Tripathy et al.(1996)]. Accession number U31364. See also C_IN3.ID#.
- 52) **IN1.ID#:** These four sequences were isolated in Hyderabad, Andhra Pradesh, in southern India. The C2V3 region of env was amplified by nested priming from DNA from PBLs from fresh blood samples. Date of sampling and health status of HIV-1 infected individuals is unknown. [Baskar et al.(1994)]. Accession numbers L29091-L29094.
- 53) **IT.ID#:** These two sequences are consensus sequences from 4 clones each. They were obtained from PCR-amplified proviral DNA from Langerhans cells from skin patches from a deceased AIDS victim in Italy [Sala et al.(1995)]. Small V1-V2 region sequences and V3-loop sequences from the same skin samples were published in [Sala et al.(1994)]. Entries with accession numbers U20670-U20677 were used here. Entries with accession numbers Z34376-Z34458, Z34470-Z34513 and Z34515 were V1-V2 and V3-loop sequences from the same patient.
- 54) **IT.PD:** This sequence is from the thymus of a 26 year old male with CDC stage III disease. Proviral DNA was PCR amplified from genomic DNA extracted from patient thymocytes, and the PCR product was directly sequenced. A V3 sequence from PBMC proviral DNA was also determined, as were gp41 sequences from both thymus and blood. [Calabro et al.(1995)]. Accession numbers U09254-U09255.
- 55) **IT.RM1:** This sequence is from a patient from Milan, Italy who was homozygous for the delta-32 deletion in CCR5 [Balotta et al.(1997b)]. Accession number U92491.
- 56) **IT1.ID#:** These 10 sequences are from infants infected in utero. The sequences came from PCR amplified DNA of uncultured PBMCs, PCR amplified DNA of cultured PBMCs, or from RNA from serum collected at or shortly after delivery. [Scarlatti et al.(1993)]. Accession numbers L08277-L08286. Sequences from the mothers of these infants are also available in entries with accession numbers L08287-L08372.
- 57) **IT2.ID#:** These 3 sequences are from unpublished sequences in entries with accession numbers X92424-X92426.
- 58) **IT3.ID#:** These 12 sequences are consensus sequences from 12 individual patients in a study of long term non-progressors and rapid progressors, from Milan, Italy [Balotta et al.(1997a)]. Accession numbers U95381-U95497.
- 59) **JP.ETR:** A Japanese isolate from long-term cell culture with a truncated env gene, due to a point mutation of a CAG codon to a TAG stop codon. [Shimizu et al.(1992)]. Accession numbers D12582, D01205-D01207, D12584 and D12571.
- 60) **JP.GUNA:** A Japanese 1989 isolate HIVGUN, infectious to T cells, was adapted to grow in fibroblast-like BT cells. A single amino acid change at the tip of the V3 loop was shown to be responsible for the change in tropism, GPGR to GSGR. [Takeuchi et al.(1991)]. Accession number M59192.
- 61) **JP.JH32:** This is a sequence from a lambda clone of Japanese isolate JH3, which was isolated in 1986 from a 10 year old Japanese Hemophiliac. [Komiya et al.(1989)]. Accession number M21138.
- 62) **JP.JH23A:** This sequence is from a Japanese patient dually infected with HIV-1 subtypes B and E [Xin et al.(1995a)]. Accession number D67089. See also D67090, E_JP.JH23B.
- 63) **JP.KM03:** This sequence is from a 28 year old hemophilia B patient with CDC stage IV disease and T-cell count of 20, living in Japan. The authors [Hattori et al.(1991)] also sequenced the V3 region from 28 other Japanese individuals, but only the V3-loop amino acid sequence is available from the other patients. Accession number S70936.
- 64) **JP.JNIH1M:** This sequence is from a Myanmar (Burmese) individual living in Japan, obtained by direct sequencing of PCR-amplified proviral DNA from peripheral blood mononuclear cells. [Weniger et al.(1994)]. Accession number L32084.
- 65) **JP1.ID#:** These 10 sequences are from 12 patients with varying rates of disease progression. Patients 1 and 2, both extremely rapid progressors who died of AIDS within 8 months (patient 1) and 3 years (patient 2) of infection by sexual contact in 1991. These two shared very similar sequences and were suspected to be epidemiologically linked, so only patient 2 sequence is presented here. Patients 8, 9, 15, 19, 47, 63 and 65 were infected between 1983 and 1985 by contaminated blood products. Patients 20 and 28 were infected by sexual contact. Sequences were determined by either RT-PCR from plasma viral RNA or PCR from uncultured PBMC proviral DNA. PCR products were cloned prior to sequencing. Each of these 11 sequences is a single clone, taken as representative of that patient. [Shioda et al.(1997)]. Accession numbers AB002829-AB002974, AB002988-AB003019. Patient 43 was infected with subtype E.

- 66) **KPKr111** This sequence is from an unpublished database entry with accession number X93580.
- 67) **LT.LIT#:** These 5 sequences are from Lithuania. [Lukashov et al.(1995)]. Accession numbers: L38417, LIT18A; L38412, LIT11A; L38419, LIT21A; L38416, LIT17A; L38420, LIT17A.
- 68) **MM1.ID#:** These two sequences are from Myanmarese (Burmese) individuals living in Myanmar, obtained by direct sequencing of PCR-amplified proviral DNA from peripheral blood mononuclear cells. [Weniger et al.(1994)]. Patient 02 is a male IV drug user, and 05 is a female prostitute, both were from Mandalay. Accession numbers L32088, L32089.
- 69) **MM2.ID#:** These 5 sequences are from dried blood spots collected in 1992 from a male STD patient (1782), a female prostitute (1739), and 3 IV drug users (1748, 1755 and 1763) in Myanmar. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. Accession numbers U53304–U53308.
- 70) **MY1.ID#:** These 11 sequences are from IV drug using prisoners in a prison in Kuala Lumpur, Malaysia. PCR products amplified from uncultured PBMCs were directly sequenced. [Brown et al.(1996)]. This report also included subtypes C and E in Malaysia. Accession numbers U65538–U65548.
- 71) **NL.114C** This consensus sequence represents sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. Samples were collected from the infant at birth, at 6 weeks and at 9 months of age. Samples were also collected from the mother before birth, at birth and after birth. Mother sequences are not included in this consensus. [Mulder-Kampinga et al.(1993)]. [Mulder-Kampinga et al.(1995)]. Infant 114 is from Accession numbers L21111–L21153. Mother 114 sequences are from Accession numbers L21028–L21110. Infant 127 sequences are from Accession numbers Z47817–Z47832. Mother 127 sequences are from Accession numbers Z47833–Z47880. Gag gene sequences from mother/child pairs are also available in the entries with accession numbers Z47903–Z47911; Z47912–Z47928; Z47929–Z47935; Z47936–Z47950. The second mother/child pair was also from the Netherlands, see G_NL.127C. The third mother/infant pair in this study was from Rwanda, see A_RW.564C.
- 72) **NL.168:** This is a consensus sequence of 3 clones after culturing in PBMC. The isolate was originally from an AIDS patient in Amsterdam. [Wrin et al.(1995)]. Accession numbers U15030–U15032. A V3-loop (105 bp) segment from the original isolate has been previously reported. [Fouchier et al.(1992)]. Accession number L06694.
- 73) **NL.wolfson:** This is a consensus sequence from 6 patients infected with the same unit of blood in Amsterdam. [et al.(1992)]. Accession numbers U04530–U04537.
- 74) **NL.X1con:** This is a consensus sequence of 10 clones from a recipient in a donor-recipient study. Sequences from donor Y and recipient X2 are also part of this study, but are not included here. [Cornelissen et al.(1995)]. Accession numbers Z47505–Z47514 are from X1. Other new sequences analyzed in this paper include Z47411–Z47540. Sequences M91828–M91838 (donor H and recipient O, referred to as patients A14 and A13, respectively in [Wolfs et al.(1992)] see B_NL1.A13) were also re-analyzed in this study.
- 75) **NL1.ID#:** These nine sequences are part of a study of presumed donor-recipient pairs from an HIV-1 transmission study conducted in the Netherlands. If pairs were extremely close or identical, only the recipient is included here. Recipient samples were from the first sample to be antibody positive, and are numbers 1,3,5,7,9, and 13. These sequences are consensus sequences of multiple clones from PCR amplified serum RNA. [Wolfs et al.(1992)]. Recipient A1 was also studied as patient H1 in [Kuiken et al.(1993)] and so is not included here (see B_NL4.H1). Accession numbers M91819–M91827, M91829, M91831–M91832, M91839, M91857–M91870, M91872, M91874, M91881–M91884, M91891, M91893, 91895–M91908, M91910, M91911–M91926. Number 13, and the donor were also analyzed in [Cornelissen et al.(1995)].
- 76) **NL2.ID#:** These two sequences are part of a Dutch study of mutations occurring over a five year period (starting in 1985) in two patients. Serum RNA was PCR amplified and multiple clones were sequenced. The consensus for each patient is shown. [Wolfs et al.(1991)] and [Zwart et al.(1992)]. Accession numbers M74591–M74684.
- 77) **NL3.NET-ID#:** These six consensus sequences from the Netherlands are samples from AIDS patients, using serum samples to generate PCR clones from viral RNA for sequencing. [Zwart et al.(1993)]. Accession numbers L01282–L01297.
- 78) **NL4.ID#:** These 74 sequences represent a study of early seroconverters from different times with different risk factors for transmission during the AIDS epidemic in the Netherlands. The year the sample was taken

Sequence Descriptions

is indicated in the last part of the sequence name. The risk group of the individual from whom the virus is derived is indicated in the first letter of the sequence name (I, B and H for IVDUs, hemophiliacs, and homosexuals, respectively). Viral genomic RNA from sera was PCR amplified and amplification product was directly sequenced. [Kuiken et al.(1993)]. Accession numbers Z29219–Z29225, Z29256–Z29258, and Z29262–Z29325.

- 79) **NL5.ID#:** These 18 consensus sequences are from a study of patients with, and without, AIDS dementia complex (ADC) in the Netherlands. Not all patients were from the Netherlands. Samples were collected between 1986 and 1992. Viral genomic RNA from sera and/or cerebral spinal fluid was reverse transcribed and PCR amplified and clones were sequenced. [Kuiken et al.(1995)] Accession numbers Z37531–Z37534, Z37734–Z37963, Z37970–Z37971
- 80) **NL6.ID#:** 16 is a consensus sequences of four sequences used in a study of HIV-1 envelope-mediated syncytium formation. The consensus represents four clones from one patient, two clones of the consensus are SI and two are NSI. 320 is a single SI clone. The sequences were derived from PCR amplified DNA from provirus cultured in PHA-stimulated PBMCs [Andeweg et al.(1992)]. Accession numbers L08655–L08662. Complete genomes from patient 320 SI and NSI strains are found in entries with accession numbers U34603 and U34604 [Guillon et al.(1995)], tat and other genes from patient 320, some of the same isolates, are found with Accession numbers M64489–M64492 [Groenink et al.(1992)].
- 81) **NL7.ID#:** These two consensus sequences are from sets of sequences (Accession numbers U05797, U13240, U13241, U13243–U13247 for consensus 537 and U13242, U13248–U13252 for 1058) used in a study on the dynamics of HIV sequence changes in vivo and the utility of heteroduplex analysis. Both sequences were derived from PCR amplified PBMC DNA. Consensus 537 represents a set of sequences from a Dutch patient with a relatively stable CD4+ cell count at 62 months post-seroconversion. Consensus 1058 represents sequences from another Dutch patient whose CD4+ cell count at 73 months post-seroconversion was declining faster than 537's. [Delwart et al.(1994)]. See also US10.ID#.
- 82) **NL8.672** This is a sequence from a patient early in infection, before, or around the time of seroconversion. Three other patients studied in this paper (537, 1058 and 594), had previously been reported. See B_NL4.594, B_NL7.1058con and B_NL7.537con. [Shpaer et al.(1994)] and [Delwart et al.(1995)]. Accession numbers U23651–U23663, U23667, U23670. See also US11.ID#.
- 83) **NL9.ID#:** These 19 sequences are from a cohort of homosexual men living in Amsterdam who seroconverted between 1985 and 1989. The sequences are from direct sequencing of PCR products after RT-PCR from serum RNA. Samples for h1, h139, h491, h1140 and h1234 were obtained at seroconversion. Samples for h138 and h1136 were obtained 12 months and 29 months after seroconversion respectively. The sequence for h320 was obtained from proviral DNA, 2 months after seroconversion. [Zwart et al.(1994a)] and [Zwart et al.(1994b)]. Accession numbers L25884 and U05786–U05808. More sequences from this same cohort of men were published in [Kuiken et al.(1996b)] and [Kuiken et al.(1996a)] Accession numbers Z67875–Z67876, Z67885–Z67939, Z67941–Z67960, Z68015–Z68089, Z68109–Z68110. Some vpr, vpu and other region sequences are available from some of these patients as well. Some of the database entries for this set appear to be duplicates of sequences reported in other studies.
- 84) **NL10.ID#:** These 13 sequences are from recent immigrants to The Netherlands from various countries. The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. [Lukashov et al.(1996)]. Accession numbers L76878, SR903853; L76897, SR9304737; L76879, SR911515; L76874, UM893272; L76912, UM9403860; L76888, SR925752; L76890, SR926969; L76894, GQ9301341; L76882, MA913670; L76886, SR923572; L76873, BR891413; L76910, UM9403051; L76885, GF921953.
- 85) **NL11.ID#:** These 2 sequences are consensus sequences of 81 clones (patient N) and 105 clones (patient F) from serum, sigmoid tissue and fecal matter from each patient. All sequences from patient N were more similar to other sequences from patient N than to any other sequence in the database. Likewise all sequences from patient F were most similar to other patient F sequences. [van der Hoek et al.(1996)]. Accession numbers Z76463–Z76648.
- 86) **NL12.ID#:** These sequences are part of a set containing 15 Dutch homosexuals, 19 Dutch intravenous drug users, 2 German homosexuals, 2 German intravenous drug users, 5 Scottish homosexuals and six Scottish intravenous drug users, from which regions of vpr, vpu and env were sequenced. The authors

- found consistent differences in the sequences between the homosexuals and IV drug users. Only 34 of the 47 patients' sequences are reported in the publication. [Kuiken et al.(1996b)]. See also B_GB5 and B_DE2 sets. Some of the patients in this study have been previously studied. For example, entries with accession numbers Z68061 and U05787 are both from the same patient.
- 87) **NOL.ID#:** These 36 sequences are from Norwegian patients who were part of the Oslo HIV cohort study [Engelstad(1996)]. Uncultured PBMC DNA was PCR amplified in two nested PCR reaction steps. PCR products were directly sequenced. Where two peaks of equal height were observed at a single position, IUPAC ambiguity codes were used. Health, sex, year of sample (1989-1992), and risk group (IVDU, Het, Homo, Hemo) for each patient were noted in a table in the publication. Four sub sequences were also part of this set (X92913, X92914, X92917, X92918). Accession numbers for the 36 subtype B sequences are X92902-X92912, X92915, X92915, X92919-X92941.
 - 88) **PRLD.ID#:** These four sequences are from Puerto Rico, and were generated as part of the DAIDS variation program in the laboratory of Dr. Marcia Kalish at the the Centers for Disease Control, Atlanta, GA. The C2V3 region was directly sequenced from PCR amplification of DNA from viral culture. The sequence ID numbers are abbreviated; for example DZPR732 can be read as DAIDS sequence (D), isolated in 1992 (2), Puerto Rico (PR), patient 301732 (732). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. Accession numbers U04926-U04929.
 - 89) **PY.ID#:** Ten sequences from 10 patients living in Asuncion, Paraguay. All 10 were male patients with symptoms of AIDS. Virus was propagated in tissue culture for an unstated length of time prior to harvesting proviral DNA for PCR and sequencing. PCR products were directly sequenced. PY.3614, PY.3615 and PY.12837 were syncytium-inducing and the other 7 were not. The tenth sequence, PY.3614p was PCR amplified directly from patient PBMCs with no culturing. The significant differences between the cultured sequence from this patient (PY.3614c) and the direct sequence, indicates that the virus that grew out in culture was a minority of the virus present in PBMC. The sequence of PY.3614c is not included in this alignment. [Cabello et al.(1995)]. Accession numbers U28949-U28959.
 - 90) **RU1.RUS#:** These 2 sequences are from [Lukashov et al.(1995)]. Accession numbers: L38405, RUS3A; L38407, RUS4A.
 - 91) **RU2.ID#:** These 15 sequences are from homosexuals living in St. Petersburg, Russia. Matykh et al unpublished (1995). Accession numbers U40283-U40285 and U40319-U40330.
 - 92) **RU3.ID#:** These 13 sequences are from Rostov, St. Petersburg, Sochy, Ecaterinberg, Komi Republic, and Nizhny Tagil, Russia. Sequences were determined by direct sequencing of PCR product from uncultured PBMC proviral DNA. Although several of these cases were suspected to have epidemiological linkage (RU9 and RU10 are beleived to have been infected by the same individual, sequences from the index case are not available; RU11, RU13 and RU27 were all thought to share a common index case, although none of them was directly infected by the index case), the sequences do not clearly indicate such epidemiological linkage and all 13 sequences are presented here. [Leitner et al.(1996b)]. Accession numbers U69646-U69652, U69654, U69657, U69660, U69662, U69666 and U69667.
 - 93) **SE.pt11s113:** This sequence is from one (patient 11 sample 113, collected in 1988) of a set of 13 sample from 9 epidemiologically linked individuals. The index case (patient 1) was a Swedish male who is believed to have contracted HIV while visiting Haiti in 1980. Six Swedish females were infected (patients 2,4,5,7,8 and 11) by patient 1. Two males (patients 6 and 10) were then infected by these females, and two HIV-infected children (patients 3 and 9) were born to the women. Sequences from each patient were determined by PCR amplification from uncultured PBMCs and direct sequencing. Heterogeneous sites were indicated with IUPAC codes. Extensive phylogenetic analysis was done to determine which methods accurately reconstructed the true phylogeny. [Leitner et al.(1996a)]. Accession numbers U68496-U68521.
 - 94) **SE1.ID#:** Seven sequences that are consensus sequences of blood and CSF samples taken from each patient. The CDC disease stage class for the patients are as follows: II pts 930, 2815; III pt 931; IV-B pt 2951; IV-A pt 1032; and IV-C2 pts 1433, 1866. [Keys et al.(1993)]. Accession numbers Z23178-Z23181, Z23185-Z23187, Z23192-Z23195, Z23200-Z23219, Z23224-Z23227, Z23232-Z23235, and Z23240-Z23255.
 - 95) **SE2.ID#:** These five sequences are from patients in Goetebotg, Sweden each of whom had recently seroconverted at the time of sampling in 1985-1990. Sequences from the sexual partners of all five patients who were believed to have transmitted the virus to these recipients were also determined, but are

Sequence Descriptions

- not shown here due to the great similarity to the recipient sequences (94% to 99% identity). Three of the recipients (R1, R3 and R4) were homosexual, and the other two were heterosexual. [Furuta et al.(1994)]. Accession numbers U10929-U10950.
- 96) **SE3.ID#**: These 5 sequences are from a study of the ability of beta-chemokines RANTES, MIP-1alpha and MIP-1beta to inhibit primary isolates with SI and NSI phenotypes. The conclusion was that NSI viruses were inhibited by these beta-chemokines, whereas SI viruses were not. Each viral isolate was cocultured on PHA-stimulated donor PBMCs for 3 days, PBMC DNA was harvested, PCR amplified, and the PCR product directly sequenced. For each of these 5 patients (labelled patients A-E) one SI and one NSI isolate were sequenced, only the NSI sequence is presented here. [Jansson et al.(1996)]. Accession numbers U76078-U76087.
- 97) **SK1.ID#**: These nine sequences are from homosexual men living in the inner city of Bratislava, in the Slovak Republic. Patients 11, 12, 15 and 23 were classified as CDC stage A1 and were not taking any medication. Patients 18 and 20 were CDC stage B2 and were taking AZT. Patients 9 and 28 were CDC stage C3 and were taking AZT. No information is available for patient 22. Two other patients (10 and 51) were included in this study, but their sequences are not shown here because after their sequences proved to be similar to sequences from patients 9 and 28, respectively, epidemiological investigation indicated that they (9 and 10; 28 and 51) had been sexual partners. DNA from cocultured PBMCs was PCR amplified and the PCR product cloned. Eight clones were sequenced from each patient, although only one sequence is presented in the publication and in the databases. It is not stated whether the single sequence is a consensus of the eight clones or a single clone.[Zachar et al.(1996b)]. Accession numbers U53192-U53194, U53196-U53203.
- 98) **TH.93TH067**: This sequence is from Thailand. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. [Penny et al.(1996)]. Accession numbers U39258 and U39259.
- 99) **JP.JH23A**: This sequence is from a Japanese patient dually infected with HIV-1 subtypes B and E [Kin et al.(1995a)]. Accession number D67089. See also D67090, E_JP.JH23A.
- 100) **TH.T8174**: This sequence comes from a study of the genetic heterogeneity and epidemiological distribution of HIV1 in Thailand. The host was an intravenous drug user and the sequence was obtained from PCR amplified PBMC DNA. [Ou et al.(1993)]. Accession number L19238. See also E_TH.T8178.
- 101) **TH1.ID#**: These ten sequences are from individuals from Thailand. PCR-direct, peripheral blood PBMC DNA. [Ou et al.(1992b)] and [Ou et al.(1993)]. (Published erratum appears in *Lancet* 342:250 (1993).) Accession numbers L07442, L07449-L07456 and L07460.
- 102) **TH2.ID#**: The TB132 sequence is from a set of isolates from HIV seropositive individuals from Thailand. PCR, PBMC co-culture, DNA. Full env sequence is available. [McCutchan et al.(1992)]. Please note: the TB132 locus name in the database corresponds to the McCutchan, et al. "BK132" isolate. Accession number L03697. The CM237 sequence is from PBMC proviral DNA. [Mascola et al.(1994)]. Accession number L14570. See also B_US14
- 103) **TH3.W2TH-ID#**: 2 sequences from Thailand from asymptomatic individuals. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April 1994 supplement to the *Human Retroviruses and AIDS 1993* database. [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. Accession numbers U08715-U08718, U08801-U08802, and U08783-U08784.
- 104) **TH4.ID#**: These twelve sequences are B subtype sequences from Thailand. Ten were genetically most similar to HIV-1 found in the Americas and Europe; these sequences were derived from people infected prior to 1988 (diagnosed in 1986 or 1987). The other two (N762 and N763) were designated B' and were isolated from people with more recent infections, 1988 and 1992. The sequences were obtained from PCR amplified PBMC DNA. The naming of the sequences submitted to the databases does not correspond with the naming of the sequences in the paper. [Kalish et al.(1994)]. Accession numbers for the entire set of thirteen sequences studied in this publication: U15576-U15588.
- 105) **TH5.ID#**: These three sequences are B subtype sequences from Thailand. Two individuals believed to be dually infected with subtypes B and E were analyzed. It is not clear from the paper or the database

- entries, which sequences came from individual 1 and which from 2. [Artenstein et al.(1995)]. Accession numbers U21471, U21473, U21475. See also E_TH6.ID#.
- 106) **TH6.ID#**: These 4 entries are from 1992 dried blood spot samples from Thailand. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Samples came from 3 previously identified HIV seropositive IV drug users (58, 11, 15) and a homosexual (86). Other samples from the same region were all subtype E (see E_TH8.#). [Cassol et al.(1996)]. Accession numbers U53310, U53311, U53314 and U53315.
- 107) **TH7.ID#**: These 70 sequences are from IV drug users in Bangkok, Thailand, who were undergoing methadone treatment at 14 treatment clinics. Blood samples were collected between January and April, 1994. Uncultured PBMC DNA from each patient was PCR amplified, and the PCR product was directly sequenced (except for patient 108, in which PCR product was cloned and 2 clones were sequenced, one shown here). Of the 84 patients sampled, 69 were Thai B, one (091) was typical subtype B, and 14 were subtype E. [Kalish et al.(1995)]. Accession numbers U22543-U22547, U22549-U22552, U22554-U22556, U22558-U22560, U22562-U22566, U22568-U22574, U22576-U22603, U22605-U22608, U22610, U22613-U22616, U22618-U22623 and U22626. See also E_TH9.
- 108) **TT.QZ4589**: This sequence is from Trinidad. W. Blattner et al. Unpublished 1995. Accession number U32396.
- 109) **TW1.ID#**: These 16 sequences are from healthy HIV-1 carriers or AIDS patients from Taiwan [Chang et al.(1997)]. Three subtype B sequences in this set were greater than 97% identical to the HXB2/LAI lab strain (see B_FRLAI) of HIV-1, and are not included here (TW83, U73049; TW271, U73059; and TW335, U73061). The manuscript reports that 123 of 143 sequences from Taiwan were subtype B, but only 27 of the 143 sequences were submitted to the sequence databases. Other subtypes found in Taiwan in this study were B (17 cases), C (1 case), F (1 case) and G (1 case). Accession numbers for B subtype are U73045-U73054, U73056, U73057, U73059, U73061 and U73063-U73069.
- 110) **US.ACP1**: This virus was cultured from a seronegative man with Kaposi's sarcoma. (See: [Ho et al.(1989)]). ACP1 was the sequenced after one passage in PBMCs. Accession number M80660. The sequence AC-H9 (M80661) was also derived from this patient. [Ashkenazi et al.(1991)].
- 111) **US.AD39**: This sequence is one recipient sequence from a set of 36 clones from a homosexual donor-recipient pair. Other donor-recipient pairs were studied by quantitative homoduplex tracking assay (QHTA) but only this pair was sequenced. [Zhu et al.(1996)]. Accession numbers U50780-U50815.
- 112) **US.ADA**: This sequence is from the monocytropic U.S. isolate ADA. [Westervelt et al.(1991)]. The complete genome of a derivative of this isolate has been reported to have been sequenced in [Theodore et al.(1996)]. The complete genome is from a macrophage tropic HIV-AD8 isolate, derived from HIV-AD87, which was in turn derived from HIV-ADA. Accession numbers M60472, AF004394.
- 113) **US.ALA1**: This sequence is from an infectious clone of the 1985 U.S. isolate AL-1, taken from a patient with AIDS. Buckler-White, A. et al., Unpublished (1988). Accession number M38430.
- 114) **US.BAL1**: This sequence is from the macrophage tropic U.S. isolate BAL, harvested from lung alveolar tissue. Reitz M, et al., Unpublished (1990). Accession numbers M68893, M68894. The entry with accession number M63929 is 98% identical to this sequence and is also derived from the BAL isolate [Hwang et al.(1991)].
- 115) **US.BCSG3**: This is a fragment of a full genomic sequence from the provirus SG3, cloned as a single proviral unit. This clone replicates more efficiently in chimpanzee than in human lymphocytes, and is extremely cytopathic and syncytium inducing (SI) in immortalized human T-cell lines. [Ghosh et al.(1993)]. Accession number L02317.
- 116) **US.BRVA**: This sequence is from U.S. isolate BRVA, which was taken from the brain tissue of a AIDS patient with neurological disorders. [Anand et al.(1989)]. Accession number M21098.
- 117) **US.BWB**: This consensus sequence is from sequences derived from PCR amplified PBMC DNA from brain tissue. [Monken et al.(1995)]. Accession numbers L17088-L17126.
- 118) **US.CDC42**: This sequence is from an infectious clone of the U.S. isolate CDC-451 [Desai et al.(1986)]. It was isolated in 1984 from a 16 year old hemophilia A patient, who died of AIDS in June, 1984. The viral isolate was propagated on H9 cells prior to cloning into Lambda phage and M13 phage for sequencing. Accession number M13137.

Sequence Descriptions

- 119) **US.Diaz:** This sequence is one of a set of 223 closely related sequences. All 223 sequences came from 3 patients with a common source of HIV, a blood donor and two recipients of this donor's blood [Zhang et al.(1997)]. Accession numbers for all 223 sequences are U29433-U29437, U29956-U30145, and U43035-U43054. Gag sequences from 12 clones are also available in entries U31573-U31584.
- 120) **US.FASH:** This sequence is also known as 91US005.11 from the WHO Global program on AIDS. It is from a 17 year old female with primary symptomatic infection (PSI) from Birmingham, Alabama, USA. The risk factor for this individual was reported as heterosexual contact. The sequence indicates that this clone is subtype B. Blood was drawn in 1991. Although this clone has a 34 bp deletion followed by a premature stop codon in the envelope gene, relative to other subtype B sequences, the env protein was strongly positive in a CAT complementation assay. The patient from which this clone was isolated experienced rapid CD4 decline. In addition to the 17 amino acid truncation of the gp41 peptide, the env gene has a mutation at amino acid position 721 (bases 2157-2159) replacing tyrosine with cysteine (Y721C). This mutation has been shown in SIV to increase cell surface concentrations envelope glycoprotein [LaBranche, C.C. et al, *J. Virol.* 69: 5217- 5227 1995]. The complete gp160 coding region of this isolate was sequenced along with those of others collected at major epicenters of the AIDS epidemic [Gao et al.(1996a)]. Accession number U27434.
- 121) **US.HOBR:** This sequence is also known as 91US006.10 from the WHO Global program on AIDS. It is from a 28 year old male with primary symptomatic infection (PSI) from Birmingham, Alabama, USA. The risk factor for this individual was reported as homosexual contact. The complete gp160 coding region of this isolate was sequenced along with those of others collected at major epicenters of the AIDS epidemic [Gao et al.(1996a)]. Accession number U27443.
- 122) **US.JFL:** This sequence is from a non-infectious clone from the monocytotropic U.S. isolate JFL [McNearney et al.(1990)]. Accession number M31451. Other sequences from HIV-1 isolates epidemiologically linked to this isolate can be found in database entries with accession numbers L06256-L06273. [McNearney et al.(1993)].
- 123) **US.JM:** This sequence, along with B_US.WM, came from viral isolates after short term culture in PBMCs, PCR amplification, and cloning of PCR products. Both are from asymptomatic, seropositive individuals. [Ashkenazi et al.(1991)]. Accession number M80662.
- 124) **US.JRCSF:** This sequence is from an infectious clone of the 1986 U.S. isolate JRCSF, derived from from the CSF of a patient who died with Kaposi's sarcoma and severe AIDS encephalopathy. The infectious clone JRFL was isolated from the brain of the same patient. JRFL does not replicate in Jurkat, U937 or HUT78 cells. JRFL does replicate in mononuclear phagocytes, and the macrophage-tropic region of the virus was determined by domain swapping with NL4-3 to reside in a 157 amino acid region of gp120 including the V3 loop. [O'Brien et al.(1990)]. Accession numbers U63632, M38429, U45960. Also see: [Pang et al.(1990)], [Pang et al.(1991)], [Koyanagi et al.(1987)], [Klasse et al.(1996)].
- 125) **US.MN:** This sequence is from U.S. isolate MN, taken from a 6 year old male pediatric AIDS patient from the Newark, New Jersey area in 1984. His mother was an IV drug user who died of pneumonia in 1982. His father was also HIV seropositive. A complete genome of isolate MN is found in M17449. Other sequences from this patient from the 1984 blood sample and from a 1987 sample taken shortly before death (U72495) are available also. [Jr et al.(1992)] [Gurgo et al.(1988)]. Accession number M17449. See also L48364-L48379 [Lukashov & Goudsmit(1995)].
- 126) **US.NY5CG:** This sequence is from the 1984 U.S. isolate NY5. [Willey et al.(1986)]. Accession number M38431. See also GenBank accession number K03346. A recombinant between NY5 and LAI has also been extensively studied, see B_FR.LAI entry.
- 127) **US.P896:** This sequence represents a molecular clone from an primary isolate derived from a Jamaican man who immigrated to Philadelphia 15 years earlier. At the time of viral isolation, he had no antiviral therapy, but was an AIDS patient with < 10 CD4 cells per mm³. The infectious molecular clone from which this sequence was derived is both macrophage-tropic and extremely cytopathic in lymphocytes. [Collman et al.(1992)] and [Kim et al.(1995)]. Accession numbers M96155, U39362.
- 128) **US.RJS:** This is a consensus of six biologically characterized clones from patient RJS, isolate 4. The HIV-1 infected individual had been infected for five years at the time of isolation. Patient RJS was a 31 year old homosexual male from California, who reported having sexual encounters with at least 1000 partners while HIV-infected, from 1980 to 1985 [Hahn et al.(1986)]. Virus was isolated via coculture on

- donor PBMCs prior to cloning and sequencing. Complete env sequence is available. [Daniels et al.(1991)] and [Fisher et al.(1988)]. Accession numbers M37491 and M37573-M37577.
- 129) **US.SB(A-C)**: These three sequences are from 1988 U.S. isolates taken from a woman, her daughter and her sexual partner. The three viruses are epidemiologically linked, however the amino acids sequences appeared sufficiently divergent in this region to merit the inclusion of all three samples. Sequences were directly sequenced from PCR amplification products after the virus was briefly cultured. [Burger et al.(1991)]. GenBank accession numbers M77228-M77230.
- 130) **US.SC**: This sequence is from the 1984 U.S. isolate SC, from an AIDS patient. [Gurgo et al.(1988)]. Accession number M17450.
- 131) **US.SF128**: This clone was isolated from the spinal cord tissue of a patient with dementia, after coculture with PBMCs. It is macrophage tropic, infecting macrophages but not T-cells [Lin et al.(1990)]. Accession numbers M95292 and M38673. The ability to infect macrophages versus HUT 78 cells was mapped to the region between a *StuI* site in env and a *XhoI* site in nef, by replacing this region in SF2 with the same region from SF128.
- 132) **US.SF162**: This sequence is from an infectious clone from the U.S. isolate SF162, cultured from the cerebrospinal fluid of a patient with toxoplasmosis. [Cheng-Mayer et al.(1990)]. Accession numbers M38428, M65024.
- 133) **US.SF2**: This sequence is from an infectious clone from the U.S. isolate ARV-2. ARV-2/SF2 was isolated from a patient with oral candidiasis after co-culture with mitogen-stimulated PBMCs in 1984 [Levy et al.(1984)]. [Sanchez-Pescador et al.(1985)]. Accession numbers K02007 and I07977. HIVSF13 (Accession number L07422) is a more infectious virus taken from the same patient five months later, when he had developed Kaposi's sarcoma and *Pneumocystis carinii* pneumonia [Cheng-Mayer et al.(1991)]. SF2 and SF13 are 98% identical to one another. The variation of SF2 in 9 years of infection in a chimpanzee has been studied by Fultz et al. unpublished, Accession numbers U56884-U56887.
- 134) **US.SF33**: This sequence is from an infectious clone from the 1984 U.S. isolate SF33. [York-Higgins et al.(1990)]. Accession number M38427.
- 135) **US.TN-ID#**: These eight sequences are from asymptomatic individuals identified after donating blood in Memphis, Tennessee, USA. [Slobod et al.(1994)]. Accession numbers U09140-U09175.
- 136) **US.twinABcom**: This sequence is a consensus of a set of 27 env sequences from a pair of heterozygotic perinatally HIV-1 infected twins who were observed during their first 2 years of life. Twin A remained asymptomatic through her first 2 years while twin B developed AIDS at six months and died at 22 months of age. Patient PBMCs were cocultured with donor PBMCs, prior to DNA extraction and PCR amplification. Approximately 500 copies of HIV proviral DNA per PCR reaction were used. PCR products were cloned prior to sequencing. Viral phenotypes from both infants, and all time points were also assessed. All were found to be non-syncytium inducing, but they differed in their ability to infect primary macrophages. Overall, it seems that the production of neutralizing antibodies by the healthy twin was the most important clinical factor. [Hutto et al.(1996)]. GenBank accession numbers U47562-U47588 for the envelope sequences and U47589-U47613 for tat sequences from the same blood samples.
- 137) **US.WEAU**: This sequence was kindly provided prior to publication by Sajal K. Ghosh. It is from a cultured isolate from a patient described as "patient 1" in [Clark et al.(1991)] and as WEAU 0575 in [Jr et al.(1993)]. The sequence is from a fully infectious complete, cloned genome. It has a high tropism for T-cells and is syncytium-inducing. Accession number U21135.
- 138) **US.WM**: This sequence, along with B_USJM, came from viral isolates after short term culture in PBMCs, PCR amplification, and cloning of PCR products. Both are from asymptomatic, seropositive individuals. [Ashkenazi et al.(1991)]. Accession number M80663.
- 139) **US.WMJ22**: This sequence is from the isolate WMJ22, isolated from a 4 year old female with AIDS, born in Florida to a woman of Haitian descent living in the U.S. Virus was cocultured on an immortalized T-cell line. [Hahn et al.(1986)] and [Starcich et al.(1986)]. Accession numbers M12507, K03457.
- 140) **US.WR27**: This sequence is from a complete genome. It represents first complete PCR-derived sequence of a U.S. clinical isolate of genotype B expanded only in primary PBMC. This provirus harbors a uniquely truncated V3 loop. WR27 was a patient with clinical progression to WR stage 5 when blood was drawn for viral isolation in 1988. This sequence is from a PCR clone from a primary isolate that was expanded in PBMC. The virus isolate had an SI phenotype. [Salminen et al.(1995)]. Accession number U26546.

Sequence Descriptions

- 141) **US.YU:** This is a consensus sequence of eight lambda phage clones and 12 PCR amplified clones derived from the uncultured brain tissue of a patient with AIDS dementia complex. A macrophage tropic clone (YU-2) is almost identical to the consensus sequence of YU in this region, with only a single amino acid change (K to N) 10 amino acids from the carboxy-terminal end of the sequence. [Li et al.(1991)]. Complete genomic sequences are available for two of the HIV1YU clones, along with biological characterizations of four of the HIV1YU clones: [Li et al.(1992)]. The GenBank accession numbers for the YU-10 and YU-2 complete genomes are M93259 and M93258, respectively. Accession numbers for the other clones are M89972-M89984.
- 142) **US1.HC-ID#:** These are forty 1990-1991 U.S. samples, from the study of the dentist who was thought to have been the source of HIV-1 infection of six of his patients. Only the dentist's viral sequence and the Florida control sequences are shown here; the six epidemiologically and genetically linked patients are excluded from this alignment because their viral sequences were very similar to the dentist's. All sequences were PCR amplified from patient PBMCs. Most are direct sequences from the amplification products, although some are consensus sequences of multiple clones of PCR products and one direct sequence. [Ou et al.(1992a)]. Accession numbers for the 75 sequences in this set: M90847-M90853, M90881-M90886, M90894-M90900, M90907-M90912, M90914-M90956, M90958-M90964, M92100-M92133, L22590-L22606 and U06872-U06919. See also [Korber & Myers(1992)], [Crandall(1995)], [Smith & Waterman(1992)], [DeBry et al.(1993)], [Palca(1992b)], [Palca(1992a)], [Abele & DeBry(1992)], [Hillis & Huelsenbeck(1994)], [Ciesielski et al.(1991)] and [Ciesielski et al.(1994)]. In a related study [Delwart et al.(1995)] heteroduplex mobility tracking assays were used to re-analyze samples from this set, again supporting the dental to patient transmission hypothesis.
- 143) **US2.D-ID#:** These 15 sequences are from the USA, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratory of Dr. Marcia Kalish at the the Centers for Disease Control, Atlanta, GA. The C2V3 region was directly sequenced from PCR amplification products of DNA from viral culture. The sequence ID numbers are abbreviated, for example D2US711 can be read as DAIDS sequence (D), isolated in 1992 (2), United States (US), patient 301711 (711). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. Full length envelope genes from some of these clones have been expressed, [Gao et al.(1996a)]. Accession numbers U04907-U04915, U04918 and U04921-U04925, U08448-U08452.
- 144) **US3.D-ID#:** These four sequences are from the US, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. gp160 sequences of clones from expanded culture stocks are available. The sequence ID numbers are abbreviated, for example D2US711 can be read as DAIDS sequence (D), isolated in 1992 (2), United States (US), patient 301711 (711). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. Accession numbers U08448-U08449 and U08451-U08452. Accession numbers for additional clones from these patients: U04916-U04917 and U04919-U04920.
- 145) **US4.ZhuPt-ID#:** Sequences from five primary seroconverters. Consensus, PCR-clones, peripheral blood PBMC DNA. Although sequences were also available from two of the donors, only sequences from five recipients are shown here. [Zhu et al.(1993)]. Accession numbers L21224-L21328, L21372-L21373, L21376-L21377, L21380-L21381, L21384-L21389, L21393-L21394, L21397-L21398, L21400-L21401, L21404, L21405, L21408, L21410-L21411, L21414, L21417-L21418, L21426-L21427, L21430-L21432, L21434-L21437, L21440-L21442, L21445, L21447, L21449-L21450, L21453-L21454, L21457, L21459, L21462-L21463, L21519-L21520, L21522-L21523, L21525-L21526, L21528-L21533, L21535-L21536, L21538-L21539, L21541, L21543, L21545-L21546, L21548, L21550-L21551, L21553-L21554, L21556, L21558-L21559, L21561-L21569, L21571-L21578, L21580, L21582, L21584, L21586, L21588 and L21590.
- 146) **US5.pt-ID#:** A consensus of PCR-clones, peripheral blood DNA. These two sequences were part of a study of blood sequences compared to brain sequences from six individuals. [Korber et al.(1994)]. Accession numbers U05360-U05368.
- 147) **US6.-ID#:** Each sequence is a consensus sequence of several cloned PCR products from PBMC proviral DNA from an individual infant. These sequences were part of a study of mother-infant transmission. Infant blood samples were taken from 1 week (infant 7) to 34 months (infant 4) post-partum. Dates of sample collection were: Infant 1, 10/25/91; Infants 2-4, 10/31/91; Infant 5, 2/6/92; Infant 6, 8/30/93;

- Infant 7 5/13/93. Maternal sequences were also reported as part of this set, but are not used in building these consensus sequences. [Ahmad et al.(1995)]. Accession numbers U16390-U16652.
- 148) **US7.-ID#:** Each sequence is a consensus sequence of several cloned PCR products from PBMC proviral DNA from an individual patient. These sequences were part of a study of early samples (1984-1986) from the San Francisco region. One of the samples (552-3) was HIV negative but was determined to be contaminated with blood from another (565-3) so the six samples from 552-3 and 565-3 are made into one consensus here (565). Another sample (552-5) from patient 552 was not contaminated, and it is presented as B_US11.552, because a larger region was reported in that publication. [Sabino et al.(1994b)] Accession numbers L20371-L20380. More V3 and tat sequences from these individuals are discussed in [Sabino et al.(1994a)]. Accession numbers U00243-U00399 and U01513-U01529.
- 149) **US8.-ID#:** These sequences were from a study of three recipients of contaminated blood. Recipient 1 (R1) and recipient 2 (R2) each received blood from different donors. A third recipient, not presented here, received blood from both donors. All three recipients were neonates. R1 received erythrocytes from donor 1 on 19 October, 1984 at the age of 3.5 weeks. For R1 the sequence of one of the two clones (2E) is presented here. R2 received erythrocytes from donor 2 on 24 September, 1984 at the age of 2 months. For R2 a consensus of 6 clones is presented here. Blood samples for this study were drawn in March 1986. R1 had slow weight gain, and R2 had lymphadenopathy at time of sample collection. [Diaz et al.(1995)]. Accession numbers U11188 = R1-2E, U11189 not used; U11203, U11196, U11199, U11192-U11194 = R2 six clones. The tat and envelope V4-V5 regions of clones from these same individuals are also available in U11173-U11178, U11180, U11205-U11209.
- 150) **US9.-ID#:** These are four consensus sequences for samples taken over a range of time from four different subjects. Blood samples for S1 were drawn in Nov '85, Jul '87, Jan '88 and May '89. Blood samples for S2 were drawn in May '85, Apr '87 and Oct '87. Blood samples for S3 were drawn in Jun '87 and Dec '87. Blood samples for S4 were drawn in Jan '85, Jan '89 and Jun '89. S2, S3 and S4 had decreasing CD4 counts during the study period. S1 had fluctuating CD4 counts. [McNearney et al.(1992)]. GenBank accession numbers L03430-L03453 and L23575-L23588 = S1; L03454-L03477 and L23618-L23633 = S2; L03478-L03490 and L23589-L23600 = S3; L03491-L03515 and L23601-L23617 = S4.
- 151) **US10.ID#:** These three consensus sequences are from sets of sequences used in a study on the dynamics of HIV sequence changes in vivo and the utility of heteroduplex analysis. All sequences were derived from PCR amplified PBMC DNA. The MA145 consensus represents sequences (GenBank accession numbers U00821, U00822, U00831-U00839) taken from an asymptomatic male from Massachusetts over a period of 4.5 years starting April 1989. Patient MA, from the US, was infected in 1984 or 1985, and had been experiencing neurological disorders prior to 1989 [Kusumi et al.(1992)]. The SFBI and SFPB consensus represent sequences (Accession numbers U13373-U13380 and U13381-U13388 respectively) taken from two patients with AIDS from San Francisco. [Delwart et al.(1994)]. Other sequences from patient MA can be found in database entries with accession numbers U00804-U00822, U00831-U00850, U00873-U00888, M79342-M79354 and M90025-M90046. Other SFBI sequences can be found in U13240-U13252. See also NL7.ID#.
- 152) **US11.ID#:** These four consensus and seven individual sequences came from patients early in infection, before, or around the time of seroconversion. Sequences for 306, 419, 349, and 074 are consensus sequences. [Shpaer et al.(1994)] and [Delwart et al.(1995)]. Accession numbers U23664-U23666, U23668, U23669, U23671-U23708, L20381. See also B_NL8.ID#.
- 153) **US12.ID#:** These six sequences were used in an investigation into the transmission of HIV-1 from one child (CHA), who had received zidovudine, to another child (CHB), who harbored a zidovudine-resistant strain. The presence of the zidovudine-resistant strain in child A and B, and the lack of such a strain in child 2's mother was used to show that child B was infected by child A and not by child B's mother. LC sequences are from children used as local controls. All sequences were derived from PCR amplified PBMC DNA. [Fitzgibbon et al.(1993)]. Accession numbers L12751-L12756, L19695, L19697, S66942. L12756 is listed as "isolate 100" in the databases, but seems to be the "group B consensus sequence" used for phylogenetic analysis.
- 154) **US13.ID#:** These three consensus sequences are from three IV drug users in Florida. Proviral DNA sequences were obtained from blood, cerebrospinal fluid and dorsal root ganglia from each of the three individuals. Sequences for V1-V5 of env, were PCR amplified, cloned and sequenced. [Shapshak et al.(1995)]. [Xin et al.(1995b)]. The sequence for patient 149 is a consensus of all 24 clones in

Sequence Descriptions

Accession numbers U16094–U16117. The sequence for 141 is a consensus of database entries labelled as being from patient 141 with the exception of: R5D, R6D, R7D, R2D and R4D, which were similar to IIB strains of HIV-1; and R1R, R3R, R7R, R8R and R9R, which were similar to samples from patient 144. The sequence for 144 is a consensus of database entries labelled as being from patient 144 with the exception of: R3R, R6R, R9R, R12R, R13R and L1D, which were similar to IIB strains of HIV-1; and C3D, C4D, C7D, C8D and C10D, which were similar to samples from patient 141. While infection of each individual with multiple strains of HIV (including one very similar to the IIB lab strain) is a possible explanation of these findings, we are only including one consensus sequence from each patient for this alignment. The authors are currently (1996) resequencing new samples from these patients. Accession numbers for patients 141 and 144 are U16032–U16093, U25191–U25261.

- 155) **US14.ID#:** These four sequences are from DNA from PBMC. [Mascola et al.(1994)]. Accession numbers L14573–L14576. See also B_TH2, E_TH2.
- 156) **US15.ID#:** These six consensus sequences are from a study of infants. Blood samples were collected from six infants over time. [Strannikova et al.(1995)]. Accession numbers U22682–U22810, U22834 and U22835.
- 157) **US16.ID#:** These two consensus sequences were used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. Both sets of sequences were from PCR amplified DNA from peripheral blood leukocytes. Patient ARTC1 was an asymptomatic individual from New York and ARTC3 was an AIDS patient from New York. [Pestano et al.(1995)]. Accession numbers U11586–U11594. See also A_UG1.964, C_UG1.45, and D_UG7.ID#.
- 158) **US17.ID#:** CB7 is one sequence from a set of 12 clones from two different samples (six clones each) from the same patient, collected in 1988 and 1990, plus two more clones as yet unpublished. The patient seroconverted in 1985. The patient did not receive any antiviral therapy until 1992. The patient's CD4 count was 1035 in 1988 and 807 in 1990. The patient's PBMCs were cocultured with donor PBMC for an unspecified length of time before cultured DNA was isolated and PCR amplified. Individual clones of PCR product were then sequenced. [Wang et al.(1995)]. Accession numbers U16324–U16335 and U19706–U19711. The other 8 sequences are individual clones from 8 different patients, all from the same clinic in Boston, MA. Individual CB7 was again included in this study. [Wang et al.(1996a)] [Wang et al.(1996b)]. Accession numbers U60152–U60162 and U27658–U27669.
- 159) **US18.ID#:** These 22 sequence are from a study of a HIV-1 infected dentist and many of his HIV-1 infected patients. In this case, no evidence was found for dentist to patient, patient to dentist, or patient to patient transmission. [Jaffe et al.(1994)]. Accession numbers U11454–U11490. Unfortunately the dentist sequences were not submitted to the databases.
- 160) **US19.ID#:** These two sequences are from a comparison of a mother who transmitted HIV to her infant and a mother who did not. One of the sequences (nontransmitter-283 accession number U07839) was 99.4% identical to the LAI strain of HIV-1 and is not included here. Sequences from the transmitter and nontransmitter were highly similar with the exception of sequence nontransmitter-217 (U07836). Thus a consensus of the similar sequences is included here as VA1-consensus, and the nontransmitter-217 outlier is presented separately as VA2. [Ayyavoo et al.(1996)]. Accession numbers U07833–U07841 and U07891–U07917.
- 161) **US20.ID#:** These three sequences are from a study of mother-infant transmission. Only child consensus sequences are used here, but mother and child clonal sequences are reported in [Roth et al.(1996)]. Proviral DNA was PCR amplified from uncultured patient PBMCs. Cloned PCR products were sequenced. Child 1 was 27 months old, child 2 was 5 months old and child 3 was 21 months old when blood was drawn for sequencing. Sequences, including those from the mothers, are in database entries with accession numbers U47745–U47807.
- 162) **US21.ID#:** These six sequences are from six different patients, all men from the Chicago, Illinois, USA MACS cohort. In the study, P1 & P2 were rapid progressors, P3 & P4 normal progressors, and P5 & P6 non-progressors. Their estimated years of seroconversion are as follows: P1, 1985; P2, 1985; P3, 1986; P4, 1986; P5, 1985; and P6, 1984. A total of 292 sequences of the C2-V5 region of envelope were completed. In the database entries the sample numbers are of the form Px.y-z, where x stands for the patient number, y is the number of months after the estimated date of seroconversion, and z is the clone number. Each of the six sequences presented here is a consensus of all clones from that patient. [Wolynsky et al.(1996)]. Accession numbers U35895–U36185.

- 163) **US22.ID#:** These 12 sequences are from a study of variability of *tat* and *env* genes. Lorenzo et al (1996) Unpublished. Accession numbers U57104–U57216. *Tat* sequences from these clones are also available (U57217–U57304).
- 164) **US23.ID#:** These 3 sequences are from a study of variability of *vif* and *env* genes. [Sova et al.(1995)]. Uncultured or short-term cultured PBMC proviral DNA was PCR amplified, and several clones from each PCR reaction were sequenced. Only one clone sequence from each patient is presented here. Accession numbers U50615–U50628. *Vif* sequences from these patients are also available (U41055, U41056, U41179–U41182 and U42229–U42282).
- 165) **US24.ID#:** These 5 sequences are from unpublished database entries by Schwartz et al. Accession numbers U49518–U49640, U51311–U51326 and U45860–U45876.
- 166) **US25.RM:** This sequence is from a blood recipient M who received a red blood cell pack from donor W in October 1984. Blood was sampled from recipient M for this sequence in January 1986, 15 months after the transmission event. Sequences from donor W (collected in January 1987, 27 months after transmission) and another recipient (O, collected in November 1986, 20 months after an RBC pack transfusion in March 1985) are also available [Delwart et al.(1995)]. Accession numbers U22826–U22828.
- 167) **US26.ID#:** These two sequences are from: A) a slow progressor, patient A, a 26 year old caucasian male who lost an average of 31 CD4+ cells/ml per year during the first 3 years of infection, and remains healthy; and B) a rapid progressor, patient B, a 38 year-old caucasian male who lost an average of 175 cells/ml per year during the first 3 years of infection. Both patients were infected from the same source but patient B picked up two divergent subtype B forms (one shared with patient A, and one unique to patient B). In the database entries, patient B is labelled as "14" and patient A as "13". [Liu et al.(1997)]. Sequences were determined by PCR from uncultured PBMC DNA and RT-PCR from plasma viral RNA. PCR products were cloned prior to sequencing. One representative clone from patient A was picked for the "13A" sequence presented here, and one of the divergent B sequences was picked for the "14B" sequence presented here. Accession numbers U56146–U56235, U79034–U79113.
- 168) **US27.ID#:** These 5 sequences are from infants infected perinatally and followed over time after birth as part of the Los Angeles Perinatal Transmission Study [Ganeshan et al.(1997)]. DNA sequences were determined from cloned PCR products from uncultured patient PBMCs.
- 169) **US28.ID#:** These 13 sequences are from studies of long-term survivors. Patient 3799 was transfused with HIV-1 infected blood in October, 1982 at the age of 30, and has remained asymptomatic, with very low viral load and fairly stable CD4 counts for over 13 years. The donor, as well as two other recipients of his blood, have all died from complications of AIDS. The husband of 3799, as well as two children born and breastfed after the 1982 transfusion, are all HIV-negative. [Michael et al.(1995)] and [Schwartz et al.(1997)]. The other patients showed varying rates of disease progression. During the study, only 101867 (sequence not available) died of AIDS. Sequences were PCR amplified directly from patient PBMC without coculture, cloned and sequenced. Only one clone from each patient is presented here. Accession numbers U60670–U60733. Entries with accession numbers U24443–U24487 are also from patient 3799, but not V3 region.
- 170) **US29.ME1:** This sequence is from a 40 year old patient in the Pittsburgh AIDS Clinical Trials Unit. [Chen et al.(1997)] describes a new method, called "progressive amplification" for obtaining full-length infectious molecular clones of HIV. In this study, two clones were generated and the *env* *gpl20* completely sequenced. One (ME1) was from the patient when asymptomatic, and both the viral isolate (367), and the molecular clone (ME1) grew on macrophages but not T-cells (Macrophage tropic) and did not show cytopathic effect on either MT-2 or PBMCs (NSI). The other (ME46) was from an isolate (828) taken from the same patient 20 months later, when the patient had AIDS. This isolate and the clone derived from it were able to grow on both macrophages and T-cells. It induced cytopathic effect in MT-2 cells and PBMCs (SI). Accession numbers U66221, U66222. The paper also reports on sequences from "patient P" in figure 5b, but those sequences are not available.
- 171) **US30.ID#:** These 10 sequences are from rapid and slow progressors who were studied in 1990-1994. The subjects were all asymptomatic and taking AZT (zidovudine) at the beginning of the study. Patients A, C, E, G and I were classified as slow progressors, while patients B, D, F, H and J were classified as rapid progressors. Two sequences appear to be mislabelled: U69410 labelled patient G clusters with patient E, and U69380 labelled as patient E clusters with patient G. Eight of the ten patient F, 1993 sequences are suspiciously similar (greater than 98% identical) to the HXB2R lab strain of LAI/IIIB. Each sequence is

Sequence Descriptions

a single clone from one patient, although several clones from each patient at two time points are available [McDonald et al.(1997)]. Accession numbers U69282-U69481.

- 172) **US31.ID#:** This study demonstrates the influence of three monoclonal antibodies IgG1b12, 2G12, and 2F5 to the HIV-1 envelope glycoprotein, and a tetrameric CD4-IgG molecule (Cd4-IgG2), for the ability to neutralize primary HIV-1 isolates from the genetic clades A through F and from group O. Each of the reagents broadly and potently neutralized B clade isolates [Trkola et al.(1996a)]. Accession numbers are U79721, U79720, U79719.
- 173) **UY1.ID#:** These 9 sequences are from 22-37 year old patients with CDC stage IV disease (except 726 at stage III). All are from Montevideo Uruguay. Four were homosexuals (270 352, 093 and 672), two were IV drug users (406 and 726), and the other 3 were heterosexuals (1699, 1193 and 376) [D. et al.(1996)]. Accession numbers U66414-U66422.
- 174) **VE.ID#:** These 8 sequences are from 8 individuals in Venezuela. Patient PBMCs were cocultured with donor PBMCs. Proviral DNA was harvested PCR-amplified. PCR products were directly sequenced. Nearly complete env gp120 sequences were determined, as well as pol gene sequences. [Quinones-Mateu et al.(1995)]. Accession numbers U16764-U16778, even numbers are env, odd numbers are pol.
- 175) **VN1.HCM9:** This sequence is from South Vietnam. [Menu et al.(1996)]. The sequence is from Ho Chi Minh city, from a woman infected by her HIV seropositive sexual partner who was thought to have been infected while traveling in Europe. Three other sequences in this study were found to be subtype E. Accession number U29209.
- 176) **ZA.ID#:** These 7 sequences are from 7 individuals in South Africa. ZA504 was from a 33 year old white male homosexual with AIDS and the virus was syncytium-inducing (SI). ZA508 was from a 32 year old white male bisexual with ARC and the virus was NSI. ZA509 was from a 30 year old white male homosexual with AIDS and the virus was SI. ZA524 was from a 49 year old white male bisexual with AIDS and the virus was NSI. ZA510 was from a 29 year old white male heterosexual with ARC and the virus was SI. ZA512 was from a 26 year old white male homosexual with ARC and the virus was SI. ZA513 was from a 3 year old black male blood transfusion recipient with AIDS and the virus was SI. All samples were collected at the Tygerberg Hospital in the Western Cape region of South Africa between 1984 and 1992. DNA was harvested from cocultured PBMCs and the env gene was PCR amplified and cloned into pBSKS+ for sequencing. Each sequence is from a single cloned PCR product. [Engelbrecht et al.(1995)]. Accession numbers L48063-L48066, L48069, L48071 and L48073. Database entries U33770 and U33774-U33779 are shorter env gene fragments from these same clones.

C Subtype

At this time there are viral sequences from 233 HIV-1 infected individuals associated with HIV-1 subtype C. The C subtype consensus sequence (C_CONSENSUS_97) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

- 1) **BR.W2BR025:** This sequence is part of a gp160 sequence from an asymptomatic individual from Brazil, sampled in 1992. A clone was derived from an expanded viral culture, expressed and sequenced. This sequence was provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences from this patient can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. Accession number U15121. This sequence is from a 23 year old male hemophilia patient from Porto Alegre, Brazil. He had seroconverted more than 1.2 months prior to the date this blood sample was collected in 1992. He was asymptomatic, and had not taken any anti-retroviral therapy prior to sampling. The HIV isolate exhibited an NSI phenotype, when assayed by the WHO [Gao et al.(1996a)]. 92BR025 is similar to another WHO sample; 91BR015. Accession number U52953. Entries with accession numbers U08720, U08785, U09126, U09132 and U09133 are also from W2BR025.
- 2) **BR1.HSP203** Although this sequence is listed as unpublished in the database, it seems to be an extension of work published in [Morgado et al.(1994)], [Sabino et al.(1994c)] and [Sabino et al.(1996)]. It is from San Paulo, Brazil. Accession number U31585.
- 3) **BR2.91BR015:** This sequence is from Brazil. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. 91BR015 is similar to another WHO sample; 92BR025. [Penny et al.(1996)]. Accession numbers U39234 and U39238.
- 4) **BU1.ID#:** These eight sequences are from Burundi. They are part of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. 91BU009 groups with subtype D in a neighbor-joining tree of the V3 region and is presented as an undetermined subtype in the section of the compendium. [Penny et al.(1996)]. Accession numbers U39252 and U39233, 91BU001; U39248 and U39237, 91BU002; U39239 and U39242, 91BU003; U39241 and U39243, 91BU004; U39240 and U39257, 91BU005; U39244 and U39246, 91BU006; U39245, U39247 and U39249, 91BU007; U39250 and U39251, 91BU008; U39253 and U39254, 91BU009.
- 5) **BY1.ID#:** These 3 sequences are from Byelorussia. [Lukashov et al.(1995)]. Accession numbers: L38410, BLR9A; L38409, BLR8A; L38408, BLR5A.
- 6) **CY.HO021:** This is a sequence from a 51 year old woman whose husband had died of AIDS. She was born and lived in Zambia, before moving to Cyprus. She was asymptomatic, with a CD4 count of 200, and she had been seropositive for at least 6 years. This sample, like others in this study (see also subtypes A, B, F and I) was collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. DNA was extracted from patient PBMCs and PCR amplified. After a second round of PCR, products were cloned and sequenced. Two clones from patient 02 were sequenced. [Kostrikis et al.(1995)]. Accession numbers U28321 and U28661.
- 7) **DJ1.DJ-ID#:** These two sequences from Djibouti were from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989-1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession numbers L22940 and L23065.
- 8) **ET2.ID#:** These nine sequences of Ethiopian isolates were part of a cohort considered to be heterosexually infected. PBMC DNA was PCR amplified and directly sequenced. [Salminen et al.(1996)]. For patient 2220, who had slim disease and AIDS, almost full length HIV-1 genome was PCR amplified from PBMC DNA and cloned. Proviral DNA from others in the cohort was PCR amplified and directly

Sequence Descriptions

- sequenced. Other regions of the genome are available for these isolates as well. [Salminen et al.(1996)]. Accession numbers U45481–U45502 (V3 region), U15060–U15066 and U45503–U45504 (LTR NF-kB/NRE regions), M64001–M64009 (gag p7 region), M64015–M64018 (env gp41 region) and U46016 (C2220 complete genome).
- 9) **ET3.ID#:** These 93 sequences are from a study that describes the distribution of HIV-1 subtypes in Ethiopia. HIV-1 RNA was collected from sera (from a majority of asymptomatic individuals) and a 284bp fragment covering the V3 region was amplified by RT-PCR and directly sequenced. All sequences were subtype C except for one subtype A [Abebe et al.(1997)]. Accession numbers U88727–U88755, U88757–U88821.
 - 10) **FR1.ID#:** These 14 sequences are from members of the French military who are believed to have been infected while deployed outside of France (Djibouti). An additional sequence (FRMP040, U58787) was listed as subtype C in the paper, but clustered with subtype B in analysis done at the HIV Database and it has been listed as subtype unclassified until more information is available. Other sequences from this study were subtypes A, B, E, and F [Lasky et al.(1997)]. Accession numbers for subtype C were U58785, U58786, U58788–U58799.
 - 11) **GA1.ID#:** These 2 sequences are from Gabon. G134 is from a 1988 or 1989 sample from a patient with AIDS living in Franceville, Gabon. LBV105 is from a 1988 sample from an asymptomatic individual sampled from the general population of Libreville. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Delaporte et al.(1996)]. Accession numbers X90912, G134; X90913, LBV105. See also subtypes A, D, F, G and O sequences from this same study.
 - 12) **GB1.00513:** This sequence is from the British isolate 93–00513. [Arnold et al.(1995c)]. Accession number U21099.
 - 13) **GM.GM3:** A sequence from Gambia, as yet unpublished. Bobkov et al. 1996 unpublished. Accession number U33098. See also Gambian sequences of subtypes B and J.
 - 14) **IN1.D-ID#:** These four sequences are from samples from high risk patients in India, PCR clones, DNA, PBMC culture. [Dietrich et al.(1993)]. Accession numbers L07651 and L07653–L07655; X65638–X65640 and X68406.
 - 15) **IN2.D-ID#:** These five sequences are from samples from high risk patients in India, primarily stage I. They were nested PCR amplified from DNA obtained from uncultured PBMC from patients serologically defined as HIV-1/HIV-2 mixed infections. [Grez et al.(1994)]. Accession numbers U07098 and U07100–U07103.
 - 16) **IN3.ID#:** These 8 sequences were isolated from Pune and New Delhi, India. All 8 sequences were from heterosexually infected patients from New Delhi, or Pune, India. DNA was isolated from cocultured PBMCs after one week of culture. PCR product was cloned and a single clone was sequenced. [Tripathy et al.(1996)]. Accession numbers U29179, U29694–U29698, U31362 and U31363. See also B_IN.ID#.
 - 17) **IN4.ID#:** These 24 sequences are from 1992 dried blood spot samples from Vellore near Madras, in Tamil Nadu state in southern India. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. Samples came from previously identified HIV seropositive STD patients (1, 3, 5, 6, 7, 11, 12, 13, 19, 20, 23, 29, 33), spouses of infected men (2, 4, 10, 16, 22, 26, 27), female prostitutes (36, 37) or bisexual men (8, 32). Two homosexual men from the same region had subtype A HIV-1 (see A_IN1.9 and 14). [Cassol et al.(1996)]. Accession numbers U53278–U53285, U53287–U53290, U53292–U53303.
 - 18) **IN5.ID#:** These 7 sequences are from India. Another sequence from this publication was subtype A (see A_IN2). [Tsuchie et al.(1993)]. Accession numbers D13420–D13424, D13426, D13427.
 - 19) **KE.NA113:** This sequence was derived from a patient who was part of a May–June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2–V3 region was amplified by RT-PCR. The PCR product was cloned and 20 clones from the patient were sequenced. Seven other patients from this study had viral subtypes A and D. [Zachar et al.(1996a)]. Accession number U33762.
 - 20) **MW.SM750:** This sequence is from cloned PCR amplified cocultured PBMC DNA. SM750 was a black male sampled in the gold mines in Malawi in 1989. [Becker et al.(1995)]. Accession number U06719.

- 21) **MW1.ID#:** These 13 sequences are from pregnant women with risk factors from Malawi. PCR-direct, peripheral blood DNA. [Orloff et al.(1993)]. Accession numbers L07427-L07441, L15721-L15735.
- 22) **MW2.D-ID#:** These two sequences are from individuals from Malawi, generated as part of the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. The C2V3 was excised from full gp160 sequences, derived from clones from expanded culture stocks. The sequence ID numbers are abbreviated, for example D3MA959 can be read as DAIDS sequence (D), isolated in 1993 (3), Malawi (MA), patient 301959 (959). Accession numbers U08453-U08454.
- 23) **MY.ID#:** These 2 sequences are from IV drug using prisoners in a prison in Kuala Lumpur, Malaysia. PCR products amplified from uncultured PBMCs were directly sequenced. Both of these prisoners had received medical treatment in India which included blood transfusion and organ transplants, and it is likely that they were infected in India. [Brown et al.(1996)]. This report also included subtypes B and E in Malaysia. Accession numbers U65549-U65550.
- 24) **NL1.ID#:** These 2 sequences are from a Dutch woman whose partner was a recent immigrant to The Netherlands from Zaire, and from a recent immigrant to The Netherlands from either Zambia or Zaire. The first two letters of the ID# represent the two letter country code for the previous residence of the patient (UN = unknown). The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. The sequence from the Dutch/Zaire patient is not unequivocally a sub sequence, it may be subtype A or unclassifiable. [Lukashov et al.(1996)]. Accession numbers L76908, NL9402418; L76898, UN9305091.
- 25) **NO1.ID#:** These 4 sequences are from Norwegian patients who were part of the Oslo HIV cohort study [Engelstad(1996)]. Uncultured PBMC DNA was PCR amplified in two nested PCR reaction steps. PCR products were directly sequenced. Where two peaks of equal height were observed at a single position, IUPAC ambiguity codes were used. Health, sex, year of sample (1989-1992), and risk group (IVDU, Het, Homo, Hemo) for each patient were noted in a table in the publication. The four subtype C sequences were all from heterosexuals who had sex with persons from Zambia (3) or South Africa (1). Thirty-six subtype B sequences were also part of this set (see B_NO1.ID#). Accession numbers X92913, X92914, X92917 and X92918.
- 26) **RU.ID#:** These 4 sequences are from Russia. [Lukashov et al.(1995)]. Accession numbers: L38418, RUS20A; L38404, RUS2A; L38406, RUS1A; L38414, RUS13A.
- 27) **RU1.YAN4:** This sequence is from Russia. Bobkov et al. 1996 Unpublished. Accession number U33109.
- 28) **RW1.ID#:** These two sequences are consensus sequences of many clones from mothers infected in Rwanda. Mulder-Kampinga et al., Unpublished (1996) and [Kampinga et al.(1997)]. Mother 566 was apparently dually infected with HIV-1 subtypes A and C. See also A_RW4.ID#. Accession numbers for mother 566 are Z76183, Z76188-Z76197, Z76284-Z76293, Z76295-Z76299, Z76301, Z76303-Z76311, Z76322-Z76342 and Z76725-Z76732; for mother 134, Z76039-Z76042 and Z76044.
- 29) **SN.SE364:** A Senegalese sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989-1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession number L22944.
- 30) **SO.1574:** This sequence is a consensus sequence of blood and CSF samples taken from the Somali patient 1574, CDC classification II. [Keys et al.(1993)]. Accession numbers Z23188, Z23190-Z23191, and Z23228-Z23231.
- 31) **SO.SM145:** A Somali sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989-1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession number L22946.
- 32) **TW1.252** This sequence is one of a set of sequences from Taiwan. Other sequences in the set were subtypes B, E, F or G. [Chang et al.(1997)]. Accession number U73055.
- 33) **TZ1.1588:** This sequence is from the Mara region of rural
- 34) **UG.45:** A single sequence used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. The sequence was derived from PCR amplified DNA.

Sequence Descriptions

- from peripheral blood leukocytes. The patient was an asymptomatic individual from Uganda. [Pestano et al.(1995)]. Accession number U11597. See also A_UG1.964, B_US17.ID#, and D_UG7.ID#.
- 35) **UG.UG268:** A Ugandan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession number L22948.
- 36) **ZA.NOI:** This sequence is from a South African individual who was part of a study of HIV-1 strains in India. This sequence was found to be closer to the Indian sequences, than are other isolates from Africa. PCR amplified DNA from PBMC cultures were sequenced. [Dietrich et al.(1993)] and [Becker et al.(1995)]. Accession numbers L07426 and U06716. See also C_IN3.ID#, B_IN.IND9.
- 37) **ZA1.ID#:** These 2 sequences are from 2 individuals in South Africa. ZA514 was from a 59 year old mixed-race male heterosexual with AIDS and the virus was NSI. ZA517 was from a 33 year old mixed-race male heterosexual with ARC and the virus was NSI. The samples were collected at the Tygerberg Hospital in the Western Cape region of South Africa between 1984 and 1992. DNA was harvested from cocultured PBMCs and the env gene was PCR amplified and cloned into pBSKS+ for sequencing. Each sequence is from a single cloned PCR product. [Engelbrecht et al.(1995)]. Accession numbers L48067–L48068. Database entries U33781 and U33782 are shorter env gene fragments from these same clones. See also B_ZA and D_ZA sequences from this same study.
- 38) **ZA2.ID:** These 3 sequences are from clones from PCR amplified cocultured PBMC DNA. Dlu was a black male sampled at the Tygerberg Hospital in Cape Town, South Africa in 1990. Gom was a black male sampled at the Tygerberg Hospital in Cape Town, South Africa in 1990. BooyD was a mixed-race female mother sampled at the Tygerberg Hospital in Cape Town, South Africa in 1990. A short sequence from a seropositive child of BooyD is presented in entry U07015. [Becker et al.(1995)]. Accession numbers U07237, U06717 and U06718.
- 39) **ZM1.ZAM-ID#:** Two sequences from Zambia, from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession numbers L22954 and L22956.
- 40) **ZW.2647:** This sequence is a consensus sequence taken from blood and CSF samples taken from Zimbabwe patient 2647, CDC classification II. [Keys et al.(1993)]. Accession numbers Z23196–Z23199 and Z23236–Z23239.

D Subtype

At this time there are viral sequences from 119 HIV-1 infected individuals associated with HIV-1 subtype D. The D subtype consensus sequence (D_CONSENSUS_97) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

- 1) **BI.BU009con**: This is a consensus sequence from 8 clones from a 36 year old female with CDC stage IV AIDS and pulmonary tuberculosis, from Bujumbura, Burundi. PBMCs from the patient were cocultured with donor PBMCs for 2 weeks, at which time large syncytia formed and the culture became antigen positive. The env region was then PCR amplified by nested PCR and cloned. [Ranjbar et al.(1995)] and [Ranjbar(1995)]. Accession numbers L35452-L35459.
- 2) **BR.RJ100** This sequence is from Rio De Janeiro, Brazil. Morgado et al. unpublished, 1997. Accession number AF000238.
- 3) **CE4020**: This sequence was the only D subtype from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990-1991. It is a consensus from cloned PCR products, derived from cultured proviral DNA. [Murphy et al.(1993)]. A full gp120 sequence from this isolate was kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg, France. It is a part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system [Richalet-Secordel et al.(1994)]. The year of isolation and health status of individuals from which the viruses were isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. Sequences of clones from sample 4093 (Accession numbers L11517, L11517) which was taken from the same patient as 4020 one year later, are closely related and are not included here. Accession numbers L11472-L11473, U43138.
- 4) **CI.CI-13**: A single D subtype sequence from a set of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-13 was symptomatic, and serologically dually reactive for HIV-1 and HIV-2. The C2V3 region is part of a 900 bp fragment that was sequenced for each individual. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3-4 clones were sequenced, and the consensus of those clones is presented here. [Janssens et al.(1994a)]. Accession numbers X72028-X72029.
- 5) **CIL.ID#**: These 2 sequences are from Abidjan, Ivory Coast. No other information is yet available. Ellenberger D.L. unpublished 1997. Subtypes A, D and G were found for Ivory Coast patients in this set. Ugandan sequences of subtype A were also part of this set. Accession numbers AF000469, AF000451.
- 6) **CM.CMR61D**: CMR61 is one of two D group sequences from a 23 year old female commercial sex worker from Cameroon who was found to be triple-infected with subtypes A and D, as well as O group HIV-1 [Takehisa et al.(1997b)]. The other sequence from the same patient is labelled as CMR709 and is from a separate blood sample. Accession numbers U58149, U58151.
- 7) **D_CM2.ID#**: These four sequences are all from Cameroon. Takehisa et al unpublished-1997. Accession numbers U70013, U70012, U70011
- 8) **GA1.ID#**: These 2 sequences are from 1988 or 1989 samples from patients with AIDS living in Franceville Gabon. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced, [Delaporte et al.(1996)]. Accession numbers X90919, G109; X90920 G141. See also subtypes A, C, F, G and O sequences from this same study.
- 9) **GB1.CPHL4**: This sequence is a consensus from the British isolate 93-43424, clones 2, 6 and 35. It was referred to as 93-43424 in [Arnold et al.(1995c)] and as CPHL4 in [Arnold et al.(1995a)]. Accession numbers U21098 (clone 35) and U23121-U23122 (clones 2 and 6 respectively). CPHL4 is a female who is believed to have contracted the virus from CPHL5 through heterosexual contact. Sequences from CPHL5 are not included in this alignment due to this epidemiological relationship.
- 10) **GH.GH3**: This sequence is from Ghana. Subtypes A and an A/G recombinant were also detected in this study [Takehisa et al.(1997a)]. Accession number U67049.
- 11) **GH1.ID#**: These 2 sequences are from Ghana. Subtypes A and G were also detected in this study [Ishikawa et al.(1996)]. Accession numbers for subtype D are U67682 and U67684.

Sequence Descriptions

- 12) **KE.NA116:** This sequence was derived from a patient who was part of a May-June 1992 study of pregnant women from the Pumwani Maternity Hospital in Nairobi, Kenya. Viral RNA was concentrated from patient serum just prior to delivery, and the envelope C2-V3 region was amplified by RT-PCR. The PCR product was cloned and 20 clones from the patient were sequenced. Seven other patients from this study had viral subtypes A and C. [Zachar et al.(1996a)]. Accession number U33765.
- 13) **KE1.KEN-ID#:** These three patients were part of a 1990-1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced. [Janssens et al.(1994c)]. Accession numbers for sequences from the entire set of 23 patients studied in this publication: U12984-U13006.
- 14) **NL.A11:** This sequence is from a Dutch study of presumed HIV-1 donor-recipient pairs. This sequence is from a recipient at the time of seroconversion; the donor was a Zairean woman living in the Netherlands (patient A12 accession numbers M91840-M91848). The sequences from both donor and recipient were extremely similar, so only the recipient (patient A11) is shown here. This sequence is a consensus sequences of multiple clones from PCR amplified serum RNA. [Wolfs et al.(1992)]. Accession numbers M91849-M91856.
- 15) **NL2.ID#:** These 7 sequences are from recent immigrants to The Netherlands from various countries. The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. [Lukashov et al.(1996)]. Accession numbers L76900, UG9307184; L76907, ZR9402261; L76892, ZR929193; L76904, UG9401525; L76895, AO9302187; L76872, ZR891183; L76876, ZR901100.
- 16) **RURUS14A** This sequence is from [Lukashov et al.(1995)]. Accession number L38415.
- 17) **SN.SE365:** A Senegalese sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989-1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession number L22945.
- 18) **TZ1.TAN-ID#:** These ten sequences are from a set of 14 Tanzanian samples from symptomatic individuals, using serum samples taken in 1988 to generate PCR clones from viral RNA for sequencing. [Zwart et al.(1993)]. Accession numbers L01298-L01339.
- 19) **TZ2.ID#:** These eight sequences were from patients at a clinic in Dar es Salaam, Tanzania. The individuals from which the virus was cultured showed clinical signs of AIDS, and the year of viral isolation was 1988. Viral cDNA was PCR amplified from donor PBMC, and one cloned PCR product per donor was sequenced. [Stwka et al.(1994)]. Accession numbers U12405, U12407, U12410-U12415.
- 20) **TZ3.ID#:** These 4 sequences are from the Mara region of rural northwest Tanzania. [Robbins et al.(1996)]. Subtype A was also found in this study. Accession numbers U61875 and U61879-U61881.
- 21) **UG.U44342:** This sequence is from a Ugandan. Consensus of PCR-clones, peripheral blood DNA. Intact env sequences are available from this sample. [Bruce et al.(1993)]. Accession numbers M98408-M98416.
- 22) **UG.UG23:** This sequence is from blood collected from the Mulago Teaching Hospital in Kampala, Uganda. Viral RNA was harvested after 10-14 days of coculture with donor PBMCs and reverse-transcribed with AMV-RT. The env V3 region was PCR amplified and cloned. This sequence is from an individual clone. [Adkin et al.(1993)], [Pestano et al.(1993)]. Accession number M98504.
- 23) **UG1.W2UG-ID#:** Twelve sequences from asymptomatic individuals from Uganda in 1992. Each sequence is a consensus from cloned PCR products derived from cell-cultured proviral DNA or culture supernatant RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. [De Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994c)]. Accession numbers U08721-U08741, U08786-U08787, U08803-U08809, U08821-U08824, U27399, U43386. Two sequences not reported to be related to these Ugandan isolates are suspected to represent lab contamination of PCR reactions with these isolates: U58827 (FRMP153) is 99% identical to 92UG046; and L38415 (RUS14A) is 99% identical to 92UG024.
- 24) **UG2.ID#:** These twelve sequences are from 1986-1987 Ugandan samples. Each sequence is a consensus from cloned PCR products derived from uncultured proviral DNA harvested directly from patient PBMCs. [Oram et al.(1991)]. No database entries exist for these sequences.

- 25) **UG3.ID#:** These 11 sequences are part of a set of sequences derived from 22 Ugandans who were attending an AIDS clinic. Blood samples were obtained in 1990. Each sequence is a consensus from cloned PCR products derived from uncultured proviral DNA harvested directly from patient PBMCs. [Albert et al.(1992)]. Accession numbers L00614-L00618, L00733-L00737, M98894-M98899, M98901, M98906-M98907, M98911-M98913, M98918, M98920-M98923, M98929-M98937, M98942-M98945 and M98967-M98975.
- 26) **UG4.ID#:** One to 4 clones of each these Ugandan isolates were sequenced, but only one clone is shown here. [Douglas et al.(1996)]. Other Ugandan isolates sequenced in this study were subtype D/A recombinant. London subtype B clones were also reported. Complete envelope gp160 sequences were reported for all isolates. Accession numbers U36867, U36868, U36871, U36884-U36887.
- 27) **UG5.UG-ID#:** Three Ugandan sequences from a set of HIV-1 viral isolates from Africa. Health status of the individuals from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989-1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. Accession numbers L22947 and L22949-L22950.
- 28) **UG6.ID#:** Three Ugandan sequences from a set of HIV-1 viral isolates from Africa. All three individuals from which the virus was cultured had AIDS, and the year of viral isolation was 1987. Viruses were cultured with HUT-78 cells for an unspecified length of time. The V3 region of env (gp160) was amplified, cloned and sequenced. [von A. et al.(1995)]. Accession numbers U15005, U15006 and U15007.
- 29) **UG7.ID#:** These sequences were used in a study of the impact of sequence variation on the distribution and seroreactivity of linear antigenic epitopes. Both sets of sequences were from PCR amplified DNA from peripheral blood leukocytes. All patients were asymptomatic individuals reporting for regular blood drawing at the Nakasero blood transfusion service, Kampala, Uganda. [Pestano et al.(1995)]. Accession numbers U11595, U11596, and U11598. See also A_UG1.964, C_UG1.45, and D_UG7.ID#.
- 30) **UG8.#:** These 2 sequences are from Gulu, northern Uganda. They are from direct sequence of PCR product amplified from uncultured PBMCs. Blood samples were drawn from 217 pregnant women attending the clinic in Gulu, northern Uganda. Ages ranged from 17 to 37 years. The 29 seropositive women (13.4%) were all asymptomatic. [Buonaguro et al.(1995)]. Database accession numbers U44881 and U44884. Eight subtype A sequences were also found in this study (see A_UG5).
- 31) **US.AMK:** This sequence comes from a student living in Alabama, who moved from Zaire to the US in 1988. Virus was isolated from the patients PBMCs; this isolate was PCR amplified, and amplification products from both gag (U08192) and env were subcloned and sequenced. His CD4 count was < 5 cells/mm³, and he was symptomatic at the time of viral isolation. [Gao et al.(1994c)] and [Gao et al.(1996a)]. AMK is also known as 93ZR001. Accession numbers U08193, U27419.
- 32) **ZA.ID#:** These 5 sequences are from 5 individuals in South Africa. ZA500 was from a 41 year old white male homosexual with ARC and the virus was non-syncytium-inducing (NSI). ZA501 was from a 24 year old white male bisexual with AIDS and the virus was SI. ZA505 was from a 36 year old white male homosexual with AIDS and the virus was SI. ZA506 was from a 33 year old white male homosexual with AIDS and the virus was SI. ZA507 was from a 37 year old white male homosexual with AIDS and the virus was SI. All samples were collected at the Tygerberg Hospital in the Western Cape region of South Africa between 1984 and 1992. DNA was harvested from cocultured PBMCs and the env gene was PCR amplified and cloned into pBSKS+ for sequencing. Each sequence is from a single cloned PCR product. [Engelbrecht et al.(1995)]. Accession numbers L47608, L48061, L48062, L48070 and L48072. Database entries U33769, U33771-U33773 and U33780 are shorter env gene fragments from these same clones. See also B_ZA and C_ZA1 sequences from this same publication.
- 33) **ZR.ELI:** This sequence is from the Zairean isolate ELI. [Alizon et al.(1986)] and [Goodenow et al.(1989)]. The complete genomic sequence and an infectious clone are available. In the 1995 Compendium (pages III-45 and III-47), ELI was listed as a possible D/A mosaic, with gp41 being A-like. Accession numbers M27949, K03454 and X04414.
- 34) **ZR.JY1:** This sequence is from Zairean isolate Z-84, clone JY1. [Youno et al.(1988)]. Accession number J03653.
- 35) **ZR.MAD:** This sequence is from an asymptomatic Zairean woman who was seropositive for HIV-1 by several French-approved HIV tests and by HIV-1 western blot. The viral sequences obtained for

Sequence Descriptions

- the env V3 region and gp41 show that this isolate belongs to M-group subtype D. [Cohen et al.(1995)]. Accession number X83216. A region of gp41 is also sequenced, see X83215.
- 36) **ZR.MAL**: This sequence is from a non-infectious clone of the Zairean isolate MAL. [Alizon et al.(1986)]. The complete genomic sequence and an infectious clone from the isolate MAL are available. MAL is known to be recombinant between subtypes A, D and I. Accession numbers K03456 and X04415.
- 37) **ZR.NDK**: This sequence is from an infectious clone of the Zairean isolate NDK. The molecular clone is highly cytopathic in vitro. [Spire et al.(1989)]. The complete genomic sequence is available. Accession number M27323.
- 38) **ZR.Z2Z6**: This sequence is from an infectious clone of Zairean isolate Z2. Theodore T, and Buckler-White A, unpublished. The complete genomic sequence is available. Accession number M22639. See also [Srinivasan et al.(1987)]. The database entry with accession numbers K03458 and M16322, which is from the same isolate.
- 39) **ZR1.JD#**: These four sequences are part of a set of 14 A and D sequences from women from Zaire. 8 were healthy, 4 showed minor signs of illness, and 2 had AIDS. Sequences were determined by directly sequencing PCR products derived from uncultured proviral DNA harvested directly from patient PBMCs. [Potts et al.(1993a)]. Accession numbers L19623, L19627, L19631 and L19635.

E Subtype

At this time there are viral sequences from 135 HIV-1 infected individuals associated with HIV-1 subtype E. The E subtype consensus sequence (E_CONSENSUS_97) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

- 1) **CE90CF402** 90CF402, previously named CAR-E 4002 or 90CR402, was obtained from a man from Bangui, Central African Republic, who had lymphadenopathy, diarrhea, severe weight loss and recurrent respiratory infections. He was infected through heterosexual contact, but the year of infection is unknown. The virus was first adapted to growth in chimpanzee cells, expanded in chimpanzee cells, and then re-expanded in human PBMCs before lambda cloning and sequencing. [Gao et al.(1996b)]. The complete genome is found with Accession number U51188.
- 2) **CF1.ID#**: These eight sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990-1991. Consensus, PCR-clones, cell culture, DNA. [Murphy et al.(1993)]. Accession numbers L11459-L11460, L11463-L11468, L11476, L11480-L11481, L11504-L11505, L11511-L11513, L11519-L11521 and U43137. Another sequence from patient 4039 is found with accession number U43112.
- 3) **CF2.ID#**: These three sequences were kindly provided prior to publication by Dr. M.P. Kieny of Transgene, Strasbourg Cedex, France. They are part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C. Mathiot and B. You (Pasteur Inst., Bangui), grown by F. Barre-Sinoussi and A. Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D. Schmitt and M.P. Kieny. Database accession numbers U43110, U43170 and U43173.
- 4) **CM.CA10**: A single E subtype sequence from a set of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals, specifically, CA10 was symptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. [Nkengasong et al.(1994)]. Accession number X80439.
- 5) **CM.CMR10**: This sequence is from Cameroon. Takehisa et al unpublished-1997. Accession number U69991
- 6) **FR1.ID#**: These six sequences are from members of the French military who are believed to have been infected while deployed in Cambodia between 1992 and 1995. Other sequences from this study were subtypes A, B, C, F, and unclassified. [Lasky et al.(1997)]. Accession numbers for subtype E were U58779-U58784.
- 7) **GB1.11643**: This sequence is from the British isolate 94-11643. The sequence was determined from PCR-amplified lymphocyte DNA. The gag gene from this isolate was subtype A, as is the gag gene from all subtype E virus studied to date. The patient is thought to have contracted the virus in Thailand, but currently lives in the United Kingdom. [Arnold et al.(1995c)]. Accession number U21109.
- 8) **ID1.ID#**: These 7 sequences are from a set of 14 sequences from Indonesia [Porter et al.(1997)]. Accession numbers U68189-U68194. Subtype B was also identified in 7 other samples from Indonesia in this study. PCR products were directly sequenced from either uncultured PBMC DNA or cocultured PBMC DNA.
- 9) **JP.JH23B**: This sequence is from a Japanese patient dually infected with HIV-1 subtypes B and E [Xin et al.(1995a)]. Accession number D67090. See also D67089, B_JP.JH23A.
- 10) **JP1.ID#**: These three sequences are from one Japanese and two Thai individuals living in Japan, obtained by direct sequencing of PCR-amplified proviral DNA from peripheral blood mononuclear cells. [Weniger et al.(1994)]. NIH3J is from a male Japanese national. NIH2T and NIH4T are from Thai female prostitutes, living in Japan. Accession numbers L32085-L32087.
- 11) **JP2.Pat43**: This sequence is from a study of rapid versus slow progressors in Japan. Patient 43 was a rapid progressor infected through sexual contact in 1992 after testing negative in 1991, [Shioda et al.(1997)]. Other sequences in this study were subtype B. Accession numbers AB002975-AB002987.

Sequence Descriptions

- 12) **MM.1786**: This sequence is from a dried blood spot collected in 1992 from a Female STD patient in Myanmar. DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. Accession number U53309.
- 13) **MY.9214103**: This sequence is from a Thailand infant who was adopted by Malaysians, and now lives in Malaysia. PCR products amplified from uncultured PBMCs were directly sequenced. [Brown et al.(1996)]. This report also included subtypes B and C in Malaysia. Accession number U65551.
- 14) **NL.TH94037**: This sequence is from a recent immigrant to The Netherlands from Thailand. The blood sample was collected in 1994. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR product was directly sequenced. [Lukashov et al.(1996)]. Accession number L76911.
- 15) **RU1.1164con**: This is a consensus of 3 clones from a single patient RU1164 from Russia. [Bobkov et al.(1997a)]. The patient was a 21 year old asymptomatic commercial sex worker who first tested HIV positive in April, 1996. Blood was collected for this sequence in May, 1996. She reported having travelled to Singapore in 1994-1996 and having multiple sexual clients there. DNA from uncultured PBMCs was PCR amplified, cloned and sequenced. Patient RU1164 lives in Khabarovsk kray in the far eastern part of the Russian Federation. This is the first report of subtype E in Russia. Accession numbers U93607, U93608, U93609.
- 16) **TH.93TH253**: This sequence is from a 21-year-old man from Chiang Mai, Thailand and was previously named CMU010 or 302053. The patient had end-stage AIDS. The mode and year of infection are unknown. 93th253 was isolated in 1993 and expanded in human PBMCs, then expanded in H9 cells, and followed by lambda cloning and sequencing. The complete genome has been sequenced. [Gao et al.(1996b)]. Accession number U51189.
- 17) **THC18**: This sequence is derived from an HIV-1 infected mother enrolled in Bangkok in a perinatal transmission study. Her plasma was screened for gp120 binding antibody, CD4/gp120-binding inhibition, to subtype B (MN/H9), and Thai E (SL7/SupT1) using native gp120 antigens and neutralizing antibody to subtype B and Thai E isolates. In these assays, THC18 plasma showed a pattern of antibody reactivity similar to other E sera. The genetic subtype of this HIV-1 isolate was identified by HMA. [Louisirirotchanakul et al.(1997)]. Accession number U96782.
- 18) **TH.CM240**: This sequence is from a 21 year old asymptomatic man from northern Thailand. The route of infection is believed to be heterosexual transmission. The blood sample was collected in 1990. The patient's PBMCs were cocultured with stimulated donor PBMCs and proviral DNA was harvested for PCR amplification and sequencing of a cloned full-length genome PCR product. [Carr et al.(1996)]. The complete genome is found in the databases with accession number U54771.
- 19) **TH.N764**: This sequence is from a survey of IV drug using prisoners in Thailand. 12 of 13 sequences from Thai prisoners were of subtype B; N764, from patient (THP13) represents the only subtype E sequence identified in this set, from a prisoner infected in 1989. The sequences were obtained from PCR amplified PBMC DNA. [Kalish et al.(1994)]. Accession number U15588.
- 20) **TH.T8178**: This sequence comes from a study of the genetic heterogeneity and epidemiological distribution of HIV1 in Thailand. The host was a female prostitute and the sequence was obtained from PCR amplified PBMC DNA. [Ou et al.(1993)]. Accession number L19239. See also B_TH.T8174.
- 21) **TH1.ID#**: These twelve sequences are from a set of 23 individuals from Thailand. PCR-direct, peripheral blood PBMC DNA. Referred to as Thai subtype A in [Ou et al.(1992b)] and [Ou et al.(1993)]. (Published erratum appears in *Lancet* 342:250 (1993).) Accession numbers L07443-L07445, L07447-L07448, L07457-L07459 and L07461-L07464.
- 22) **TH2.ID#**: Six of these eight sequences are from 16 isolates from HIV seropositive individuals from Thailand. Sequences were from PCR products derived from co-cultured PBMC DNA. The full length envelope gene sequences are available. [McCutchan et al.(1992)]. Please note: the "TN-ID#" locus names in the database correspond to the McCutchan et al.'s "CM-ID#" isolates. Accession numbers L03698-L03701 and L03703-L03704. The other two (TH238, TN240) are also from Thailand, DNA from PBMC. [Mascola et al.(1994)]. Accession numbers L14571, L14572.
- 23) **TH3.W2TH-ID#**: Fifteen sequences from asymptomatic individuals from Thailand in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. [De Wolf et al.(1994)]:

- [Osmanov et al.(1994)]; and [Gao et al.(1994a)]. Accession numbers U08810-U08811, U08825-U08836 and U08742-U08761. The entry with accession number U09131 is also TH_W2TH022.
- 24) **TH4.D-ID#:** These three sequences are part of a set of sequences generated for the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. They are clones from expanded culture stocks, and are excised from full gp160 sequences. The sequence ID numbers are abbreviated, for example D3TH966 can be read as DAIDS sequence (D), isolated in 1993 (3), Thai (TH), patient 301966 (966). [Gao et al.(1996a)]. Accession numbers U08456-U08458.
 - 25) **TH5.ID#:** These seventeen consensus and five individual sequences are from twenty two patients with AIDS involved in a study of genotypic and phenotypic characteristics of Thai HIV-1. Blood samples were collected between July and December 1993. All sequences were derived from PCR amplified PBMC DNA, after patient PBMCs were cocultured with virus-free donor PBMCs. CMU01, CMU03, CMU04, CMU05, CMU07, and CMU10 are NSI, the rest are SI, as determined by syncytium formation in the cocultured cells. CM = Chang Mai University Hospital. KH = Kavila Army Hospital. All subjects were males and reported past contact with commercial female sex workers, but no history of drug injection, blood transfusion or homosexual contact. [Yu et al.(1995)]. Accession numbers U25550-U25626. Longer sequences from samples KH003, KH005, KH008, CMU02, CMU08 and CMU010 were determined in 1996 [McCutchan et al.(1996)]. Accession numbers U48264-U48269.
 - 26) **TH6.ID#:** These three sequences are E subtype sequences from Thailand. Two individuals believed to be dually infected with subtypes B and E were analyzed. It is not clear from the paper or the database entries which sequences came from individual 1 and which from 2. [Artenstein et al.(1995)]. Accession numbers U21472, U21474, U21476. See also B_TH5.ID#.
 - 27) **TH7.ID#:** These two sequences are from samples collected in 1993 in Thailand. Patients 1018 and 1110 were asymptomatic. [McCutchan et al.(1996)]. Accession numbers U48273-U48274.
 - 28) **TH8.ID#:** These two sequences were from dried blood spots collected in 1992 from a heterosexual (0289) and an IV drug user (0103). DNA was extracted from the blood spots and PCR amplified. The PCR products were directly sequenced. [Cassol et al.(1996)]. Accession numbers U53312 and U53313.
 - 29) **TH9.ID#:** These 14 sequences are from 84 IV drug users in Bangkok, Thailand, who were undergoing methadone treatment at 14 treatment clinics. Blood samples were collected between January and April, 1994. Uncultured PBMC DNA from each patient was PCR amplified, and the PCR product was directly sequenced. Of the 84 patients sampled, 69 were Thai B, one (091) was typical subtype B, and 14 were subtype E. [Kalish et al.(1995)]. Accession numbers U22542, U22548, U22553, U22557, U22561, U22567, U22575, U22604, U22609, U22611, U22612, U22617, U22624 and U22625. See also B_TH7.
 - 30) **TH10.ID#:** These sequences are from Thailand. They are part of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two clones from each isolate were sequenced. [Penny et al.(1996)]. Accession numbers U39256 and U39260, 92TH002; U39255 and U39261, 92TH011.
 - 31) **TWLID#:** These 3 sequences are from healthy HIV-1 carriers or AIDS patients from Taiwan [Chang et al.(1997)]. Other subtypes found in Taiwan in this study were B, C, F and G. Accession numbers U73060, U73062 and U73070.
 - 32) **US.POC30506:** This sequence is from a U.S. serviceman who acquired an HIV-1 infection while deployed in Thailand. He was asymptomatic when the sample for this sequence was collected in 1993. [McCutchan et al.(1996)]. Accession number U48272.
 - 33) **UY.ID#:** These four sequences are from Uruguayan servicemen who acquired HIV-1 infections while deployed as United Nations peacekeepers in Cambodia in 1993. All four were asymptomatic when samples were collected for these sequences in 1993. [McCutchan et al.(1996)]. Accession numbers U48275-U48278.
 - 34) **VN1.ID#:** These 3 sequences are from South Vietnam. [Menu et al.(1996)]. The sequences were from IV drug users in Ho Chi Minh city and Dong Nai, and a female prostitute in Can Tho. A fourth sequence, from Ho Chi Minh city, was found to be subtype B. Accession numbers U29206-U29208.
 - 35) **VN2.ID#:** These 4 sequences are from South Viet Nam. VN1 and VN2 were from healthy 17 and 25 year old female prostitutes from Can Tho and An Giang. VN3 and VN 4 were from male IV drug users. VN3 was 43, had pruritus and splenomegaly, and was from Nha Trang. VN4 was 31, healthy and was also from Nha Trang. [Nerurkar et al.(1996)]. Accession numbers U45239, U45240, U48719 and U48720.

Sequence Descriptions

NOTE:

- 1) While the sequences in this subtype were distinct over this region of env from the other four env subtypes, in the gag gene it is not possible to make a distinction between this subtype and subtype A. What this means is that the isolates for which both gag and env are sequenced which cluster together as the "A" subtype in gag, are very distinctive in env and are broken down into two subtypes, env "A" and env "E". This holds true for the E subtypes sequences that originated in Thailand, as well as the E subtype isolate from the Central African Republic for which gag sequence was obtained. [McCutchan et al.(1992)]; [Louwagie et al.(1993)]; and [Murphy et al.(1993)]. Complete genomes of subtype A and subtype E viruses became available in late 1996 [Gao et al.(1996b)], genbank accession numbers U51188-U51190.
- 2) The relative lack of diversity in the Thai sequences in this subtype relative to the other subtypes is likely to be a consequence of the short time span of the HIV-1 subtype E epidemic in Thailand. [McCutchan et al.(1992)], and [Ou et al.(1992b)].

F Subtype

At this time there are viral sequences from 80 HIV-1 infected individuals associated with HIV-1 subtype F. The F subtype consensus sequence (F_CONSENSUS_97) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published.

- 1) **AR1.ID#:** These two sequences are from direct sequencing of PCR products from uncultured PBMCs, from 1993 samples from Buenos Aires, Argentina. Patient 21280 had AIDS and reported IV drug abuse. Patient 20016 was asymptomatic and HIV risk behavior was unknown. Two other samples taken from unrelated patients in 1993 were subtypes B or B/F recombinant. [Marquina et al.(1996)]. Accession numbers U68522 and U68524.
- 2) **AR2.ID#:** These 3 sequences are from Rosario, Argentina. A total of 24 patients from different risk groups visiting a clinic in Rosario were included in this study. Of the 14 sequences determined, 11 were found to belong to subtype B and 3 were found to belong to subtype F. DNA was extracted from whole blood and PCR amplified. PCR products were directly sequenced. Subtypes of all 24 patients were also tested by HMA. [Campodonico et al.(1996)]. Accession numbers U37032, U37033 and U37043.
- 3) **BR.7944:** This sequence represents a single env F subtype sequence found among 22 Brazilian outpatients with varying degrees of disease progression. It was identified by Potts et al. as the single sequence which did not cluster with North American sequences in phylogenetic analysis. Consensus, PCR clones, peripheral blood PBMC DNA. [Potts et al.(1993b)]. Accession number L19237.
- 4) **BR.RJI03:** An F subtype sequence from Rio de Janeiro, Brazil. 26 additional B and a B-F recombinant were also observed in this set. Year of isolation for RJI03 was 1993, from a woman of CDC clinical stage II. [Morgado et al.(1994)]. DNA was amplified directly from PBMCs of an HIV infected woman with CDC stage II disease, and the PCR product was directly sequenced. Accession number U00422. See also [Sabino et al.(1994c)] Accession number U08974.
- 5) **BR1.BZ-ID#:** Three sequences from Brazil of the F subtype. Full length env (gp160) was amplified from proviral DNA of cultured PBMCs, cloned and sequenced. [Louwagie et al.(1994)]. Accession numbers L22082, L22084 and L22085. The gag gene of these same isolates is found in L22083, L22086 and L11751. Although shown in figure 2B of the paper as clustering with subtype F in the gag region, BZ126 (L22083) seems to be a subtype A outlier with a strong similarity to subtype C at the 3' end, and is 99A/C recombinant ZAM184. See also B_BR4.BZ-ID#.
- 6) **BR2.HSP#:** These 5 sequences are from Sao Paulo Brazil. They seem to be an extension of work published in [Morgado et al.(1994)] and [Sabino et al.(1994c)]. They are proviral DNA sequences from cloned PCR products, from uncultured PBMCs [Sabino et al.(1996)]. Accession numbers U31588, U31592-U31595.
- 7) **BR4.93BR02017:** This sequence is part of a set of sequences generated through the WHO Global Programme on AIDS. The virus was derived from a 52 year old asymptomatic male, from Rio de Janeiro, Brazil, whose route of infection is thought to be due to bisexual contact. The blood sample was taken in 1993. The env sequence clusters with HIV-1 subtype F sequences. The complete gp160 coding region of this isolate was sequenced along with those of others collected at major epicenters of the AIDS epidemic [Gao et al.(1996a)]. Accession number U27401.
- 8) **CM.CA-ID#:** These sequences are 3 of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals; specifically patients CA16 and CA20 were asymptomatic and patient CA4 was symptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. The F subtype designation of these sequences is tentative. Although the F subtype sequences from Cameroon and Brazil consistently form a clade in phylogenetic analyses, the branch lengths between isolates from the two countries are typical of inter-subtype distances, and sequences from the two countries each form their own distinct clade within the F subtype (HIV database and Wouter Janssens, personal communication). [Nkengasong et al.(1994)]. Accession numbers for the subtype F sequences are X80443, X80448 and X80451.
- 9) **CM1.ID#:** These four sequences are all from Cameroon. Takehisa et al unpublished-1997. Accession numbers U70001, U69999, U69998, U69997

Sequence Descriptions

- 10) **CY.HO44-1:** This is a single sequence from two individuals who were heterosexual partners of one another. Patient 16 was a 29 year old bisexual male who was born and lived in Zaire, before moving to Cyprus. He was symptomatic with a CD4 count of 60 and had been seropositive for at least 6 years. Patient 44 was a 32 year old heterosexual female. She was asymptomatic with a CD4 count of 1,136. These samples, like others in this study (see also subtypes A, B, C, and D) were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. DNA was extracted from patient PBMCs and PCR amplified. After a second round of PCR, products were cloned and sequenced. One clone from patient 16 and one from patient 44 were sequenced. [Kostrikis et al.(1995)]. Because of the close epidemiological linkage, only the clone from patient 44 is presented here. Accession numbers U28662 (16) and U28679 (44).
- 11) **FR1.ID#:** This sequence is from a member of the French military who is believed to have been infected while deployed in the Central African Republic in 1992. Other sequences from this study were subtypes A, B, C, E and unclassified. [Lasky et al.(1997)]. Accession number U58807.
- 12) **GA1.VI354:** This sequence is from a 1989 sample from a patient with AIDS living in Libreville, Gabon. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Delaporte et al.(1996)]. Accession number X90923. See also subtypes A, C, D, G and O sequences from this same study.
- 13) **NL1.ID#:** These 3 sequences are from recent immigrants to The Netherlands from Brazil and Zaire. The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. [Lukashov et al.(1996)]. Accession numbers L76899, ZR9306911; L76871, ZR890819; L76901, BR9400960.
- 14) **RO1.ID#:** These nine sequences are from isolates from Romanian children, in different clinical stages. All isolates showed cytopathic properties in peripheral blood mononuclear cells. DNA sequences are direct sequences of PCR products amplified from co-cultured PBMCs. The patients are also known as RM(A-J). [Dumitrescu et al.(1994)]. Accession numbers L19571-L19579.
- 15) **RO2.ID#:** These fifteen sequences are from isolates from Romanian children, infected nosocomially, and adults, infected through heterosexual transmission or transfusion. DNA sequences are direct sequences of PCR products amplified from co-cultured PBMCs. [Apetrei et al.(1997)]. Accession numbers Z83284-Z83303.
- 16) **RO3.RM-ID#:** These 24 sequences are from isolates from orphaned Romanian children, ranging in age from 2.5 to 6 years, admitted to a clinic in Tirgu Mures, Romania. All children were referred to this clinic with serious infections and are believed to have been infected horizontally in different orphanages. Virus was isolated after coculture with donor PBMCs. Proviral DNA from cocultured PBMCs was PCR amplified and the PCR products were directly sequenced. [Holm-Hansen et al.(1995)]. Accession numbers X77964-X77987.
- 17) **RU1.ID#:** These 3 sequences are from St. Petersburg and Armavir, Russia. Sequences were determined by direct sequencing of PCR product from uncultured PBMC proviral DNA. [Leitner et al.(1996b)]. RU26 is believed to have been infected by RU20, so only the RU26 sequence is shown here. RU20 was infected by a woman, whose husband was infected by HIV-1 while living in Congo. RU22, RU23 and RU29 were all heterosexually infected by the same HIV-1 infected man, thus only RU22 is presented here. RU30 is believed to have been infected in 1993 while traveling in the Congo. Accession numbers U69655, U69657, U69659, U69661, U69663 and U69664.
- 18) **TWL3341:** This sequence is one of a set of sequences from Taiwan. Other sequences in the set were subtypes B, C, E or G. [Chang et al.(1997)]. Accession number U67765.

G Subtype

At this time there are viral sequences from 28 HIV-1 infected individuals associated with HIV-1 subtype G. The G subtype consensus sequence (G_CONSENSUS_97) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published and/or have been made available for printing in the database by their authors.

- 1) **BJLID#:** These 2 sequences are from female prostitutes, born in either Ghana or Togo, who live in Benin. 43 is from directly sequenced PCR product, derived via RT-PCR from patient serum RNA. 259 is from cloned PCR product, also by RT-PCR from serum RNA. [Heyndrickx et al.(1996)]. Accession numbers U61872 and U61874. Subtype A sequences were also determined in this study.
- 2) **CF4067:** This sequence was associated with the C subtype in first analysis of the C2V3 region ([Murphy et al.(1993)]), but when a full gp120 sequence became available from this isolate, and phylogenetic analysis was performed including some of the new subtype G sequences, it was more closely associated with G. The full length sequence was kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg Cedex, France. It is part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. Accession numbers L11499 and L11500, U43169.
- 3) **CF115166:** A single sequence from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990-1991. It is a consensus from PCR clones from cultured proviral DNA. [Murphy et al.(1993)]. Accession number L11525. Other sequences from this study were subtypes A, D, E and unclassified. This sequence was originally listed as subtype C in the publication, as was the 4067 isolate.
- 4) **CLLID#:** These 4 sequences are from Abidjan, Ivory Coast. No other information is yet available. Ellenberger D.L. unpublished 1997. Subtypes A, D and G were found for Ivory Coast patients in this set. Ugandan sequences of subtype A were also part of this set. Accession numbers AF000458, AF000460, AF000463, AF000468.
- 5) **GA.LBV217:** A sequence from Gabon from a set of HIV-1 viral isolates from Africa. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a full length env sequence. [Janssens et al.(1994b)]. Accession number U09664.
- 6) **GA.VI525:** A sequence from Gabon from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989-1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1994)] and [Janssens et al.(1994b)]. Accession numbers L22953 and U09665. The same isolate was classified as subtype H in gag. [Louwagie et al.(1993)], Accession number L11792.
- 7) **GA1.ID#:** These 2 sequences are from Gabon. G98 is from a 1988 or 1989 sample from a patient with AIDS living in Franceville, Gabon who moved there from Niger. VI526 is from a 1990 sample from an AIDS patient at the Libreville General Hospital. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Delaporte et al.(1996)]. Accession numbers X90916, G98; X90922, VI526. See also subtypes A, C, D, F and O sequences from this same study.
- 8) **GB1.22:** This sequence is a consensus of three clones from an infected infant in a mother-infant transmission study. The sequences were obtained via PCR from cell lysates, with sequencing of cloned PCR products. The infant was 3 months old at the time of blood drawing, and had pneumonia. [Arnold et al.(1995b)], Accession numbers U26304-U26306. Envelope sequences for the mother are found in database entries U26301-U26302, and gag sequences for mother and infant are in U26303 and U26307.
- 9) **GH3.ID#:** This sequence is from Ghana. Subtypes A and D were also detected in this study [Ishikawa et al.(1996)]. Another report from Ghana [Takahisa et al.(1997a)] showed that a Ghana env subtype G isolate was A/G recombinant, with pol gene related to subtype A. Although this sequence appears to be phylogenetically related to that one, we have no pol sequence for this isolate, and thus no solid evidence of it being A/G recombinant. Accession numbers for subtype G is U67683.
- 10) **KPKr121** This sequence is from an unpublished database entry with accession number X93469.

Sequence Descriptions

- 11) **NG1.ID#:** These four sequences are G subtype sequences from Nigeria [Abimiku et al.(1994)]. JP882 and JV832 were derived from AIDS patients, and G3 and G9 from healthy women. G9 was cultured on the T cell line CEM-SS, and the other three isolates were cocultured with uninfected donor PBMCs. DNA from viral cultures was PCR amplified, cloned and sequenced. Accession numbers U13208-U13209, U13211 and U13213. Complete genomes for JV1083 (aka JV832, 92NG083) and G3 (aka 93NG003) are available in database entries U88826 and U88825, respectively.
- 12) **NL.127C** This consensus sequence represents sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. A sample was collected from the infant at 1.5 months of age. Samples were also collected from the mother before birth, at birth and after birth. Mother sequences are not included in this consensus. [Mulder-Kampinga et al.(1993)]. [Mulder-Kampinga et al.(1995)]. Infant 127 sequences are from Accession numbers Z47817-Z47832. Mother 127 sequences are from Accession numbers Z47833-Z47880. Gag gene sequences from mother/child pairs are also available in entries with accession numbers Z47903-Z47911; Z47912-Z47928; Z47929-Z47935; Z47936-Z47950. The second mother/child pair was also from the Netherlands, see B_NL.114C. The third mother/infant pair in this study was from Rwanda, see A_RW.564C.
- 13) **NL1.ID#:** These 4 sequences are from recent immigrants to The Netherlands from Brazil and Zaire. The first two letters of the ID# represent the two letter country code for the previous residence of the patient. The next two numerals represent the year of isolation. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR products were directly sequenced. [Lukashov et al.(1996)]. Accession numbers L76880, ZR911976; L76906, LR9401885; L76884, UM9210113; L76902, GH9401230.
- 14) **RU1.ID#:** RU16 is from Volgograd, Russia. The sequence was determined by direct sequencing of PCR product from uncultured PBMC proviral DNA. [Leitner et al.(1996b)]. RU16 is a 21 year old female who was infected nosocomially. The nosocomial outbreak in southern Russia began with an index case, an infant born in Elista, to a HIV-1 infected mother whose husband is believed to have been infected in the Congo. Over a period of several months in 1988-1989, over 250 patients, mostly children, were infected perenterally with contaminated needles in several hospitals [Bobkov et al.(1994a)], [Lukashov et al.(1995)], [Bobkov et al.(1997b)]. This sequence clusters with the other Russian subtype G sequences, which were also part of this nosocomial outbreak. Accession number U69653. Other sequences from this Russian outbreak can be found with accession numbers L38413, U08355-U08368, U30312, U30313, U27445, U51295. Several of these are from the WHO isolate 92RU131 (U27445), from a 3.5 year old female from Rostov-on-Don, isolated in 1992, which has been classified as G/A recombinant [Gao et al.(1996a)]. The CF4067 sequence is more closely related to these Russian G sequences, than are some other G sequences from Russia, such as BUK3a from Uzbekistan. Other sequences, with only 105 bp of V3 loop data are available in entries with accession numbers U10701-U10859, these include a mother who is thought to have been infected by her infant during breast feeding [Bobkov et al.(1994b)].
- 15) **TW1.267:** This sequence is from directly sequenced PCR product from uncultured PBMCs from Taiwan. [Chang et al.(1997)]. Most of the Taiwanese sequences determined in this study were subtype B, but subtypes C, E and F were also found. Database accession number U73058.
- 16) **UG.JW3:** This single sequence of subtype G HIV-1 is from a female patient with stage IV disease and CD4 count of 20, from Uganda. The patient had recently migrated to the United Kingdom from Uganda, but contracted the HIV in Uganda. The sequence is referred to as JW3 in [Kaleebu et al.(1995)] but was previously referred to as K1, by the HIV database. It has been given the name 92UG975.10 by the World Health Organization. Accession numbers U22010, U27426.
- 17) **UZ1.BUK3a:** This sequence is from a 39 year old male caucasian heterosexual who lived in Uzbekistan [Bobkov et al.(1996a)]. The patient tested HIV-1 seropositive in June 1991, blood was collected for this sequence in November, 1992. He reported living in Mozambique in 1984-1985 where he was admitted to the hospital several times, but reported having no sexual relations during this time. His wife in Uzbekistan was HIV-1 seropositive in 1991 and died of AIDS in 1995. He reported having sexual relations with a woman in Uzbekistan in 1986-1987 shortly after returning from Mozambique, and this woman and her husband were both found to be HIV-1 seropositive in 1991. Patient BUK died of AIDS in November, 1993. Proviral DNA from uncultured PBMCs was PCR amplified in nested double reactions. The PCR product was cloned into pUC18 plasmid prior to sequencing. Accession numbers U33095, U33096.

H Subtype

At this time there are viral sequences from 3 HIV-1 infected individuals associated with HIV-1 subtype H. The H subtype consensus sequence (H_CONSENSUS_97) generated from these 2 sequences was based on the most common amino acid found in each position of an alignment. Both of these sequences have been published and/or have been made available for printing in the database by their authors. Eight sequences that are too short for classification are closer to H than to other subtypes. The locus names (ID's) and sources of the sequences are:

- 1) **CF90CF056:** This sequence is from a 1990 blood sample from the Central African Republic. It was originally published as an unclassified subtype, [Murphy et al.(1993)], with accession number L11497. It was later classified as subtype H, [Janssens et al.(1994b)]. The complete genome was sequenced in 1997, Fong Gao unpublished 1997, and appears to be subtype H throughout the genome, accession number AF005496. L11497 contains an 8 base-pair, frameshifting insertion, relative to AF005496, near the 3' end of its sequence, which is a third copy of an 9 bp direct repeat.
- 2) **CM.CA13:** A sequence from Cameroon from a set of HIV-1 viral isolates from Africa used to define the prototype G and H env sequences. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a 900 base pair sequence. [Janssens et al.(1994b)] and [Nkengasong et al.(1994)]. The H subtype association is not always clearly apparent using some sets of background sequences for comparison, and neighbor joining trees (HIV database, Wouter Janssens, personal communication), although parsimony trees confirmed the original association documented in [Janssens et al.(1994b)]. Accession numbers X80441 and U09667.
- 3) **RU1.MLY10:** This sequence is from Russia [Bobkov et al.(1996c)]. Patient MLY was a recipient of a 1987 blood transfusion from patient SLH. Patient SLH had a history of sexual contact with an HIV-1 seropositive student from Zaire. Patient MLY gave birth to a child by caesarian section (soon after which she received the transfusion), and the child was later determined to be HIV infected, presumably via transmission during a 3-month period of breast feeding. Accession numbers: U33104 U33105, MLY; U33106 U33107 U33108, SHL.
- 4) **ZR.VI557:** A sequence from Zaire from a set of HIV-1 viral isolates from Africa used to define the prototype G and H env sequences. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a 900 base pair sequence. [Janssens et al.(1994b)]. Accession number U09666.

I Subtype

At this time there are viral sequences from 2 HIV-1 infected individuals associated with HIV-1 subtype I. The subtype I consensus sequence reflects the sequencing of two clones from one individual and one clone from the other. Both of these sequences have been published.

- 1) **CY.HOcon:** This is a consensus sequence from two individuals who were heterosexual partners of one another, and former IV drug users. They had lived for several years in Athens, Greece as well as in Cyprus. These samples, like others in this study (see also subtypes A, B, C, and F) were collected in February 1994 from the AIDS clinic in Nicosia, Cyprus. Patient HO31 was a 24 year old asymptomatic female known to have been HIV seropositive for at least 5 years. Patient HO32 was a 35 year old asymptomatic male, also seropositive for at least 5 years. DNA was extracted from patient PBMCs and PCR amplified. After a second round of PCR, products were cloned and sequenced. Two clones from HO32 and one from HO31 were sequenced. [Kostrikis et al.(1995)]. Accession numbers U28672, U28673 and U28685.

J Subtype

At this time there are viral sequences from 5 HIV-1 infected individuals associated with HIV-1 subtype J. The J subtype consensus sequence (J_CONSENSUS_97) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published or made available to the database for printing.

- 1) **GM1.GM#:** Three sequences from Gambia, as yet unpublished, which cluster with the two Swedish sequences by Leitner in the V3 loop region. Accession numbers U33099, U33100, U33102. Bobkov et al. 1996 unpublished. GM4 (U33099 U43105) has been published in [Bobkov et al.(1996b)] as a G*/C recombinant in the env V1-V5 region, with the ? region covering the V3 loop. See also Gambian sequences of subtypes B and C.
- 2) **SE1.SE#:** Two sequences from Sweden, both from patients who were recent immigrants from Zaire. [Leitner et al.(1995)]. Sample 7022 as collected in December 1993 from an asymptomatic female with a CD4 count of 184. Her first known seropositive sample is from May 1990, but epidemiological investigation indicated that she was infected in Zaire between 1981 and 1986. Sample 7887 was collected in October 1994 from an asymptomatic male who had tested seronegative in Sweden in January 1993, and who had a seropositive sample in August 1994. His CD4 count was normal at 567. Both patients were heterosexual and a thorough epidemiological investigation revealed no contact or shared contacts between the two. The two sequences were 95% identical to each other over 255 bases of env and 98% identical to each other over 460 bases of gag. Accession numbers L41176 and L41177. Gag gene sequences from these same individuals are in L41178 and L41179.

O Subtype

At this time there are viral sequences 13 HIV-1 infected individuals associated with HIV-1 subtype O that have been published and/or have been made available for printing in the database by their authors. The O subtype consensus sequence (O_CONSENSUS_97) generated from these sequences was based on the most common amino acid found in each position of the alignment; when there was no consensus in a position an "X" was used. These sequences represent a set of sequences that are extremely divergent relative to other HIV-1's. The subtypes A-H have been grouped together under the heading "M" for main. "O" sequences are as different from one another as are sequences from different "M" subtypes.

- 1) **CM.CA9:** This sequence is from an individual living in Cameroon. [Janssens et al.(1994d)]. No database entry is yet available. The pol gene from this isolate has Accession number X78476.
- 2) **CM.ANT70:** The complete viral genome has been sequenced from this viral isolate derived from a symptomatic Cameroonian, CDC stage III. [R. et al.(1990)] and [Vanden Haesevelde et al.(1994)]. Accession number M31171. LTR and partial env sequences were also presented in L20587 and L23119.
- 3) **CM.MVP5180:** The complete viral genome has been sequenced from an isolate derived from a Cameroonian woman, sampled in 1991; the donor died of AIDS in 1992. The viral isolate MVP-5180 was grown in several human T-cell lines and the monocytic U937 line. [Gurtler et al.(1994)]. Accession number L20571.
- 4) **CM.CMR61:** This sequence is one of two O group sequences from a 23 year old female commercial sex worker from Cameroon who was found to be triple-infected with subtypes A and D, as well as O group HIV-1 [Takehisa et al.(1997b)]. Accession numbers U58152, U58153.
- 5) **ES.1158:** This sequence was from a 35 year old man from Spain. Two blood samples from this same man were collected in April and September 1995. The V3 region was PCR amplified from uncultured PBMCs, cloned into pGEM-5ZF and an individual clone sequenced. The April sequence is shown here. The sequence from the September blood sample (681, Accession number U62617) is also available. Accession number U62618.
- 6) **FR.CF#:** These seven consensus sequences are from Cameroonian patients living in France. [Loussert-Ajaka et al.(1995)]. PBMC proviral DNA was PCR amplified and 3-6 clones from each patient were sequenced. The consensus of the 3-6 clones is presented. Accession numbers U24562-U24568. Gag gene sequences for these patients are also available with Accession numbers U24706-U24712.
- 7) **FR.DUR:** This sequence is as yet unpublished, by J.H.M. Cohen et al. Accession number X84327.
- 8) **FR.VAU:** This sequence was derived from an isolate from a French woman who died of AIDS in 1992. DNA was extracted from VAU infected PBMCs, PCR amplified, cloned, and gp160 env was sequenced. The viral isolate was highly cytopathic. [Charneau et al.(1994)]. Accession number X80020.
- 9) **GA1.VI686:** This sequence is from a 1992 sample from a Gabonese woman with AIDS, taken at the Libreville General Hospital in Gabon. Method of proviral DNA isolation was not described. DNA was PCR amplified and cloned. One clone per isolate was sequenced. [Janssens et al.(1994d)], [Delaporte et al.(1996)]. Accession number X96526. The pol gene from this isolate has Accession number X78477. See also subtypes A, C, D, F and G sequences from this same study.

Uncertain Classification

At this time there are viral sequences from 19 HIV-1 infected individuals that are not clearly associated with any of the HIV-1 genetic subtypes A through J. They either appeared distinct from the subtypes A-J in phylogenetic analysis, or else the subtype association was unclear, with different associations in different analyses. For some of the shorter gene fragments, subtype associations might have been established if more sequence information was available or if a different set of sequences was included in the background set used to define subtype associations. Some of these sequences may be representatives of subtypes as divergent as A-J, but only a single limited sample is yet available. Still others may represent recombinant genomes.

- 1) **AR.20021:** This B/F recombinant sequence is from direct sequencing of PCR product from uncultured PBMCs, from a 1993 sample from Buenos Aires, Argentina. The patient was asymptomatic and HIV risk behavior was unknown. Three other samples taken from unrelated patients in 1993 were subtypes F (2) or B (1). [Marquina et al.(1996)]. Accession number U68523.
- 2) **BIL91BU009:** This C/D sequence is from Burundi. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. 91BU009 groups with subtype D in a neighbor-joining tree of the V3 region and with subtype C in other regions. [Penny et al.(1996)]. Accession numbers U39253 and U39254.
- 3) **BR.93BR01904** This sequence is part of the WHO Global Programme on AIDS. The virus was derived from a twenty-year old asymptomatic male patient in Rio de Janeiro, Brazil, who presumably contracted the virus through bisexual contact. Blood sample was taken in 1993 [Gao et al.(1996a)]. This sequence appears to be a recombinant of subtypes B and F. Accession numbers U27404, U27408 and U27444. All three confirm the subtype F/B recombination in env.
- 4) **BR.93BR029:** This sequence is another B/F recombinant from Brazil. It is one of several complete env gene sequences obtained for the World Health Organization. HMA subtyping as well as sequence-determined subtyping was done on each one. Sequences were PCR amplified from cocultured PBMCs. Two to three clones from each isolate were sequenced. [Penny et al.(1996)]. Another env sequence from this same patient was determined by a second group. [Gao et al.(1996a)]. Accession numbers U27413, U39235, U39236 and AF005495 (complete genome). LTR sequence from this isolate is in the entry with accession number U51291.
- 5) **BR.RJ03** This sequence is B-F recombinant in the V3 region. DNA was amplified directly from PBMCs of an HIV infected woman with CDC stage II/A disease in August, 1992, and the PCR product was directly sequenced [Morgado et al.(1994)]. More V3 region sequences from this individual (RJ549 from April 1992) and her sexual partner (RJ548 from April 1992) were also sequenced [Sabino et al.(1994c)]. Accession numbers U00420, U08953-U08955, U08957-U08960, U08962-U08964, U10019-U10029, U08972, U08973 and U08965-U08971.
- 6) **CF1.ID#:** These four sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990-1991. Consensus, PCR-clones, cell culture, DNA. [Murphy et al.(1993)]. Accession numbers L11482-L11483, L11497, L11508-L11510 and L11514-L11515. Janssens et al. classified 4056 as an H subtype sequence [Janssens et al.(1994b)]. D. Schmitt provided an unpublished sequence of CF4081, U43174.
- 7) **FR1.ID#:** This sequence is from a member of the French military who is believed to have been infected while deployed in Djibouti in 1991. It was classified as subtype C in [Lasky et al.(1997)]. However, it has a GPCR V3 loop tip, and clusters with subtype B in phylogenetic analysis done for this section of the HIV Database Compendium. All other subtype C as of November 1997 have GPGQ at the tip of the V3 loop. Other sequences from this study were subtypes A, B, C, E, and F. [Lasky et al.(1997)]. Accession number U58787.
- 8) **GH.GH8:** This A/G recombinant sequence is from Ghana. Subtypes A and D were also detected in this study [Takehisa et al.(1997a)]. Accession number U67039.
- 9) **KE.K124:** A Kenyan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably

Sequence Descriptions

- between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. This isolate is not clearly associated with D subtype; however, Louwagie and colleagues found that it associated with the D subtype in env, and the A subtype in gag. Using parsimony analysis, we found that it was difficult to determine a clear association, and this observation was confirmed by Wouter Janssens (personal communication). Accession number L22942.
- 10) **KE.KEN976** This is a single unclassified sequence from a set of patients who were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced. [Janssens et al.(1994a)]. Accession number U12992.
 - 11) **NG.IBNG**: An envelope gp120 sequence from Nigeria was kindly provided by Dr. Tom Howard from the University of Southern California (USC) [Howard et al.(1994)] and the complete genome has now been sequenced [Howard & Rasheed(1996)]. Originally classified as subtype A, it has more recently been thought to be a mosaic between subtypes A and G. Accession numbers U48628 and L39106.
 - 12) **NL.RW94028**: This sequence is from a recent immigrant to The Netherlands from Rwanda. The blood sample was collected in 1994. Viral RNA was prepared from patient serum and RT-PCR was used to amplify the V3 region of the env gene. The PCR product was directly sequenced. [Lukashov et al.(1996)]. Database accession number L76909.
 - 13) **RW.92RW009**: This sequence is from a complete genome of a 1992 isolate from Rwanda. The env region clusters with subtype A, but overall the genome is A/C recombinant or mosaic, [Gao et al.(1994a)], [Gao et al.(1994b)], [De Wolf et al.(1994)]. Accession numbers U08793, U16220, U16221, U16222, U88823, U13441. A Gag gene is available with accession number U86545.
 - 14) **SE.KI4803**: This sequence is from patient number 24 described in [Asjo et al.(1986)] and [Fredriksson et al.(1991)]. Several molecular clones from this patient have been extensively characterized in [Tan et al.(1993)]. Complete env gp120 sequences for 8 clones were determined in [McKeating et al.(1996)] and one of the 8 sequences (clone 13) is presented here. Accession numbers for 7 of the 8 clones (clone 32 was not submitted to the databases) are U57788–U57794. A complete genome from isolate KI4803 has reportedly been sequenced, but not yet submitted to the databases (Fenyó, EM personal communication). The patient had AIDS at the time of viral isolation in 1985. The virus exhibited rapid/high phenotype [Asjo et al.(1986)] and when gp120 from this isolate was swapped into the HXB2 genome, the ability of the chimeras to grow in various cell lines correlated with the gp120 [McKeating et al.(1996)]. Although the amino acid sequences translated from these sequence entries are clearly subtype B, the DNA sequences do not cluster with any of the HIV-1 M-group subtypes. The codon usage of these sequences is similar to the codons used for optimal high-level expression in E.coli, and it is possible that these sequences were reverse-translated from amino acid sequences. The authors have been contacted, but have not responded.
 - 15) **UG1.ID#**: Two clones from each of these two D/A recombinant Ugandan isolates were sequenced, but only the sequences of 92UG035 clone 21 and C6080 clone 09 are shown here. [Douglas et al.(1996)]. The publication shows C6080 as subtype A, but analyses done at Los Alamos indicate that it is a D/A recombinant, like 92UG035. Other Ugandan isolates sequenced in this study were subtype D. London subtype B clones were also reported. Complete envelope gp160 sequences were reported for all isolates. Accession numbers U36865, U36866, U36881 and U36883.
 - 16) **ZM.ZAM184**: This Zambian sequence is an outlier, though in some phylogenetic analysis it appears most closely associated with the A subtype. In particular it is closely associated with A₁CFSAS U43171 (100/100 replicates in parsimony analysis of gp120). Health status of the individual from which the virus was cultured was a woman from Lusaka, Zambia who participated in a clinical research study. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)], [Salminen et al.(1997)]. The Salminen paper is based on full-length gag and env genes recovered directly from peripheral blood mononuclear cells or from primary virus cultures, using serial blood samples from a Zambian woman and a sample from her spouse. DNA sequencing and phylogenetic analysis established that two different A/C recombinant forms of HIV-1 predominated at two time points in the woman. A related but distinct recombinant HIV-1 was recovered from her spouse. Intersubtype recombination apparently played a central role in the evolution of HIV-1 in this couple and may contribute substantially to the rapid emergence of HIV-1 variants whenever mixed-subtype HIV-1 infections occur. Accession number L22955. A complete genome from this isolate is now

available with accession number U86780 and other clones from this patient and her spouse, from blood samples taken in 1989 and 1990, are found with accession numbers U86768-U86781.

- 17) **ZR.Z3:** This sequence is from the 1983 Zairean isolate Z-3 (non-infectious, possibly due to frame-shift). [Willey et al.(1986)]. Accession number K03347.
- 18) **ZR.Z321:** This sequence is from the 1976 Zairean isolate Z321. [Srinivasan et al.(1989)]. Accession number M15896. Earlier listed as subtype A, it was subsequently shown to be recombinant between subtypes A and G [Choi et al.(1997)]. Accession numbers M15896, U76035, U50207, U50208.

References

- [Abebe et al.(1997)] A. Abebe, C. L. Kuiken, J. Goudsmit, M. Valk, T. Messele, T. Sahlu, H. Yeneneh, A. Fontanet, W. D. F., & D. R. Wit. HIV type 1 subtype C in Addis Ababa, Ethiopia. *LA - Bag. AIDS Res Hum Retroviruses* 13:1071-5, 1997. OTE: (Medline: 97407546).
- [Abele & DeBry(1992)] L. G. Abele & R. W. DeBry. Florida dentist case: research affiliation and ethics. *Science* 255:903, 1992. OTE: (Medline: 92188144).
- [Abiniku et al.(1994)] A. G. Abiniku, T. L. Stern, A. Zwandor, P. D. Markham, C. Calef, S. Kyari, W. C. Saxinger, R. C. Gallo, M. Robert-Guroff, & M. S. Reitz. Subgroup G HIV type 1 isolates from Nigeria. *AIDS Res Hum Retroviruses* 10:1581-1583, 1994. OTE: (Medline: 95194721).
- [Adachi et al.(1986)] A. Adachi, H. E. Gendelman, S. Koenig, T. Folks, R. Willey, A. Rabson, & M. A. Martin. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J Virol* 59:284-91, 1986. OTE: (Medline: 86281827), (Genbank: M19921).
- [Ahmad et al.(1995)] N. Ahmad, B. M. Baroudy, R. C. Baker, & C. Chappey. Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. *J Virol* 69:1001-1012, 1995. OTE: (Medline: 95115054), (Genbank: U16390 U16391 U16392 U16393 U16394 U16395 U16396 U16397 U16398 U16399 U16400 U16401 U16402 U16403 U16404 U16405 U16406 U16407 U16408 U16409 U16410 U16411 U16412 U16413 U16414 U16415 U16416 U16417 U16418).
- [Ait-Khaled & Emery(1993)] M. Ait-Khaled & V. C. Emery. Sequence variation within the human immunodeficiency virus V3 loop at seroconversion. *J Med Virol* 41:270-274, 1993. OTE: (Medline: 94149438).
- [Albert et al.(1992)] J. Albert, L. Franzen, M. Jansson, G. Scarlatti, P. K. Kataaha, E. Katabira, F. Mubiro, M. Rydaker, P. Rossi, & U. P. et al. Ugandan HIV-1 V3 loop sequences closely related to the U.S./European consensus. *Virology* 190:674-81, 1992. OTE: (Medline: 92391084), (Genbank: M98894 M98895 M98896 M98897 M98898 M98899 M98900 M98901 M98902 M98903 M98904 M98905 M98906 M98907 M98908 M98909 M98910 M98911 M98912 M98913 M98914 M98915 M98916 M98917 M98918 M98919 M98920 M98921 M98922 M98923).
- [Alizon et al.(1986)] M. Alizon, S. Wain-Hobson, L. Montagnier, & P. Sonigo. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* 46:63-74, 1986. OTE: (Medline: 86245056), (Genbank: K02013 K03454 K03456).
- [Anand et al.(1989)] R. Anand, R. Thayer, A. Srinivasan, S. Nayyar, M. Gardner, P. Luciw, & S. Dandekar. Biological and molecular characterization of human immunodeficiency virus (HIV-1BR) from the brain of a patient with progressive dementia. *Virology* 168:79-89, 1989. OTE: (Medline: 89085613), (Genbank: M21098).
- [Andeweg et al.(1992)] A. C. Andeweg, M. Groenink, P. Leeftang, G. de R. E., A. D. Osterhaus, M. Tersmette, & M. L. Bosch. Genetic and functional analysis of a set of HIV-1 envelope genes

Sequence Descriptions

- obtained from biological clones with varying syncytium-inducing capacities. *AIDS Res Hum Retroviruses* 8:1803-13, 1992. OTE: (Medline: 93090463).
- [Antonioli et al.(1995)] I. M. Antonioli, C. Baumberger, S. Yerly, & L. Perrin. V3 sequences in primary HIV-1 infection. *AIDS* 9:11-17, 1995. OTE: (Medline: 95200567).
- [Apetrei et al.(1997)] C. Apetrei, I. Loussert-Ajaka, G. Collin, F. Letourneur, M. Duca, S. Saragosti, F. Simon, & F. Brun-Vezinet. HIV type 1 subtype F sequences in Romanian children and adults. *LA - Eng. AIDS Res Hum Retroviruses* 13:363-5, 1997. OTE: (Medline: 97225064), (Genbank: Z83284 Z83285 Z83286 Z83287 Z83288 Z83289 Z83290 Z83291 Z83292 Z83293 Z83294 Z83295 Z83296 Z83297 Z83298 Z83299 Z83300 Z83301 Z83302 Z83303).
- [Arnold et al.(1995a)] C. Arnold, P. Balfe, & J. P. Clewley. Sequence distances between env genes of HIV-1 from individuals infected from the same source: implications for the investigation of possible transmission events. *Virology* 211:198-203, 1995a. OTE: (Medline: 95373135), (Genbank: U23112 U23113 U23114 U23115 U23116 U23117 U23118 U23119 U23120 U23121 U23122 U23123 U23124 U23125 U23126 U23127 U23128 U23129 U23130 U23131 U23132 U23133 U23134 U23135 U23136 U23137 U23138).
- [Arnold et al.(1995b)] C. Arnold, K. L. Barlow, S. Kaye, C. Loveday, P. Balfe, & J. P. Clewley. HIV type 1 sequence subtype G transmission from mother to infant: failure of variant sequence species to amplify in the Roche Amplicor Test. *AIDS Res Hum Retroviruses* 11:999-1001, 1995b. OTE: (Medline: 96020103), (Genbank: U26301 U26302 U26303 U26304 U26305 U26306 U26307).
- [Arnold et al.(1995c)] C. Arnold, K. L. Barlow, J. V. Parry, & J. P. Clewley. At least five HIV-1 sequence subtypes (A, B, C, D, A/E) occur in England. *AIDS Res Hum Retroviruses* 11:427-429, 1995c. OTE: (Medline: 95306148), (Genbank: U21095 U21096 U21097 U21098 U21099 U21100 U21106 Z33479).
- [Artenstein et al.(1995)] A. W. Artenstein, T. C. VanCott, J. R. Mascola, J. K. Carr, P. A. Hegerich, J. Gaynes, E. Sanders-Buell, M. L. Robb, D. E. Dayhoff, & S. T. et al. Dual infection with human immunodeficiency virus type 1 of distinct envelope subtypes in humans. *J Infect Dis* 171:805-810, 1995. OTE: (Medline: 95221984), (Genbank: U21471 U21472 U21473 U21474 U21475 U21476).
- [Ashkenazi et al.(1991)] A. Ashkenazi, D. H. Smith, S. A. Marsters, L. Riddle, T. J. Gregory, D. D. Ho, & D. J. Capon. Resistance of primary isolates of human immunodeficiency virus type 1 to soluble CD4 is independent of CD4-rgp120 binding affinity. *Proc Natl Acad Sci U S A* 88:7056-60, 1991. OTE: (Medline: 91334404).
- [Asjo et al.(1986)] B. Asjo, L. Morfeldt-Manson, J. Albert, G. Biberfeld, A. Karlsson, K. Lidman, & E. M. Fenyo. Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet* 2:660-662, 1986. OTE: (Medline: 87013480).
- [Atkin et al.(1993)] A. Atkin, G. Pestano, D. Serwadda, A. M. Prince, D. Pascual, N. Sewankambo, & W. M. Boto. Phylogenetic and serological characterization of two Ugandan HIV-1 isolates. *AIDS Res Hum Retroviruses* 9:351-6, 1993. OTE: (Medline: 93290930), (Genbank: M98503 M98504).
- [Audoly et al.(1996)] G. Audoly, S. Filion, P. Pinay, & C. Desgranges. In vivo and in vitro sequence diversity of the V3 env region of HIV type 1 from the Ivory Coast. *AIDS Res Hum Retroviruses* 12:1073-1075, 1996. OTE: (Medline: 96424763).
- [Ayyavoo et al.(1996)] V. Ayyavoo, K. E. Ugen, L. S. Fernandes, J. J. Goedert, A. Rubinstein, W. V. Williams, & D. B. Weiner. Analysis of genetic heterogeneity, antigenicity, and biological characteristics of HIV-1 in a maternal transmitter and nontransmitter patient pair. *DNA Cell Biol* 15:571-580, 1996. OTE: (Medline: 96326292), (Genbank: U07833 U07834 U07835 U07836 U07837 U07839 U07840 U07841 U07891 U07892 U07893 U07894 U07895 U07896 U07897 U07898 U07899 U07900 U07901 U07902 U07903 U07904 U07905 U07906 U07907 U07908 U07909 U07910 U07911 U07912).
- [Balfe et al.(1990)] P. Balfe, P. Simmonds, C. A. Ludlam, I. O. Bishop, & A. J. Brown. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate

- of sequence change and low frequency of inactivating mutations. *J Virol* 64:6221-33, 1990. OTE: (Medline: 91056593), (Genbank: M36997).
- [Balotta et al.(1997a)] C. Balotta, P. Bagnarelli, C. Riva, A. Valenza, S. Antinori, M. C. Colombo, R. Sam-paolesi, M. Violin, P. de M. P., M. Moroni, M. Clementi, & M. Galli. Comparable biological and molecular determinants in HIV type 1-infected long-term nonprogressors and recently infected individuals. LA - Eng. *AIDS Res Hum Retroviruses* 13:337-41, 1997a. OTE: (Medline: 97225060).
- [Balotta et al.(1997b)] C. Balotta, P. Bagnarelli, M. Violin, A. L. Ridolfo, D. Zhou, A. Berlusconi, S. Corvasce, M. Corbellino, M. Clementi, M. Clerici, M. Moroni, & M. Galli. Homozygous delta 32 deletion of the CCR-5 chemokine receptor gene in an HIV-1-infected patient. LA - Eng. *AIDS* 11:F67-71, 1997b. OTE: (Medline: 97401097).
- [Barin et al.(1996)] F. Barin, Y. Lahbabi, L. B. B. Lejeune, A. Baillou-Beaufils, F. Denis, C. Mathiot, S. M'Boup, V. Vithayasai, U. Dietrich, & A. Goudeau. Diversity of antibody binding to V3 peptides representing consensus sequences of HIV type 1 genotypes A to E: an approach for HIV type 1 serological subtyping. *AIDS Res Hum Retroviruses* 12:1279-1289, 1996.
- [Baskar et al.(1994)] P. V. Baskar, S. C. Ray, R. Rao, T. C. Quinn, J. E. Hildreth, & R. C. Bollinger. Presence in India of HIV type 1 similar to North American strains. *AIDS Res Hum Retroviruses* 10:1039-1041, 1994. OTE: (Medline: 95110628), (Genbank: L29091 L29092 L29093 L29094).
- [Becker et al.(1995)] M. L. Becker, J. D. G., & W. B. Becker. Analysis of partial gag and env gene sequences of HIV type 1 strains from southern Africa. *AIDS Res. Hum. Retroviruses* 11:1265-1267, 1995. OTE: (Medline: 96157217), (Genbank: X52154).
- [Bex et al.(1994)] F. Bex, P. Hermans, S. Sprecher, A. Achour, R. Badjou, C. Desgranges, J. Cogniaux, P. Franchioli, C. Vanhulle, & A. Lachgar. Syngeneic adoptive transfer of anti-human immunodeficiency virus (HIV-1)-primed lymphocytes from a vaccinated HIV-seronegative individual to his HIV-1-infected identical twin. *Blood* 84:3317-26, 1994. OTE: (Medline: 95036224).
- [Bieniasz et al.(1997)] P. D. Bieniasz, R. A. Fridell, I. Aramori, S. S. Ferguson, M. G. Caron, & B. R. Cullen. HIV-1-induced cell fusion is mediated by multiple regions within both the viral envelope and the CCR-5 co-receptor. *EMBO J* 16:2599-2609, 1997.
- [Blouin et al.(1996)] J. C. Blouin, E. A. Guzman, & B. T. Foley. Global Variation in the HIV-1 V3 region. In *Human Retroviruses and AIDS 1996*, pages III 77-201, Los Alamos, NM. Los Alamos National Laboratory: Theoretical Biology and Biophysics, 1996.
- [Bobkov et al.(1994a)] A. Bobkov, R. Cheingsong-Popov, M. Garaev, A. Rzhaniyeva, P. Kaleebu, S. Beddows, M. H. Bachmann, J. I. Mullins, J. Louwagie, & W. J. et al. Identification of an env G subtype and heterogeneity of HIV-1 strains in the Russian Federation and Belarus. *AIDS* 8:1649-1655, 1994a. OTE: (Medline: 95194609), (Genbank: U08356 U08357 U08358 U08359 U08360 U08361 U08362 U08363 U08364 U08365 U08366 U08367 U08368).
- [Bobkov et al.(1996a)] A. Bobkov, R. Cheingsong-Popov, M. Garaev, & J. Weber. Glycoprotein 120 polymorphism in an HIV type 1 epidemic originating from a point source: nucleotide sequence analysis of variants with conserved V3 loop sequences. *AIDS Res Hum Retroviruses* 12:251-253, 1996a. OTE: (Medline: 96432138).
- [Bobkov et al.(1996b)] A. Bobkov, R. Cheingsong-Popov, M. Salminen, F. McCutchan, J. Louwagie, K. Ariyoshi, H. Whittle, & J. Weber. Complex mosaic structure of the partial envelope sequence from a Gambian HIV type 1 isolate. LA - Eng. *AIDS Res Hum Retroviruses* 12:169-171, 1996b. OTE: (Medline: 96431385).
- [Bobkov et al.(1996c)] A. Bobkov, R. Cheingsong-Popov, L. Selimova, E. Kazemova, N. Karasyova, A. Kravchenko, N. Ladnaya, V. Pokrovsky, & J. Weber. Genetic heterogeneity of HIV type 1 in Russia: identification of H variants and relationship with epidemiological data. LA - Eng. *AIDS*

Sequence Descriptions

- Res Hum Retroviruses* 12:1687-90, 1996c. OTE: (Medline: 97118361), (Genbank: U33104 U33105 U33106 U33107 U33108).
- [Bobkov et al.(1997a)] A. Bobkov, R. Cheingsong-Popov, L. Selimova, N. Ladanaya, E. Kazennova, A. Kravchenko, V. Pokrovsky, & J. Weber. HIV type 1 subtype E in Russia. LA - Eng. *AIDS Res Hum Retroviruses* 13:725-7, 1997a. OTE: (Medline: 97311525), (Genbank: U93607 U93608 U93609 U93610).
- [Bobkov et al.(1994b)] A. Bobkov, M. M. Garaev, A. Rzhabinova, P. Kaleebu, R. Pitman, J. N. Weber, & R. Cheingsong-Popov. Molecular epidemiology of HIV-1 in the former Soviet Union: analysis of env V3 sequences and their correlation with epidemiologic data. *AIDS* 8:619-624, 1994b. OTE: (Medline: 94338597).
- [Bobkov et al.(1997b)] A. F. Bobkov, V. V. Pokrovskii, L. M. Selimova, E. V. Kazennova, N. G. Karaseva, N. N. Ladanaya, A. V. Kravchenko, R. Cheingsong-Popov, & D. Veber. (Genotyping and phylogenetic analysis of HIV-1 isolates circulating in Russia) LA - Rus. *Vopr Virusol* 42:13-6, 1997b. OTE: (Medline: 97216819).
- [Briant et al.(1995)] L. Briant, C. M. Wade, J. Puel, A. J. Brown, & M. Guyader. Analysis of envelope sequence variants suggests multiple mechanisms of mother-to-child transmission of human immunodeficiency virus type 1. *J Virol* 69:3778-3788, 1995. OTE: (Medline: 95264471), (Genbank: U24717 U24718 U24719 U24720 U24721 U24722 U24723 U24724 U24725 U24726 U24727 U24728 U24729 U24730 U24731 U24732 U24733 U24734 U24735 U24736 U24737 U24738 U24739 U24740 U24741 U24742 U24743 U24744 U24745 U24746).
- [Brown et al.(1996)] T. M. Brown, K. E. Robbins, M. Sinniah, T. S. Saraswathy, V. Lee, L. S. Hooi, B. Vijayamalar, C. C. Luo, C. Y. Ou, J. Rapier, G. Schochetman, & M. L. Kalish. Human immunodeficiency virus type 1 subtypes in Malaysia include B C, and E. *AIDS Res Hum Retroviruses* 12:1655-1657, 1996.
- [Bruce et al.(1993)] C. Bruce, C. Clegg, A. Featherstone, J. Smith, & J. Oram. Sequence analysis of the gp120 region of the env gene of Ugandan human immunodeficiency proviruses from a single individual. *AIDS Res Hum Retroviruses* 9:357-63, 1993. OTE: (Medline: 93290931).
- [Buonaguro et al.(1995)] L. Buonaguro, G. D. E., M. Monaco, D. Greco, P. Corti, E. Beth-Giraldo, F. M. Buonaguro, & G. Giraldo. Heteroduplex mobility assay and phylogenetic analysis of V3 region sequences of human immunodeficiency virus type 1 isolates from Gulu, northern Uganda. The Italian-Ugandan Cooperation AIDS Program. *J Virol* 69:7971-7981, 1995. OTE: (Medline: 96079047).
- [Burger et al.(1991)] H. Burger, B. Weiser, K. Flaherty, J. Gulla, P. N. Nguyen, & R. A. Gibbs. Evolution of human immunodeficiency virus type 1 nucleotide sequence diversity among close contacts. *Proc Natl Acad Sci U S A* 88:11235-40, 1991. OTE: (Medline: 92107924), (Genbank: M77229 M77230 M77278).
- [Cabello et al.(1995)] A. Cabello, M. Cabral, M. B. Vera, R. Kiefer, R. M. Azorero, J. Eberle, L. Gurtler, & B. V. A. Analysis of the V3 loop sequences from 10 HIV type 1-infected AIDS patients from Paraguay. *AIDS Res. Hum. Retroviruses* 11:1135-7, 1995. OTE: (Medline: 96089222).
- [Calabro et al.(1995)] M. L. Calabro, C. Zanolto, F. Calderazzo, C. Crivellaro, M. D. A., R. D. A., & L. Chieco-Bianchi. HIV-1 infection of the thymus: evidence for a cytopathic and thymotropic viral variant in vivo. *AIDS Res Hum Retroviruses* 11:11-19, 1995. OTE: (Medline: 95251923).
- [Campodonico et al.(1996)] M. Campodonico, W. Janssens, L. Heyndrickx, K. Franssen, A. Leonaers, F. F. Fay, M. Taborda, der Van Groen, & O. H. Fay. HIV type 1 subtypes in Argentina and genetic heterogeneity of the V3 region. LA - Eng. *AIDS Res Hum Retroviruses* 12:79-81, 1996. OTE: (Medline: 96423022), (Genbank: U37030 U37031 U37032 U37033 U37034 U37035 U37036 U37037 U37038 U37039 U37040 U37041 U37042 U37043).
- [Carr et al.(1996)] J. K. Carr, M. O. Salminen, C. Koch, D. Gotte, A. W. Arntstein, P. A. Hegerich, L. S. D., D. S. Burke, & F. E. McCutchan. Full-length sequence and mosaic structure of a human

- immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* 70:5935-5943, 1996. OTE: (Medline: 96323109), (Genbank: U54771).
- [Cassol et al.(1996)] S. Cassol, B. Weniger, P. Babu, M. Htoon, A. Delaney, M. O'Shaughnessy, & C.-Y. Ou. Detection of hiv-1 env subtypes a, b, c, and e in asia using dried blood spots: A new surveillance tool for molecular epidemiology. *AIDS Res. Hum. Retroviruses* 12:1435-1441, 1996.
- [Chang et al.(1997)] K. Chang, C. Lin, J. Chen, C. Shih, H. C. Lin, R. Lin, S. Twu, & M. Salminen. Hiv-1 env gene diversity detected in taiwan. *AIDS Res. Hum. Retroviruses* 13(2):201-204, 1997. OTE: (Medline: 97159690).
- [Charneau et al.(1994)] P. Charneau, A. M. Borman, C. Quillent, D. Guetard, S. Chamaret, J. Cohen, G. Remy, L. Montagnier, & F. Clavel. Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group. *Virology* 205:247-253, 1994. OTE: (Medline: 95065659).
- [Chen et al.(1997)] M. Chen, M. K. Singh, R. Balachandran, & P. Gupta. Isolation and characterization of two divergent infectious molecular clones of HIV type 1 longitudinally obtained from a seropositive patient by a progressive amplification procedure. LA - Eng. *AIDS Res Hum Retroviruses* 13:743-50, 1997. OTE: (Medline: 97315044).
- [Cheng-Mayer et al.(1988)] C. Cheng-Mayer, J. Homsy, L. A. Evans, & J. A. Levy. Identification of human immunodeficiency virus subtypes with distinct patterns of sensitivity to serum neutralization. *Proc Natl Acad Sci U S A* 85:2815-9, 1988. OTE: (Medline: 88190159).
- [Cheng-Mayer et al.(1990)] C. Cheng-Mayer, M. Quiroga, J. W. Tung, D. Dina, & J. A. Levy. Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. *J Virol* 64:4390-8, 1990. OTE: (Medline: 90347835), (Genbank: M63024).
- [Cheng-Mayer et al.(1991)] C. Cheng-Mayer, T. Shioda, & J. A. Levy. Host range, replicative, and cytopathic properties of human immunodeficiency virus type 1 are determined by very few amino acid changes in tat and gp120. *J Virol* 65:6931-41, 1991. OTE: (Medline: 92046357).
- [Choi et al.(1997)] D. J. Choi, S. Dube, T. P. Spicer, H. B. Slade, F. C. Jensen, & B. J. Poiesz. HIV type 1 isolate Z321, the strain used to make a therapeutic HIV type 1 immunogen, is intersubtype recombinant. *AIDS Res Hum Retroviruses* 13:357-61, 1997. OTE: (Medline: 97225063), (Genbank: K02007 U50207 U50208 U76035).
- [Ciesielski et al.(1991)] C. A. Ciesielski, D. M. Bell, & D. W. Marianos. Transmission of HIV from infected health-care workers to patients. LA - Eng. *AIDS* 5 Suppl 2:S93-7, 1991. OTE: (Medline: 93264041).
- [Ciesielski et al.(1994)] C. A. Ciesielski, D. W. Marianos, G. Schochetman, J. J. Witte, & H. W. Jaffe. The 1990 Florida dental investigation: The press and the science. *Ann Intern Med* 121:886-888, 1994. OTE: (Medline: 95069397).
- [Clark et al.(1991)] S. J. Clark, M. S. Saag, W. D. Decker, S. Campbell-Hill, J. L. Roberson, P. J. Veldkamp, J. C. Kappes, B. H. Hahn, & G. M. Shaw. High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. LA - Eng. *N Engl J Med* 324:954-60, 1991. OTE: (Medline: 91163598).
- [Cleland et al.(1996)] A. Cleland, H. G. Watson, P. Robertson, C. A. Ludlam, & A. J. Brown. Evolution of zidovudine resistance-associated genotypes in human immunodeficiency virus type 1-infected patients. LA - Eng. *J Acquir Immune Defic Syndr Hum Retrovirol* 12:6-18, 1996. OTE: (Medline: 96242958), (Genbank: U45026 U45027 U45028 U45029 U45030 U45031 U45032 U45033 U45034 U45035 U45036 U45037 U45038 U45039 U45040 U45041 U45042 U45043 U45044 U45045 U45046 U45047 U45048 U45049 U45050 U45051 U45052 U45053 U45054 U45055).

Sequence Descriptions

- [Cocchi et al.(1996)] E. Cocchi, A. L. DeVico, A. Garzino-Demo, A. Cara, R. C. Gallo, & P. Lusso. The V3 domain of the HIV-1 gp120 envelope glycoprotein is critical for chemokine-mediated blockage of infection. *Nat Med* 270:1811-1815, 1996.
- [Cohen et al.(1995)] J. H. Cohen, D. Guetard, F. Philbert, S. Chamaret, T. Tabary, R. Deleys, & L. Montagnier. Atypical HIV serological profile of novel HIV-1 variant distinct from subtype O (letter) LA - Eng. *Lancet* 345:856, 1995. OTE: (Medline: 95206029), (Genbank: X83215 X83216).
- [Collman et al.(1992)] R. Collman, J. W. Balliet, S. A. Gregory, H. Friedman, D. L. Kolson, N. Nathanson, & A. Srinivasan. An infectious molecular clone of an unusual macrophage-tropic and highly cytopathic strain of human immunodeficiency virus type 1. *J Virol* 66:7517-21, 1992. OTE: (Medline: 93059708), (Genbank: M96155).
- [Contag et al.(1997)] C. H. Contag, A. Ehrnst, J. Duda, A. B. Bohlin, S. Lindgren, G. H. Learn, & J. I. Mullins. Mother-to-infant transmission of human immunodeficiency virus type 1 involving five envelope sequence subtypes. LA - Eng. *J Virol* 71:1292-300, 1997. OTE: (Medline: 97151117), (Genbank: U56263 U56264 U56265 U56266 U56267 U56268 U56269 U56270 U56271 U56272 U56273 U56274 U56275 U56276 U56277 U56278 U56279 U56280 U56281 U56282 U56283 U56284 U56285 U56286 U56287 U56288 U56289 U56290 U56291 U56292).
- [Cornelissen et al.(1995)] M. Cornelissen, G. Mulder-Kampinga, J. Veenstra, F. Zorgdrager, C. Kuiken, S. Hartman, J. Dekker, der van Hoek, C. Sol, & R. C. et al. Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. *J Virol* 69:1810-1818, 1995. OTE: (Medline: 95156613), (Genbank: Z47411 Z47540).
- [Couto-Fernandez et al.(1994)] J. C. Couto-Fernandez, W. Janssens, L. Heyndrickx, J. Motte, K. Franssen, M. Peeters, E. Delaporte, B. Galvao-Castro, P. Piot, & G. van der Groen. Genetic and antigenic variability of HIV type 1 in Brazil. *AIDS Res Hum Retroviruses* 10:1157-1163, 1994. OTE: (Medline: 95127298).
- [Crandall(1995)] K. A. Crandall. Intraspecific phylogenetics: support for dental transmission of human immunodeficiency virus. *J Virol* 69:2351-2356, 1995. OTE: (Medline: 95191010).
- [D. et al.(1996)] F. M. R. D., M. Jansson, E. Halapi, J. C. Russi, O. Libonatti, & H. Wigzell. Genetic analysis of V3 domain sequences obtained from Uruguayan HIV type 1-infected individuals. LA - Eng. *AIDS Res Hum Retroviruses* 12:1491-3, 1996. OTE: (Medline: 97048221).
- [da Costa et al.(1995)] S. M. da Costa, M. Schechter, N. Shindo, A. C. Vicente, E. F. Oliveira, M. E. Pinto, & A. Tanuri. Sequence and phylogenetic analysis of glycoprotein 120 of an HIV type 1 variant (GWOR) prevalent in Brazil. *AIDS Res. Hum. Retroviruses* 11:1143-5, 1995. OTE: (Medline: 96089224).
- [Daniels et al.(1991)] R. S. Daniels, M. H. Smith, & A. G. Fisher. Molecular characterization of biologically diverse envelope variants of human immunodeficiency virus type 1 derived from an individual. *J Virol* 65:5574-8, 1991. OTE: (Medline: 91374618).
- [de Jong et al.(1992)] J. J. de Jong, J. Goudsmit, W. Keulen, B. Klaver, W. J. A. Krone, M. Tersmette, & A. de Ronde. Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. *J Virol* 66:757-765, 1992.
- [De Wolf et al.(1994)] F. De Wolf, B. Hogervorst, J. Goudsmit, E. M. Fenyo, H. Rubsamen-Waigmann, H. Holmes, B. Galvao-Castro, C. W. E. Karita, & S. D. Sempala. Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics. WHO Network for HIV Isolation and Characterization. *AIDS Res Hum Retroviruses* 10:1387-1400, 1994. OTE: (Medline: 95194697).
- [Deacon et al.(1995)] N. J. Deacon, A. Tsykin, A. Solomon, K. Smith, M. Ludford-Menting, D. J. Hooker, D. A. McPhee, A. L. Greenway, A. Ellett, & C. C. et al. Genomic structure of an attenuated quasi

- species of HIV-1 from a blood transfusion donor and recipients. *Science* 270:988-991, 1995. OTE: (Medline: 96069819).
- [DeBry et al.(1993)] R. W. DeBry, L. G. Abele, S. H. Weiss, M. D. Hill, M. Bouzas, E. Lorenzo, F. Graebnitz, & L. Resnick. Dental HIV transmission. *Nature* 361:691, 1993. OTE: (Medline: 93180905).
- [Delaporte et al.(1996)] E. Delaporte, W. Janssens, M. Peeters, A. Buve, G. Dibanga, J. Perret, V. Ditsambou, M. G. Courbot, A. Georges, A. Bourgeois, B. Samb, D. Henzel, L. Heyndrickx, K. Fransen, G. V. der Groen, B. Larouz, & J. Mbet. Epidemiological and molecular characteristics of hiv infection in gabon (1986 - 1994). *AIDS* 10:903-910, 1996.
- [Delwart et al.(1995)] E. L. Delwart, M. P. Busch, M. L. Kalish, J. W. Mosley, & J. I. Mullins. Rapid molecular epidemiology of human immunodeficiency virus transmission. *AIDS Res. Hum. Retroviruses* 11:1081-93, 1995. OTE: (Medline: 96089215).
- [Delwart et al.(1994)] E. L. Delwart, H. W. Sheppard, B. D. Walker, J. Goudsmit, & J. I. Mullins. Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays. *J Virol* 68:6672-6683, 1994. OTE: (Medline: 94365971), (Genbank: U00821 U00822 U00832 U00833 U00834 U00837 U00839 U13240 U13241 U13242 U13243 U13244 U13245 U13246 U13247 U13248 U13249 U13250 U13251 U13252 U13373 U13374 U13375 U13376 U13377 U13378 U13379 U13380 U13381 U13382).
- [Desai et al.(1986)] S. M. Desai, V. S. Kalyanaraman, J. M. Casey, A. Srinivasan, P. R. Andersen, & S. G. Devare. Molecular cloning and primary nucleotide sequence analysis of a distinct human immunodeficiency virus isolate reveal significant divergence in its genomic sequences. *Proc Natl Acad Sci U S A* 83:8380-4, 1986. OTE: (Medline: 87041461), (Genbank: M13136 M13137).
- [Diaz et al.(1995)] R. S. Diaz, E. C. Sabino, A. Mayer, J. W. Mosley, & M. P. Busch. Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. The Transfusion Safety Study Group. *J Virol* 69:3273-3281, 1995. OTE: (Medline: 95264414), (Genbank: U11124 U11125 U11126 U11127 U11128 U11129 U11130 U11131 U11132 U11133 U11134 U11135 U11136 U11137 U11138 U11139 U11140 U11141 U11142 U11143 U11144 U11145 U11146 U11147 U11148 U11149 U11150 U11151 U11152 U11153).
- [Dietrich et al.(1993)] U. Dietrich, M. Grez, H. von Briesen, B. Panhans, M. Geissendorfer, H. Kuhnelt, J. Maniar, G. Mahambre, W. B. Becker, & M. L. B. et al. HIV-1 strains from India are highly divergent from prototypic African and US/European strains, but are linked to a South African isolate. *AIDS* 7:23-7, 1993. OTE: (Medline: 93183429), (Genbank: X65638 X65639 X65640 X68406 X68407 X65638 X65639 X65640 X68406 X68407).
- [Douglas et al.(1996)] N. Douglas, A. Knight, A. Hayhurst, W. Barrett, M. Kevany, & R. Daniels. An efficient method for the rescue and analysis of functional hiv-1 env genes: evidence for recombination in the vicinity of the tat/rev splice site. *AIDS* 10:39-46, 1996.
- [Duensing et al.(1995)] T. D. Duensing, H. Fang, D. W. Dorward, & S. H. Pincus. Processing of the envelope glycoprotein gp160 in immunotoxin-resistant cell lines chronically infected with human immunodeficiency virus type 1. *J Virol* 69:7122-7131, 1995. OTE: (Medline: 96013815).
- [Dumitrescu et al.(1994)] O. Dumitrescu, M. L. Kalish, S. C. Kliks, C. I. Bandea, & J. A. Levy. Characterization of human immunodeficiency virus type 1 isolates from children in Romania: identification of a new envelope subtype. *J Infect Dis* 169:281-8, 1994. OTE: (Medline: 94149311).
- [Engelbrecht et al.(1995)] S. Engelbrecht, J. D. Laten, T. L. Smith, & R. van E. J. Identification of env subtypes in fourteen HIV type 1 isolates from south Africa. LA - Eng. *AIDS Res Hum Retroviruses* 11:1269-71, 1995. OTE: (Medline: 96157218), (Genbank: L47608 L48061 L48062 L48063 L48064 L48065 L48066 L48067 L48068 L48069 L48070 L48071 L48072 L48073 U33769 U33770 U33771 U33772 U33773 U33774 U33775 U33776 U33777 U33778 U33779 U33780 U33781 U33782).
- [Engelstad(1996)] M. Engelstad. Genetic diversity of env V3 loop sequences in HIV type 1-infected homosexuals, heterosexuals, intravenous drug users, and hemophilia patients in Norway: high prevalence

Sequence Descriptions

- of subtype B and detection of subtype C. LA - Eng. *AIDS Res Hum Retroviruses* 12:1733-8, 1996. OTE: (Medline: 97118369), (Genbank: X92902 X92903 X92904 X92905 X92906 X92907 X92908 X92909 X92910 X92911 X92912 X92913 X92914 X92915 X92916 X92917 X92918 X92919 X92920 X92921 X92922 X92923 X92924 X92925 X92926 X92927 X92928 X92929 X92930 X92931).
- [et al(1992)] T. F. W. W. et al. Genetic variation of the third variable envelope region in natural HIV-1 infection. *Doctoral Thesis, Amsterdam* 1992.
- [Fisher et al.(1988)] A. G. Fisher, B. Ensoli, D. Looney, A. Rose, R. C. Gallo, M. S. Saag, G. M. Shaw, B. H. Hahn, & F. Wong-Staal. Biologically diverse molecular variants within a single HIV-1 isolate. *Nature* 334:444-7, 1988. OTE: (Medline: 88302422).
- [Fitzgibbon et al.(1993)] J. E. Fitzgibbon, S. Gaur, L. D. Frenkel, F. Laraque, B. R. Edlin, & D. T. Rubin. Transmission from one child to another of human immunodeficiency virus type 1 with a zidovudine-resistance mutation (see comments). *N Engl J Med* 329:1835-41, 1993. OTE: (Medline: 94067210).
- [Fouchier et al.(1992)] R. A. Fouchier, M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, & H. Schuitemaker. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol* 66:3183-7, 1992. OTE: (Medline: 92219412).
- [Fredriksson et al.(1991)] R. Fredriksson, P. Stalhanske, G. von A., B. Lind, P. Aman, E. Rassart, & E. M. Fenyo. Biological characterization of infectious molecular clones derived from a human immunodeficiency virus type-1 isolate with rapid/high replicative capacity. LA - Eng. *Virology* 181:55-61, 1991. OTE: (Medline: 91135020).
- [Furuta et al.(1994)] Y. Furuta, T. Bergstrom, G. Norkrans, & P. Horal. HIV type 1 V3 sequence diversity in contact-traced Swedish couples at the time of sexual transmission. *AIDS Res Hum Retroviruses* 10:1187-1189, 1994. OTE: (Medline: 95127302), (Genbank: U10929 U10930 U10931 U10932 U10933 U10934 U10935 U10936 U10937 U10938 U10939 U10940 U10941 U10942 U10943 U10944 U10945 U10946 U10947 U10948 U10949 U10950).
- [Ganeshan et al.(1997)] S. Ganeshan, R. E. Dickover, B. T. Korber, Y. J. Bryson, & S. M. Wolinsky. Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. LA - Eng. *J Virol* 71:663-77, 1997. OTE: (Medline: 97138372), (Genbank: U47950 U47951 U47952 U47953 U47954 U47955 U47956 U47957 U47958 U47959 U47960 U47961 U47962 U47963 U47964 U47965 U47966 U47967 U47968 U47969 U47970 U47971 U47972 U47973 U47974 U47975 U47976 U47977 U47978 U47979).
- [Gao et al.(1996a)] F. Gao, S. G. Morrison, D. L. Robertson, C. L. Thornton, S. Craig, G. Karlsson, J. Sodroski, M. Morgado, B. Galvao-Castro, H. von Briesen, S. Beddows, J. Weber, P. M. Sharp, G. M. Shaw, & B. H. Hahn. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. *J. Virol.* 70:1651-1657, 1996a. OTE: (Medline: 96190564), (Genbank: L34667 U02739 U04908 U08442 U08443 U08444 U08445 U08446 U08447 U08448 U08450 U08451 U08452 U08453 U08454 U08455 U08456 U08457 U08794 U08797 U08801 U09126 U09127 U09131 U27419 U27434 U27443 U43386).
- [Gao et al.(1996b)] F. Gao, D. L. Robertson, S. G. S. G. Morrison, H. Hui, S. Craig, P. N. Fultz, J. Decker, M. Girard, G. M. Shaw, B. H. Hahn, & P. M. Sharp. The heterosexual HIV-1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virology* 70:7013-7029, 1996b.
- [Gao et al.(1994a)] F. Gao, L. Yue, S. Craig, C. L. Thornton, D. L. Robertson, F. E. McCutchan, J. A. Bradac, P. M. Sharp, & B. H. Hahn. Genetic variation of HIV type 1 in four World Health Organization-sponsored vaccine evaluation sites: generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. WHO Network for HIV Isolation and Characterization. *AIDS Res Hum Retroviruses* 10:1359-1368, 1994a. OTE: (Medline: 95194694), (Genbank: L34667 U08793 U08794 U08795 U08796 U08797 U08802 U08804 U08810 U09131).

- [Gao et al.(1994b)] F. Gao, L. Yue, S. Craig, C. L. Thornton, D. L. Robertson, F. E. McCutchan, J. A. Bradac, P. M. Sharp, B. H. Hahn, & the WHO Network for HIV Isolation and Characterization. HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. *AIDS Res Hum Retroviruses* 10:1327-1343, 1994b. OTE: (Medline: 95194691).
- [Gao et al.(1994c)] F. Gao, L. Yue, S. C. Hill, D. L. Robertson, A. H. Graves, M. S. Saag, G. M. Shaw, P. M. Sharp, & B. H. Hahn. HIV-1 sequence subtype D in the United States. *AIDS Res Hum Retroviruses* 10:625-627, 1994c. OTE: (Medline: 95000937), (Genbank: U08192 U08193).
- [Ghosh et al.(1993)] S. K. Ghosh, P. N. Pultz, E. Keddle, M. S. Saag, P. M. Sharp, B. H. Hahn, & G. M. Shaw. A molecular clone of HIV-1 tropic and cytopathic for human and chimpanzee lymphocytes. *Virology* 194:858-64, 1993. OTE: (Medline: 93276579), (Genbank: L02317).
- [Gomez et al.(1996)] C. E. Gomez, J. R. Fernandez, E. Iglesias, A. E. Lopez, L. Lobaina, E. Noa, H. Diaz, A. Herrera, F. Rolo, & C. A. Duarte. Complete DNA sequence of the gene encoding the external glycoprotein (gp120) from a Cuban HIV type 1 isolate. *AIDS Res. Hum. Retroviruses* 12:553-555, 1996. OTE: (Medline: 96296968).
- [Goodenow et al.(1989)] M. Goodenow, T. Huet, W. Saurin, S. Kwok, J. Sninsky, & S. Wain-Hobson. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J Acquir Immune Defic Syndr* 2:344-52, 1989. OTE: (Medline: 89328783).
- [Grez et al.(1994)] M. Grez, U. Dietrich, P. Balfe, H. von Briesen, J. K. Maniar, G. Mahambre, E. L. Delwart, J. I. Mullins, & H. Rubsamen-Waigmann. Genetic analysis of human immunodeficiency virus type 1 and 2 (HIV-1 and HIV-2) mixed infections in India reveals a recent spread of HIV-1 and HIV-2 from a single ancestor for each of these viruses. *J Virol* 68:2161-8, 1994. OTE: (Medline: 94187055).
- [Groenink et al.(1992)] M. Groenink, A. C. Andeweg, R. A. Fouchier, S. Broersen, der van Jagt, H. Schuitemaker, G. de R. B., M. L. Bosch, H. G. Huisman, & M. Tersmette. Phenotype-associated env gene variation among eight related human immunodeficiency virus type 1 clones: evidence for in vivo recombination and determinants of cytotropism outside the V3 domain. *LA - Eng. J Virol* 66:6175-80, 1992. OTE: (Medline: 92408013).
- [Guillon et al.(1995)] C. Guillon, F. Bedin, R. A. Fouchier, H. Schuitemaker, & R. A. Gruters. Completion of nucleotide sequences of non-syncytium-inducing and syncytium-inducing HIV type 1 variants isolated from the same patient. *LA - Eng. AIDS Res Hum Retroviruses* 11:1537-41, 1995. OTE: (Medline: 96288454), (Genbank: L06678 L06706 M64490 M64491 U34603 U34604).
- [Guo et al.(1991)] H. G. Guo, J. C. Chermann, D. Waters, L. Hall, A. Louie, R. C. Gallo, H. Streicher, M. S. Reitz, M. Popovic, & W. Blattner. Sequence analysis of original HIV-1. *Nature* 349:745-6, 1991. OTE: (Medline: 91156044), (Genbank: X57446 X57447 X57448 X57449 X57450 X57451 X57452 X57453 X57454 X57455).
- [Gurgo et al.(1988)] C. Gurgo, H. G. Guo, G. Franchini, A. Aldovini, E. Collalti, K. Farrell, F. Wong-Staal, R. C. Gallo, & M. S. R. Jr. Envelope sequences of two new United States HIV-1 isolates. *Virology* 164:531-6, 1988. OTE: (Medline: 88219542), (Genbank: M17449 M17450).
- [Gurtler et al.(1994)] L. G. Gurtler, P. H. Hauser, J. Eberle, A. von Brunn, S. Knapp, L. Zekeng, J. M. Tsague, & L. Kaptue. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J Virol* 68:1581-5, 1994. OTE: (Medline: 94149848).
- [Hahn et al.(1986)] B. H. Hahn, G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, & W. P. Parks. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* 232:1548-53, 1986. OTE: (Medline: 86235450), (Genbank: M12507).
- [Hattori et al.(1991)] T. Hattori, K. Shiozaki, Y. Eda, S. Tokiyoshi, S. Matsushita, H. Inaba, M. Fujimaki, T. Meguro, K. Yamada, & M. H. et al. Characteristics of the principal neutralizing determinant of

Sequence Descriptions

- HIV-1 prevalent in Japan. *AIDS Res Hum Retroviruses* 7:325-30, 1991. OTE: (Medline: 92075338), (Genbank: M81693 M81694 S70936).
- [Henikoff & Henikoff(1992)] S. Henikoff & J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915-10919, 1992.
- [Henikoff & Henikoff(1993)] S. Henikoff & J. G. Henikoff. Performance evaluation of amino acid substitution matrices. *Proteins: Structure, Function and Genetics* 17:49-61, 1993.
- [Heyndrickx et al.(1996)] L. Heyndrickx, W. Janssens, M. Alary, K. Franssen, K. Vereecken, S. Coppens, B. Willems, N. Davo, A. Guedeme, E. Baganizi, J. Joly, & der Van Groen. Genetic variability of HIV type 1 in Benin. LA - Eng. *AIDS Res Hum Retroviruses* 12:1495-7, 1996. OTE: (Medline: 97048222), (Genbank: U61854 U61855 U61856 U61857 U61858 U61859 U61860 U61861 U61862 U61863 U61864 U61865 U61866 U61867 U61868 U61869 U61870 U61871 U61872 U61873 U61874).
- [Hillis & Huelsenbeck(1994)] D. M. Hillis & J. P. Huelsenbeck. Support for dental HIV transmission. *Nature* 369:24-5, 1994. OTE: (Medline: 94217803), (Genbank: U06872 U06873 U06874 U06875 U06876 U06877 U06878 U06879 U06880 U06881 U06882 U06883 U06884 U06885 U06886 U06887 U06888 U06889 U06890 U06891 U06892 U06893 U06894 U06895 U06896 U06897 U06898 U06899 U06900 U06901).
- [Ho et al.(1989)] D. D. Ho, H. S. Robin, M. Alam, B. J. Wallace, & Y. Mizrahi. Human immunodeficiency virus type 1 in a seronegative patient with visceral Kaposi's sarcoma and hypogammaglobulinemia. *American J of Med* 86:349-51, 1989. OTE: (Medline: 89148307).
- [Holm-Hansen et al.(1995)] C. Holm-Hansen, D. Grothues, S. Rustad, B. Rosok, F. R. Pascu, & B. Asjo. Characterization of HIV type 1 from Romanian children: lack of correlation between V3 loop amino acid sequence and syncytium formation in MT-2 cells. *AIDS Res Hum Retroviruses* 11:597-603, 1995. OTE: (Medline: 96093895), (Genbank: X77968 X77969 X77970 X77971 X77972 X77973 X77974 X77975 X77976 X77977 X77978 X77979 X77980 X77981 X77982 X77983 X77984 X77985 X77986 X77987).
- [Holmes et al.(1992)] E. C. Holmes, L. Q. Zhang, P. Simmonds, C. A. Ludlam, & A. J. Brown. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc Natl Acad Sci U S A* 89:4835-9, 1992. OTE: (Medline: 92279221), (Genbank: M84240 M84241 M84242 M84243 M84244 M84245 M84246 M84247 M84248 M84249 M84250 M84251 M84252 M84253 M84254 M84255 M84256 M84257 M84258 M84259 M84260 M84261 M84262 M84263 M84264 M84265 M84266 M84267 M84268 M84269).
- [Howard et al.(1994)] T. M. Howard, D. O. Olayele, & S. Rasheed. Sequence analysis of the glycoprotein 120 coding region of a new HIV type 1 subtype A strain (HIV-1IbNg) from Nigeria. *AIDS Res Hum Retroviruses* 10:1755-1757, 1994. OTE: (Medline: 95194745).
- [Howard & Rasheed(1996)] T. M. Howard & S. Rasheed. Genomic structure and nucleotide sequence analysis of a new hiv type 1 subtype a strain from nigeria. *AIDS Res. Hum. Retroviruses* 12:1413-1425, 1996.
- [Huet et al.(1989)] T. Huet, M. C. Dazza, F. Brun-Vezinet, G. E. Roelants, & S. Wain-Hobson. A highly defective HIV-1 strain isolated from a healthy Gabonese individual presenting an atypical western blot. *AIDS* 3:707-15, 1989. OTE: (Medline: 90148544).
- [Hutto et al.(1996)] C. Hutto, Y. Zhou, J. He, R. Geffin, M. Hill, W. Scott, & C. Wood. Longitudinal studies of viral sequence, viral phenotype, and immunologic parameters of human immunodeficiency virus type 1 infection in perinatally infected twins with discordant disease courses. *J. Virol* 70:3589-3598, 1996. OTE: (Medline: 96211491).
- [Hwang et al.(1991)] S. S. Hwang, T. J. Boyle, H. K. Lyerly, & B. R. Cullen. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* 253:71-4, 1991. OTE: (Medline: 91289160).

- [Ishikawa et al.(1996)] K. Ishikawa, W. Janssens, J. Brandful, L. Heyndrickx, Y. Takebe, W. Ampofo, T. Sata, S. Yamazaki, M. Osei-Kwasi, N. Yamamoto, Y. Koyanagi, der Van Groen, & T. Kurata. Genetic and phylogenetic analysis of HIV type 1 env subtypes in Ghana, west Africa. LA - Eng. *AIDS Res Hum Retroviruses* 12:1575-8, 1996. OTE: (Medline: 97068257), (Genbank: U67682 U67683 U67684 U67685 U67686 U67687 U67688 U67689 U67690 U67691 U67692 U67693 U67694 U67696 U67697 U67698 U67699 U67700).
- [Jaffe et al.(1994)] H. W. Jaffe, J. M. McCurdy, M. L. Kalish, T. Liberti, G. Metellus, B. H. Bowman, S. B. Richards, A. R. Neasman, & J. J. Witte. Lack of HIV transmission in the practice of a dentist with AIDS. *Ann Intern Med* 121:855-859, 1994. OTE: (Medline: 95069391).
- [Janssens et al.(1994a)] W. Janssens, L. Heyndrickx, de Van Peer, A. Bouckaert, K. Fransen, J. Motte, G. M. Gershy-Damet, M. Peeters, P. Piot, & G. van der Groen. Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d'Ivoire. *AIDS* 8:21-6, 1994a. OTE: (Medline: 94280700).
- [Janssens et al.(1994b)] W. Janssens, L. Heyndrickx, K. Fransen, J. Motte, M. Peeters, J. N. Nkengasong, P. M. Ndumbe, E. Delaporte, J. L. Perret, & C. A. et al. Genetic and phylogenetic analysis of env subtypes G and H in central Africa. *AIDS Res Hum Retroviruses* 10:877-879, 1994b. OTE: (Medline: 95078007), (Genbank: U09664 U09665 U09666 U09667).
- [Janssens et al.(1994c)] W. Janssens, L. Heyndrickx, K. Fransen, M. Temmerman, A. Leonaers, T. Ivens, J. Motte, P. Piot, & G. van der Groen. Genetic variability of HIV type 1 in Kenya. *AIDS Res Hum Retroviruses* 10:1577-1579, 1994c. OTE: (Medline: 95194720), (Genbank: U12984 U12985 U12986 U12987 U12988 U12989 U12990 U12991 U12992 U12993 U12994 U12995 U12996 U12997 U12998 U12999 U13000 U13001 U13002 U13003 U13004 U13005 U13006).
- [Janssens et al.(1994d)] W. Janssens, J. N. Nkengasong, L. Heyndrickx, K. Fransen, P. M. Ndumbe, E. Delaporte, M. Peeters, J. L. Perret, A. Ndoumou, & C. Atende. Further evidence of the presence of genetically aberrant HIV-1 strains in Cameroon and Gabon. *AIDS* 8:1012-3, 1994d. OTE: (Medline: 95032906), (Genbank: X78476 X78477).
- [Jansson et al.(1996)] M. Jansson, M. Popovic, A. Karlsson, F. Cocchi, P. Rossi, J. Albert, & H. Wigzell. Sensitivity to inhibition by beta-chemokines correlates with biological phenotypes of primary HIV-1 isolates. LA - Eng. *Proc Natl Acad Sci U S A* 93:15382-7, 1996. OTE: (Medline: 97140339), (Genbank: U76078 U76079 U76080 U76081 U76082 U76083 U76084 U76085 U76086 U76087).
- [Jr et al.(1996)] G. H. L. Jr, B. T. Korber, B. Foley, B. H. Hahn, S. M. Wolinsky, & J. I. Mullins. Maintaining the integrity of human immunodeficiency virus sequence databases. LA - Eng. *J Virol* 70:5720-30, 1996. OTE: (Medline: 96357112).
- [Jr et al.(1993)] M. P. Jr, M. S. Saag, L. C. Yang, S. J. Clark, J. C. Kappes, K. C. Luk, B. H. Hahn, G. M. Shaw, & J. D. Lifson. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. LA - Eng. *Science* 259:1749-54, 1993. OTE: (Medline: 93206118).
- [Jr et al.(1992)] M. S. R. Jr, H. G. Guo, J. Oleske, J. Hoxie, M. Popovic, E. Read-Connole, P. Markham, H. Streicher, & R. C. Gallo. On the historical origins of HIV-1 (MN) and (RF). *AIDS Res Hum Retroviruses* 8:1950, 1992. OTE: (Medline: 93144005).
- [Jr et al.(1994)] M. S. R. Jr, L. Hall, M. Robert-Guroff, J. Lautenberger, B. M. Hahn, G. M. Shaw, L. I. Kong, S. H. Weiss, D. Waters, & R. C. G. et al. Viral variability and serum antibody response in a laboratory worker infected with HIV type 1 (HTLV type IIIB). *AIDS Res Hum Retroviruses* 10:1143-1155, 1994. OTE: (Medline: 95127297).
- [Kaleebu et al.(1995)] P. Kaleebu, A. Bobkov, R. Cheingsong-Popov, P. Bieniasz, M. Garaev, & J. Weber. Identification of HIV-1 subtype G from Uganda. *AIDS Res Hum Retroviruses* 11:657-659, 1995. OTE: (Medline: 96093903), (Genbank: U22010).
- [Kalish et al.(1995)] M. L. Kalish, A. Baldwin, S. Raktham, C. Wasi, C. C. Luo, G. Schochetman, T. D. Mastro, N. Young, S. Vanichseni, & H. R. et al. The evolving molecular epidemiology of HIV-1

Sequence Descriptions

- envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. *AIDS* 9:851-857, 1995. OTE: (Medline: 96014957).
- [Kalish et al.(1994)] M. L. Kalish, C. C. Luo, B. G. Weniger, K. Limpakarnjanarat, N. Young, C. Y. Ou, & G. Schochetman. Early HIV type 1 strains in Thailand were not responsible for the current epidemic. *AIDS Res Hum Retroviruses* 10:1573-1575, 1994. OTE: (Medline: 95194719), (Genbank: L15724 L15728 L15730 L19237 L19238 L19239 L19571 L19579 L19634 U15576 U15577 U15578 U15579 U15580 U15581 U15582 U15583 U15584 U15585 U15586 U15587 U15588).
- [Kampinga et al.(1997)] G. A. Kampinga, A. Simonon, de Van Perre, E. Karita, P. Msellati, & J. Goudsmit. Primary infections with HIV-1 of women and their offspring in Rwanda: findings of heterogeneity at seroconversion, coinfection, and recombinants of HIV-1 subtypes A and C. *Virology* 227:63-76, 1997. OTE: (Medline: 97167518), (Genbank: Z47881 Z47882 Z47883 Z47884 Z47885 Z47886 Z47887 Z47888 Z47889 Z47890 Z47891 Z47951 Z47952 Z47953 Z47954 Z47955 Z47956 Z47957 Z47958 Z47959 Z47960 Z47961 Z47962 Z47963 Z47964 Z47965 Z75958 Z75959 Z75960 Z75961).
- [Kasper et al.(1994)] P. Kasper, R. Kaiser, J. Oldenburg, H. H. Brackmann, B. Matz, & K. E. Schneweis. Parallel evolution in the V3 region of HIV type 1 after infection of hemophiliacs from a homogeneous source. *AIDS Res Hum Retroviruses* 10:1669-1678, 1994. OTE: (Medline: 95194733).
- [Keys et al.(1993)] B. Keys, J. Karis, B. Fadeel, A. Valentin, G. Norkrans, L. Hagberg, & P. Chiodi. V3 sequences of paired HIV-1 isolates from blood and cerebrospinal fluid cluster according to host and show variation related to the clinical stage of disease. *Virology* 196:475-83, 1993. OTE: (Medline: 93383372), (Genbank: Z23177 Z23178 Z23179 Z23180 Z23181 Z23182 Z23183 Z23184 Z23185 Z23186 Z23187 Z23188 Z23190 Z23191 Z23192 Z23193 Z23194 Z23195 Z23196 Z23197 Z23198 Z23199 Z23200 Z23201 Z23202 Z23203 Z23204 Z23205 Z23206 Z23207).
- [Kim et al.(1995)] F. M. Kim, D. L. Kolson, J. W. Balliet, A. Srinivasan, & R. G. Collman. V3-independent determinants of macrophage tropism in a primary human immunodeficiency virus type 1 isolate. *J Virol* 69:1755-1761, 1995. OTE: (Medline: 95156606).
- [Klasse et al.(1996)] P. J. Klasse, M. T. Boyd, R. A. Weiss, & T. F. Schulz. Mutations in the vpu, env and nef genes of a syncytium-inducing variant of HIV type 1 JR-CSF that infects a range of T-cell lines. *AIDS Res. Hum. Retroviruses* 12:347-350, 1996.
- [Komiyama et al.(1989)] N. Komiyama, N. Hattori, J. Inoue, S. Sakuma, T. Kurimura, & M. Yoshida. Nucleotide sequences of gag and env genes of a Japanese isolate of HIV- 1 and their expression in bacteria. *AIDS Res Hum Retroviruses* 5:411-9, 1989. OTE: (Medline: 89352108).
- [Korber & Myers(1992)] B. Korber & G. Myers. Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Res Hum Retroviruses* 8:1549-60, 1992. OTE: (Medline: 93090472), (Genbank: M90847 M90848 M90849 M90850 M90851 M90852 M90853 M90854 M90855 M90856 M90857 M90858 M90859 M90860 M90861 M90862 M90863 M90864 M90865 M90866 M90867 M90868 M90869 M90870 M90871 M90872 M90873 M90874 M90875 M90876).
- [Korber et al.(1994)] B. T. Korber, K. J. Kunstman, B. K. Patterson, M. Furtado, M. M. McEvilly, R. Levy, & S. M. Wolinsky. Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain- derived sequences. *J Virol* 68:7467-7481, 1994. OTE: (Medline: 95018660), (Genbank: U05360 U05361 U05362 U05363 U05364 U05365 U05366 U05367 U05368 U05369 U05370 U05371 U05372 U05373 U05374 U05375 U05376 U05377 U05378 U05379 U05380 U05381 U05382 U05383 U05384 U05385 U05386 U05387 U05388 U05389).
- [Korber et al.(1995)] B. T. Korber, G. Learn, J. I. Mullins, B. H. Hahn, & S. Wolinsky. Protecting HIV databases. *Nature* 378:242-244, 1995. OTE: (Medline: 96069755).
- [Kostrikis et al.(1995)] L. G. Kostrikis, E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, & D. D. Ho. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a

- new subtype designated subtype I. *J Virol* 69:6122-6130, 1995. OTE: (Medline: 95395946), (Genbank: L19573 L19576 L22084 L22954 L22955 M15896 M38428 M66533 U28321 U28661 U28662 U28663 U28664 U28665 U28666 U28667 U28668 U28669 U28670 U28671 U28672 U28673 U28674 U28675 U28676 U28677 U28678 U28679 U28680 U28719).
- [Koyanagi et al.(1987)] Y. Koyanagi, S. Miles, R. T. Mitsuyasu, J. E. Merrill, H. V. Vinters, & I. S. Chen. Dual infection of the central nervous system by AIDS viruses with distinct cellular tropisms. *LA - Eng. Science* 236:819-22, 1987. OTE: (Medline: 87206194).
- [Kreutz et al.(1992)] R. Kreutz, U. Dietrich, H. Kuhnel, K. Nieselt-Struwe, M. Eigen, & H. Rubsamen-Waigmann. Analysis of the envelope region of the highly divergent HIV-2ALT isolate extends the known range of variability within the primate immunodeficiency viruses. *AIDS Res Hum Retroviruses* 8:1619-29, 1992. OTE: (Medline: 93090480), (Genbank: X61240 X61240).
- [Kuiken et al.(1996a)] C. Kuiken, V. Lukashov, E. Baan, J. dekker, J. Leunissen, & J. Goudsmit. Evidence for limited intra-subject evolution of the v3 domain of the hiv-1 envelope in the amsterdam population. *AIDS*. In Press, 1996a.
- [Kuiken et al.(1996b)] C. L. Kuiken, M. T. Cornelissen, F. Zorgdrager, S. Hartman, A. J. Gibbs, & J. Goudsmit. Consistent risk group-associated differences in human immunodeficiency virus type 1 vpr, vpu and V3 sequences despite independent evolution. *J. Gen. Virol.* 77:783-792, 1996b. OTE: (Medline: 96203966), (Genbank: Z29280 Z29296 Z29298 Z29299 Z29302 Z29303 Z29304 Z29306 Z29307 Z29310 Z29311 Z29312 Z29315 Z29319 Z29320 Z29321 Z29322 Z29323 Z29324 Z68033 Z68047 Z68062 Z68070 Z68074 Z68086 Z68508 Z68509 Z68510 Z68511 Z68512).
- [Kuiken et al.(1995)] C. L. Kuiken, J. Goudsmit, G. F. Weiller, J. S. Armstrong, S. Hartman, P. Portegies, J. Dekker, & M. Cornelissen. Differences in human immunodeficiency virus type 1 V3 sequences from patients with and without AIDS dementia complex. *J Gen Virol* 76:175-180, 1995. OTE: (Medline: 95146947), (Genbank: Z37734 Z37735 Z37736 Z37737 Z37738 Z37739 Z37740 Z37741 Z37742 Z37743 Z37744 Z37745 Z37746 Z37747 Z37748 Z37749 Z37750 Z37751 Z37752 Z37753 Z37754 Z37755 Z37756 Z37757 Z37758 Z37759 Z37760 Z37761 Z37762 Z37763).
- [Kuiken et al.(1993)] C. L. Kuiken, G. Zwart, E. Baan, R. A. Coutinho, J. A. van den Hoek, & J. Goudsmit. Increasing antigenic and genetic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of the AIDS epidemic. *Proc Natl Acad Sci U S A* 90:9061-5, 1993. OTE: (Medline: 94022315).
- [Kusumi et al.(1992)] K. Kusumi, B. Conway, S. Cunningham, A. Berson, C. Evans, A. K. Iversen, D. Colvin, M. V. Gallo, S. Coutre, & E. G. S. et al. Human immunodeficiency virus type 1 envelope gene structure and diversity in vivo and after cocultivation in vitro. *J Virol* 66:875-85, 1992. OTE: (Medline: 92114191), (Genbank: M79342 M79343 M79344 M79345 M79346 M79347 M79348 M79349 M79350 M79351 M79352 M79353 M79354).
- [Lasky et al.(1997)] M. Lasky, J. L. Perret, M. Peeters, F. Bibollet-Ruche, F. Liegeois, D. Patrel, S. Molinier, C. Gras, & E. Delaporte. Presence of multiple non-B subtypes and divergent subtype B strains of HIV-1 in individuals infected after overseas deployment. *LA - Eng. AIDS* 11:43-51, 1997. OTE: (Medline: 97264202).
- [Leigh Brown & Cleland(1996)] A. J. Leigh Brown & A. Cleland. Independent evolution of the env and pol genes of HIV-1 during zidovudine therapy. *LA - Eng. AIDS* 10:1067-73, 1996. OTE: (Medline: 97028609), (Genbank: U58119 U58393 U58394 U58395 U58396 U58397 U58398 U58399 U58400 U58401 U58402 U58403 U58404 U58405 U58406 U58407 U58408 U58409 U58410 U58411 U58412 U58413 U58414 U58415 U58416 U58417 U58418 U58420 U58421 U58422).
- [Leitner et al.(1995)] T. Leitner, A. Alaeus, S. Marquina, E. Lilja, K. Lidman, & J. Albert. Yet another subtype of HIV type 1. *AIDS Res Hum Retroviruses* 11:995-997, 1995. OTE: (Medline: 96020102), (Genbank: L41176 L41177 L41178 L41179).

Sequence Descriptions

- [Leitner et al.(1996a)] T. Leitner, D. Escanilla, C. Franzen, M. Uhlen, & J. Albert. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc. Natl. Acad. Sci. U.S.A.* 20:10864-10869, 1996a.
- [Leitner et al.(1996b)] T. Leitner, G. Korovina, S. Marquina, T. Smolskaya, & J. Albert. Molecular epidemiology and MT-2 cell tropism of Russian HIV type 1 variant. LA - Eng. *AIDS Res Hum Retroviruses* 12:1595-603, 1996b. OTE: (Medline: 97102922), (Genbank: U69646 U69647 U69648 U69649 U69650 U69651 U69652 U69653 U69654 U69655 U69656 U69657 U69658 U69659 U69660 U69661 U69662 U69663 U69664 U69665 U69667).
- [Lepage et al.(1996)] P. Lepage, P. Msellati, D. G. Hitimana, A. Bazubagira, G. V. C., A. Simonon, E. Karita, L. Dequae-Merchadou, de Van Perre, & F. Dabis. Growth of human immunodeficiency type 1-infected and uninfected children: a prospective cohort study in Kigali, Rwanda, 1988 to 1993. LA - Eng. *Pediatr Infect Dis J* 15:479-85, 1996. OTE: (Medline: 96377503).
- [Levy et al.(1984)] J. A. Levy, A. D. Hoffman, S. M. Kramer, J. A. Landis, J. M. Shimabukuro, & L. S. Oshiro. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* 225:840-2, 1984. OTE: (Medline: 84300260).
- [Li et al.(1992)] Y. Li, H. Hui, C. J. Burgess, R. W. Price, P. M. Sharp, B. H. Hahn, & G. M. Shaw. Complete nucleotide sequence, genome organization, and biological properties of human immunodeficiency virus type 1 in vivo: evidence for limited defectiveness and complementation. *J Virol* 66:6587-600, 1992. OTE: (Medline: 93021387), (Genbank: M93258 M93259).
- [Li et al.(1991)] Y. Li, J. C. Kappes, J. A. Conway, R. W. Price, G. M. Shaw, & B. H. Hahn. Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication-competent and -defective viral genomes. *J Virol* 65:3973-85, 1991. OTE: (Medline: 91303644).
- [Liu et al.(1997)] S. L. Liu, T. Schacker, L. Musey, D. Shriner, M. J. McElrath, L. Corey, & J. I. Mullins. Divergent patterns of progression to AIDS after infection from the same source: human immunodeficiency virus type 1 evolution and antiviral responses. LA - Eng. *J Virol* 71:4284-95, 1997. OTE: (Medline: 97296232), (Genbank: U56146 U56147 U56148 U56149 U56150 U56151 U56152 U56153 U56154 U56155 U56156 U56157 U56158 U56159 U56160 U56161 U56162 U56163 U56164 U56165 U56166 U56167 U56168 U56169 U56170 U56171 U56172 U56173 U56174 U56175).
- [Liu et al.(1990)] Z. Q. Liu, C. Wood, J. A. Levy, & C. Cheng-Mayer. The viral envelope gene is involved in macrophage tropism of a human immunodeficiency virus type 1 strain isolated from brain tissue. *J Virol* 64:6148-53, 1990. OTE: (Medline: 91056585), (Genbank: M38673 M38673).
- [Lockey et al.(1996)] T. D. Lockey, K. S. Slobod, S. D. Rencher, R. V. Srinivas, & J. L. Hurwitz. Fluctuating diversity in the HTLV-IIIb virus stock: implications for neutralization and challenge experiments. LA - Eng. *AIDS Res Hum Retroviruses* 12:1297-9, 1996. OTE: (Medline: 97024624), (Genbank: U54646 U54647 U54648 U54649 U54650 U54651 U54652 U54653 U54654 U54655 U54656 U54657 U54658 U54659 U54660 U54661 U54662 U54663 U54664 U54665 U54666 U54667 U54668 U54669 U54670 U54671 U54672 U54673 U54674 U54675).
- [Louisirirotchanakul et al.(1997)] S. Louisirirotchanakul, A. Bobkov, N. Shaffer, T. D. Mastro, C. Wasi, J. Weber, & R. Cheingsong-Popov. Sequence analysis of an HIV type 1 env subtype E isolate from Thailand with discordant V3-loop serotyping. LA - Eng. *AIDS Res Hum Retroviruses* 13:807-9, 1997. OTE: (Medline: 97315052), (Genbank: U96782 U96783).
- [Loussert-Ajaka et al.(1995)] I. Loussert-Ajaka, M. L. Chaix, B. Korber, F. Letourneur, E. Gomas, E. Allen, T. D. Ly, F. Brun-Vezinet, F. Simon, & S. Saragosti. Variability of human immunodeficiency virus type 1 group O strains isolated from Cameroonian patients living in France. *J Virol* 69:5640-5649, 1995. OTE: (Medline: 95363977), (Genbank: U24562 U24563 U24564 U24565 U24566 U24567 U24568 U24706 U24707 U24708 U24709 U24710 U24711 U24712).

- [Louwagie et al.(1994)] J. Louwagie, E. L. Delwart, J. I. Mollins, F. E. McCutchan, G. Eddy, & D. S. Burke. Genetic analysis of HIV-1 isolates from Brazil reveals presence of two distinct genetic subtypes. *AIDS Res Hum Retroviruses* 10:361-367, 1994. OTE: (Medline: 95000928), (Genbank: L11751 L11752 L11753 L11754 L22082 L22083 L22084 L22085 L22086 L22087 L22088).
- [Louwagie et al.(1995)] J. Louwagie, W. Janssens, J. Mascola, L. Heyndrickx, P. Hegerich, der van Groen, F. E. McCutchan, & D. S. Burke. Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. *J Virol* 69:263-271, 1995. OTE: (Medline: 95074874), (Genbank: L22939 L22957 L23064 L23065).
- [Louwagie et al.(1993)] J. Louwagie, F. E. McCutchan, M. Peeters, T. P. Brennan, E. Sanders-Buell, G. A. Eddy, G. van der Groen, K. Fransen, G. M. Gershy-Damet, & R. D. et al. Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* 7:769-80, 1993. OTE: (Medline: 93371703), (Genbank: L03696 L03702 L03705 L03707 L11751 L11752 L11753 L11754 L11755 L11756 L11757 L11758 L11759 L11760 L11761 L11762 L11763 L11764 L11765 L11766 L11767 L11768 L11769 L11770 L11771 L11772 L11773 L11774 L11775 L11776).
- [Ludlam et al.(1985)] C. A. Ludlam, J. Tucker, C. M. Steel, R. S. Tedder, R. Cheingsong-Popov, R. A. Weiss, D. B. McClelland, I. Philp, & R. J. Prescott. Human T-lymphotropic virus type III (HTLV-III) infection in seronegative haemophiliacs after transfusion of factor VIII. *Eng. Lancet* 2:233-6, 1985. OTE: (Medline: 85266713).
- [Lukashov et al.(1996)] V. Lukashov, C. Kuiken, K. Boer, & J. Goudsmit. Hiv type 1 subtypes in the netherlands circulating among women originating from aids-endemic regions. *AIDS Res. Hum. Retroviruses* 12:951-953, 1996.
- [Lukashov et al.(1995)] V. V. Lukashov, M. T. Cornelissen, J. Goudsmit, M. N. Papuashvili, P. G. Rytik, R. M. Khaitov, E. V. Karamov, & W. de F. Simultaneous introduction of distinct HIV-1 subtypes into different risk groups in Russia, Byelorussia and Lithuania. *AIDS* 9:435-439, 1995. OTE: (Medline: 95367213).
- [Lukashov & Goudsmit(1995)] V. V. Lukashov & J. Goudsmit. Increasing genotypic and phenotypic selection from the original genomic RNA populations of HIV-1 strains LAI and MN (NM) by peripheral blood mononuclear cell culture, B-cell-line propagation and T-cell-line adaptation. *AIDS* 9:1307-1311, 1995. OTE: (Medline: 96188304).
- [Marquina et al.(1996)] S. Marquina, T. Leitner, R. D. Rabinovich, J. Benetucci, O. Libonatti, & J. Albert. Co-existence of subtypes B, F and a B/F env recombinant of HIV-1 in Buenos Aires, Argentina. *AIDS Res. Hum. Retroviruses* 12:1651-1654, 1996.
- [Mascola et al.(1994)] J. R. Mascola, J. Louwagie, F. E. McCutchan, C. L. Fischer, P. A. Hegerich, K. E. Wagner, A. K. Fowler, J. G. McNeil, & D. S. Burke. Two antigenically distinct subtypes of human immunodeficiency virus type 1: viral genotype predicts neutralization serotype. *J Infect Dis* 169:48-54, 1994. OTE: (Medline: 94103665), (Genbank: L14570 L14571 L14572 L14573 L14574 L14575 L14576).
- [McCutchan et al.(1996)] F. E. McCutchan, A. W. Artenstein, E. Sanders-Buell, M. O. Salminen, J. K. Carr, J. R. Mascola, X. F. Yu, K. E. Nelson, C. Khamboonruang, D. Schmitt, M. P. Kienny, J. G. McNeil, & D. S. Burke. Diversity of the envelope glycoprotein among human immunodeficiency virus type 1 isolates of clade E from Asia and Africa. *J. Virol.* 70:3331-3338, 1996. OTE: (Medline: 96211461), (Genbank: U48264 U48265 U48266 U48267 U48268 U48269 U48272 U48723 U48724 U48725 U48726 U48727 U48728).
- [McCutchan et al.(1992)] F. E. McCutchan, P. A. Hegerich, T. P. Brennan, P. Phanuphak, P. Singharaj, A. Jugsuddee, P. W. Berman, A. M. Gray, A. K. Fowler, & D. S. Burke. Genetic variants of HIV-1 in Thailand. *AIDS Res Hum Retroviruses* 8:1887-95, 1992. OTE: (Medline: 93143998).

Sequence Descriptions

- [McDonald et al.(1997)] R. A. McDonald, D. L. Mayers, R. C. Chung, K. F. Wagner, S. Ratto-Kim, D. L. Birx, & N. L. Michael. Evolution of human immunodeficiency virus type 1 env sequence variation in patients with diverse rates of disease progression and T-cell function. *LA - Eng. J Virol* 71:1871-9, 1997. OTE: (Medline: 97184515), (Genbank: U69282 U69283 U69284 U69285 U69286 U69287 U69288 U69289 U69290 U69291 U69292 U69293 U69294 U69295 U69296 U69297 U69298 U69299 U69300 U69301 U69302 U69303 U69304 U69305 U69306 U69307 U69308 U69309 U69310 U69311).
- [McKeating et al.(1996)] J. A. McKeating, Y. J. Zhang, C. Arnold, R. Frederiksson, E. M. Fenyo, & P. Balfe. Chimeric viruses expressing primary envelope glycoproteins of human immunodeficiency virus type I show increased sensitivity to neutralization by human sera. *Virology* 220:450-460, 1996. OTE: (Medline: 96251941), (Genbank: U57788 U57789 U57790 U57791 U57792 U57793 U57794).
- [McNearney et al.(1993)] T. McNearney, Z. Hornickova, B. Kloster, A. Birdwell, G. A. Storch, S. H. Polmar, M. Arens, & L. Ratner. Evolution of sequence divergence among human immunodeficiency virus type 1 isolates derived from a blood donor and a recipient. *Pediatr Res* 33:36-42, 1993. OTE: (Medline: 93165379).
- [McNearney et al.(1992)] T. McNearney, Z. Hornickova, R. Markham, A. Birdwell, M. Arens, A. Saah, & L. Ratner. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. *Proc Natl Acad Sci U S A* 89:10247-51, 1992. OTE: (Medline: 93066216).
- [McNearney et al.(1990)] T. McNearney, P. Westervelt, B. J. Thielan, D. B. Trowbridge, J. Garcia, R. Whittier, & L. Ratner. Limited sequence heterogeneity among biologically distinct human immunodeficiency virus type 1 isolates from individuals involved in a clustered infectious outbreak. *Proc Natl Acad Sci U S A* 87:1917-21, 1990. OTE: (Medline: 90175407), (Genbank: M31451).
- [Menu et al.(1996)] E. Menu, T.X., Truong, M. Lafon, T. Nguyen, M. Muller-Trutwin, T. Nguyen, A. Deslandres, G. Chaouat, Q. Duong, B. Ha, H. Fleury, & F. Barre-Sinoussi. Hiv type 1 thai subtype e is predominant in south vietnam. *AIDS Res. Hum. Retroviruses* 12:629-633, 1996.
- [Michael et al.(1997)] N. L. Michael, A. E. Brown, R. F. Voigt, S. S. Frankel, J. R. Mascola, K. S. Brothers, M. Louder, D. L. Birx, & S. A. Cassol. Rapid disease progression without seroconversion following primary human immunodeficiency virus type 1 infection-evidence for highly susceptible human hosts. *LA - Eng. J Infect Dis* 175:1352-9, 1997. OTE: (Medline: 97323978), (Genbank: U52888 U52889 U52890 U52891 U52892 U52893 U52894 U52895 U52896 U52897 U52898 U52899 U52900 U52901 U52902 U52903 U52904 U52905 U52906 U53103 U53104 U53105 U53106 U53107 U53108 U53109 U53110 U53111 U53112 U53113).
- [Michael et al.(1995)] N. L. Michael, G. Chang, L. A. d'Arcy, P. K. Ehrenberg, R. Mariani, M. P. Busch, D. L. Birx, & D. H. Schwartz. Defective accessory genes in a human immunodeficiency virus type I- infected long-term survivor lacking recoverable virus. *LA - Eng. J Virol* 69:4228-36, 1995. OTE: (Medline: 95287475), (Genbank: U24443 U24444 U24445 U24446 U24447 U24448 U24449 U24450 U24451 U24452 U24453 U24454 U24455 U24456 U24457 U24458 U24459 U24460 U24461 U24462 U24463 U24464 U24465 U24466 U24467 U24468 U24469 U24470 U24471 U24472).
- [Monken et al.(1995)] C. E. Monken, B. Wu, & A. Srinivasan. High resolution analysis of HIV-1 quasispecies in the brain. *AIDS* 9:345-349, 1995. OTE: (Medline: 95314789), (Genbank: L17088 L17089 L17090 L17091 L17092 L17093 L17094 L17095 L17096 L17097 L17098 L17099 L17100 L17101 L17102 L17103 L17104 L17105 L17106 L17107 L17108 L17109 L17110 L17111 L17112 L17113 L17114 L17115 L17116 L17117).
- [Montpetit(1995)] M. Montpetit. HIV-1 subtype A in Canada. *AIDS Res. Hum. Retroviruses* 11:1421-1422, 1995. OTE: (Medline: 96159142).
- [Morgado et al.(1994)] M. G. Morgado, E. C. Sabino, E. G. Shpaer, V. Bongertz, L. Brigido, M. D. Guimaraes, E. A. Castilho, B. Galvao-Castro, J. I. Mullins, & R. M. H. et al. V3 region polymorphisms in HIV-1 from Brazil: prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F. *AIDS Res Hum Retroviruses* 10:569-576, 1994. OTE: (Medline: 95000929), (Genbank: L19237 L19571 L19575 L19576 L19579).

- [Mulder-Kampinga et al.(1993)] G. A. Mulder-Kampinga, C. Kuiken, J. Dekker, H. J. Scherpbier, K. Boer, & J. Goudsmit. Genomic human immunodeficiency virus type 1 RNA variation in mother and child following intra-uterine virus transmission. *J Gen Virol* 74:1747-56, 1993. OTE: (Medline: 93389428), (Genbank: L21028 L21153).
- [Mulder-Kampinga et al.(1995)] G. A. Mulder-Kampinga, A. Simonon, C. L. Kuiken, J. Dekker, H. J. Scherpbier, de van Perre, K. Boer, & J. Goudsmit. Similarity in env and gag genes between genomic RNAs of human immunodeficiency virus type 1 (HIV-1) from mother and infant is unrelated to time of HIV-1 RNA positivity in the child. *J Virol* 69:2285-2296, 1995. OTE: (Medline: 95191002), (Genbank: L21028 L21029 L21030 L21031 L21032 L21033 L21034 L21035 L21036 L21037 L21038 L21039 L21040 L21041 L21042 L21043 L21044 L21045 L21046 L21047 L21048 L21049 L21050 L21051 L21052 L21053 Z47817 Z47818 Z47819 Z47820).
- [Murphy et al.(1993)] E. Murphy, B. Korber, M. C. Georges-Courbot, B. You, A. Pinter, D. Cook, M. P. Kierny, A. Georges, C. Mathiot, & F. B. et al. Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the central African Republic. *AIDS Res Hum Retroviruses* 9:997-1006, 1993. OTE: (Medline: 94107601), (Genbank: L11457 L11458 L11459 L11460 L11461 L11462 L11463 L11464 L11465 L11466 L11467 L11468 L11469 L11470 L11471 L11472 L11473 L11474 L11475 L11476 L11477 L11478 L11479 L11480 L11481 L11482 L11483 L11484 L11485 L11486).
- [Nerurkar et al.(1996)] V. R. Nerurkar, H. T. Nguyen, W. M. Dashwood, P. R. Hoffmann, C. Yin, D. M. Morens, A. H. Kaplan, R. Detels, & R. Yanagihara. HIV type 1 subtype B in commercial sex workers and injection drug users in Southern Vietnam. *AIDS Res. Hum. Retroviruses* 12:841-843, 1996.
- [Nkengasong et al.(1994)] J. N. Nkengasong, W. Janssens, L. Heyndrickx, K. Franssen, P. M. Ndumbe, J. Motte, A. Leonaers, M. Ngolle, J. Ayuk, & P. P. et al. Genotypic subtypes of HIV-1 in Cameroon. *AIDS* 8:1405-1412, 1994. OTE: (Medline: 95118531).
- [O'Brien et al.(1990)] W. A. O'Brien, Y. Koyanagi, A. Namazie, J. Q. Zhao, A. Diagne, K. Idler, J. A. Zack, & L. S. Chen. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. *Nature* 348:69-73, 1990. OTE: (Medline: 91043044).
- [Olivares et al.(1997)] I. Olivares, G. Shaw, & C. Lopez-Galindez. Phenotypic switch in a Spanish HIV type 1 isolate on serial passage on MT-4 cells (In Process Citation) LA - Eng. *AIDS Res Hum Retroviruses* 13:979-84, 1997. OTE: (Medline: 97366584).
- [Oram et al.(1990)] J. D. Oram, R. G. Downing, M. Roff, J. C. Clegg, D. Serwadda, & J. W. Carswell. Nucleotide sequence of a Ugandan HIV-1 provirus reveals genetic diversity from other HIV-1 isolates. *AIDS Res Hum Retroviruses* 6:1073-8, 1990. OTE: (Medline: 91090981), (Genbank: M62320).
- [Oram et al.(1991)] J. D. Oram, R. G. Downing, M. Roff, N. Serwankambo, J. C. Clegg, A. S. Featherstone, & J. C. Booth. Sequence analysis of the V3 loop regions of the env genes of Ugandan human immunodeficiency proviruses. *AIDS Res Hum Retroviruses* 7:605-14, 1991. OTE: (Medline: 92118457).
- [Orloff et al.(1993)] G. M. Orloff, M. L. Kalish, J. Chipangwi, K. E. Potts, C. Y. Ou, G. Schochetman, G. Dallabetta, A. L. Saah, & P. G. Miotti. V3 loops of HIV-1 specimens from pregnant women in Malawi uniformly lack a potential N-linked glycosylation site. *AIDS Res Hum Retroviruses* 9:705-6, 1993. OTE: (Medline: 93378795), (Genbank: L15721 L15722 L15723 L15724 L15725 L15726 L15727 L15728 L15729 L15730 L15731 L15732 L15733 L15734 L15735).
- [Osmanov et al.(1994)] S. Osmanov, W. L. Heyward, & J. Esparza. The World Health Organization Network for HIV Isolation and Characterization: summary of a pilot study. *AIDS Res Hum Retroviruses* 10:1325-1326, 1994. OTE: (Medline: 95194690).
- [Ou et al.(1992a)] C. Y. Ou, C. A. Ciesielski, G. Myers, C. I. Banda, C. C. Luo, B. T. Korber, J. I. Mullins, G. Schochetman, R. L. Berkelman, & A. N. E. et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* 256:1165-71, 1992a. OTE: (Medline: 92271245), (Genbank: M90847 M90848 M90849 M90850 M90851 M90852 M90853 M90854 M90855 M90856 M90857 M90858).

Sequence Descriptions

M90859 M90860 M90861 M90862 M90863 M90864 M90865 M90866 M90867 M90868 M90869 M90870 M90871 M90872 M90873 M90874 M90875 M90876).

- [Ou et al.(1992b)] C. Y. Ou, Y. Takebe, C. C. Luo, M. Kalish, W. Auwanit, C. Bandea, la de Torre, J. L. Moore, G. Schochetman, & S. Y. et al. Wide distribution of two subtypes of HIV-1 in Thailand. *AIDS Res Hum Retroviruses* 8:1471-2, 1992b. OTE: (Medline: 93103845).
- [Ou et al.(1993)] C. Y. Ou, Y. Takebe, B. G. Weniger, C. C. Luo, M. L. Kalish, W. Auwanit, S. Yamazaki, H. D. Gayle, N. L. Young, & G. Schochetman. Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. *Lancet* 341:1171-4, 1993. OTE: (Medline: 93254133).
- [Palca(1992a)] J. Palca. AIDS. CDC closes the case of the Florida dentist. *Science* 256:1130-1, 1992a. OTE: (Medline: 92271233).
- [Palca(1992b)] J. Palca. The case of the Florida dentist. *Science* 255:392-4, 1992b. OTE: (Medline: 92132539).
- [Pang et al.(1990)] S. Pang, Y. Koyanagi, S. Miles, C. Wifey, H. V. Vinters, & I. S. Chen. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* 343:85-9, 1990. OTE: (Medline: 90114416).
- [Pang et al.(1991)] S. Pang, H. V. Vinters, T. Akashi, W. A. O'Brien, & I. S. Chen. HIV-1 env sequence variation in brain tissue of patients with AIDS-related neurologic disease. *J Acquir Immune Defic Syndr* 4:1082-92, 1991. OTE: (Medline: 92092169).
- [Penny et al.(1996)] M. A. Penny, S. J. Thomas, N. W. Douglas, S. Ranjbar, H. Holmes, & R. S. Daniels. Env-gene sequences of primary HIV-1 isolates of subtypes B, C, D, E and F obtained from the WHO Network for HIV Isolation and Characterisation. *AIDS Res Hum Retroviruses* 12:741-747, 1996.
- [Pestano et al.(1993)] G. Pestano, A. Prince, J. Guyden, J. M. Ntambi, A. Atkin, & W. M. Boto. Independent divergences in the CD4 binding site and V3 loop encoded in two seroprevalent Ugandan HIV-1 clinical isolates. *J Acquir Immune Defic Syndr* 6:872-80, 1993. OTE: (Medline: 93301830), (Genbank: M98503 M98504).
- [Pestano et al.(1995)] G. A. Pestano, K. S. Hosford, A. I. Spira, J. Riley, J. M. Xie, N. Sewankambo, L. Brown, D. D. Ho, & W. M. Boto. Seroreactivity of analogous antigenic epitopes in glycoprotein 120 expressed in HIV-1 subtypes A, B, C, and D. *AIDS Res Hum Retroviruses* 11:589-596, 1995. OTE: (Medline: 96093894).
- [Pincus et al.(1994)] S. H. Pincus, K. G. Messer, P. L. Nara, W. A. Blattner, G. Colclough, & M. Reitz. Temporal analysis of the antibody response to HIV envelope protein in HIV-infected laboratory workers. *J Clin Invest* 93:2505-13, 1994. OTE: (Medline: 94259810).
- [Popovic et al.(1984)] M. Popovic, N. Flomenberg, D. J. Volkman, D. Mann, A. S. Fauci, B. Dupont, & R. C. Gallo. Alteration of T-cell functions by infection with HTLV-I or HTLV-II. *Science* 226:459-62, 1984. OTE: (Medline: 85040344).
- [Porter et al.(1997)] K. R. Porter, J. R. Mascola, H. Hupudic, D. Ewing, T. C. VanCott, R. L. Anthony, A. L. Corwin, S. Widodo, S. Ertono, F. E. McCutchan, D. S. Burks, C. G. Hayes, F. S. Wignall, & R. R. Graham. Genetic, antigenic and serologic characterization of human immunodeficiency virus type 1 from Indonesia. LA - Eng. *J Acquir Immune Defic Syndr Hum Retrovirol* 14:1-6, 1997. OTE: (Medline: 97143192).
- [Potts et al.(1993a)] K. E. Potts, M. L. Kalish, C. I. Bandea, G. M. Orloff, L. S. M., C. Brown, N. Malanda, M. Kavuka, G. Schochetman, & C. Y. O. et al. Genetic diversity of human immunodeficiency virus type 1 strains in Kinshasa, Zaire. *AIDS Res Hum Retroviruses* 9:613-8, 1993a. OTE: (Medline: 93378782).
- [Potts et al.(1993b)] K. E. Potts, M. L. Kalish, T. Lott, G. Orloff, C. C. Luo, M. A. Bernard, C. B. Alves, R. Badaro, J. Suleiman, & O. F. et al. Genetic heterogeneity of the V3 region of the HIV-1 envelope

- glycoprotein in Brazil. Brazilian Collaborative AIDS Research Group. *AIDS* 7:1191-7, 1993b. OTE: (Medline: 94030749).
- [Quinones-Mateu et al.(1995)] M. E. Quinones-Mateu, J. Dopazo, J. A. Este, T. R. Rota, & E. Domingo. Molecular characterization of human immunodeficiency virus type 1 isolates from Venezuela. *AIDS Res Hum Retroviruses* 11:605-616, 1995. OTE: (Medline: 96093896), (Genbank: U16764 U16779).
- [Quinones-Mateu et al.(1996)] M. E. Quinones-Mateu, A. Holguin, V. Soriano, & E. Domingo. Env gene diversity of HIV type 1 isolates from Spain. *AIDS Res. Hum. Retroviruses* 12:955-957, 1996.
- [R. et al.(1990)] L. D. R., B. Vanderborcht, M. Vanden Haesevelde, L. Heyndrickx, A. van Geel, C. Wauters, R. Bernaerts, E. Saman, P. Nijs, & B. W. et al. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J Virol* 64:1207-16, 1990. OTE: (Medline: 90156518), (Genbank: M31171).
- [Ranjbar(1995)] S. Ranjbar. Molecular characterization of an HIV type 1 isolate from Burundi. *AIDS Res. Hum. Retroviruses* 11:1146, 1995. OTE: (Medline: 96089225).
- [Ranjbar et al.(1995)] S. Ranjbar, A. Slade, A. Jenkins, A. Heath, P. Kitchin, N. Almond, S. Osmanov, & H. Holmes. Molecular characterization of an HIV type 1 isolate from Burundi. *AIDS Res Hum Retroviruses* 11:981-984, 1995. OTE: (Medline: 96020099), (Genbank: L35452 L35453 L35454 L35455 L35456 L35457 L35458 L35459).
- [Richalet-Secordel et al.(1994)] P. M. Richalet-Secordel, A. Deslandres, S. Plauc, B. You, F. Barre-Sinoussi, & R. V. M. H. Cross-reactive potential of rabbit antibodies raised against a cyclic peptide representing a chimeric V3 loop of HIV-1 gp120 studied by biosensor technique and ELISA. *FEMS Immunol Med Microbiol* 9:77-87, 1994. OTE: (Medline: 95004283).
- [Robbins et al.(1996)] K. E. Robbins, C. I. Banda, A. Levin, J. J. Goedert, W. A. Blattner, G. Brubaker, T. M. Brown, G. Schochetman, M. L. Kalish, J. Shao, & T. R. O'Brien. Genetic variability of human immunodeficiency virus type 1 in rural northwest Tanzania. *AIDS Res. Hum. Retroviruses* 12:1389-1391, 1996. OTE: (Medline: 97046198), (GenBank: U61875 U61879 U61880 U61881).
- [Roth et al.(1996)] W. W. Roth, J. A. Zuberi, H. G. Stringer, S. K. Davidson, & V. C. Bond. Examination of HIV type 1 variants in mother-child pairs. *AIDS Res. Hum. Retroviruses* 12:925-930, 1996.
- [Sabino et al.(1994a)] E. Sabino, L. Z. Pan, C. Cheng-Mayer, & A. Mayer. Comparison of in vivo plasma and peripheral blood mononuclear cell HIV-1 quasi-species to short-term tissue culture isolates: an analysis of tat and C2-V3 env regions. *AIDS* 8:901-9, 1994a. OTE: (Medline: 95032920), (Genbank: U01357 U01358 U01359 U01360 U01361 U01362 U01363 U01364 U01365 U01366 U01367 U01368 U01369 U01370 U01371 U01372 U01373 U01374 U01375 U01379 U01380 U01381 U01382 U01383 U01384 U01385 U01386 U01387 U01388 U01389).
- [Sabino et al.(1994b)] E. C. Sabino, E. Delwart, T. H. Lee, A. Mayer, J. I. Mullins, & M. P. Busch. Identification of low-level contamination of blood as basis for detection of human immunodeficiency virus (HIV) DNA in anti-HIV-negative specimens. *J Acquir Immune Defic Syndr* 7:853-859, 1994b. OTE: (Medline: 94293167), (Genbank: L20371 L20372 L20373 L20374 L20375 L20376 L20377 L20378 L20379 L20380).
- [Sabino et al.(1996)] E. C. Sabino, R. S. Diaz, L. F. Brígido, G. H. Learn, J. I. Mullins, A. L. Reingold, A. J. S. Duarte, A. Mayer, & M. P. Busch. Distribution of HIV-1 subtypes seen in an AIDS clinic in Sao Paulo City, Brazil. *AIDS* 10:1579-1584, 1996.
- [Sabino et al.(1994c)] E. C. Sabino, E. G. Shpaer, M. G. Morgado, B. T. Korber, R. S. Diaz, V. Bongertz, S. Cavalcante, B. Galvao-Castro, J. I. Mullins, & A. Mayer. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol* 68:6340-6346, 1994c. OTE: (Medline: 94365938), (Genbank: U08955 U08956 U08957 U08958 U08959 U08960 U08962 U08963 U08964 U08965 U08966 U08967 U08968

Sequence Descriptions

- U08969 U08970 U08971 U08972 U08973 U08974 U08975 U10019 U10020 U10021 U10022 U10023 U10024 U10025 U10026).
- [Saksena et al.(1997)] N. K. Saksena, Y. C. Ge, B. Wang, S. H. Xiang, J. Ziegler, P. Palasanthiran, W. Bolton, & A. L. Cunningham. RNA and DNA sequence analysis of the *nef* gene of HIV type 1 strains from the first HIV type 1-infected long-term nonprogressing mother-child pair. LA - Eng. *AIDS Res Hum Retroviruses* 13:729-32, 1997. OTE: (Medline: 97311526), (Genbank: U73339 U73340 U73341 U73342 U73343 U73344 U73345 U73346 U73347 U73348 U73349 U73350 U73351 U73352 U73353 U73354 U73355 U73356 U73357 U73358 U73359 U73360 U73361 U73362 U73363 U73364 U73365 U73366 U73367 U73368).
- [Sala et al.(1995)] M. Sala, E. Pelletier, & S. Wain-Hobson. HIV-1 *gp120* sequences from a doubly infected drug user. *AIDS Res Hum Retroviruses* 11:653-655, 1995. OTE: (Medline: 96093902), (Genbank: U20670 U20671 U20672 U20673 U20674 U20675 U20676 U20677).
- [Sala et al.(1994)] M. Sala, G. Zambruno, J. P. Vartanian, A. Marconi, U. Bertazzoni, & S. Wain-Hobson. Spatial discontinuities in human immunodeficiency virus type 1 quasispecies derived from epidermal Langerhans cells of a patient with AIDS and evidence for double infection. *J Virol* 68:5280-5283, 1994. OTE: (Medline: 94309197).
- [Salminen et al.(1997)] M. O. Salminen, J. K. Carr, D. L. Robertson, P. Hegerich, D. Gotte, C. Koch, E. Sanders-Buell, F. Gao, P. M. Sharp, B. H. Hahn, D. S. Burke, & F. E. McCutchan. Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. LA - Eng. *J Virol* 71:2647-55, 1997. OTE: (Medline: 97213928), (Genbank: K03455 M19921 U86768 U86781).
- [Salminen et al.(1996)] M. O. Salminen, B. Johansson, A. Sonnerborg, S. Aychunie, D. Gotte, P. Leinikki, D. S. Burke, & F. E. McCutchan. Full-length sequence of an Ethiopian human immunodeficiency virus type 1 (HIV-1) isolate of genetic subtype C. *AIDS Res. Hum. Retroviruses* 12:1329-1339, 1996.
- [Salminen et al.(1995)] M. O. Salminen, C. Koch, E. Sanders-Buell, P. K. Ehrenberg, N. L. Michael, J. K. Carr, D. S. Burke, & F. E. McCutchan. Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. *Virology* 213:80-86, 1995. OTE: (Medline: 96036482), (Genbank: M17449 M19921 U26546 U26942).
- [Sanchez-Palomino et al.(1993)] S. Sanchez-Palomino, J. M. Rojas, M. A. Martinez, E. M. Fenyo, R. Nagera, E. Domingo, & C. Lopez-Galindez. Dilute passage promotes expression of genetic and phenotypic variants of human immunodeficiency virus type 1 in cell culture. LA - Eng. *J Virol* 67:2938-43, 1993. OTE: (Medline: 93233267).
- [Sanchez-Pescador et al.(1985)] R. Sanchez-Pescador, M. D. Power, P. J. Barr, K. S. Steimer, M. M. Stempien, S. L. Brown-Shimer, W. W. Gee, A. Renard, A. Randolph, & J. A. L. et al. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science* 227:484-92, 1985. OTE: (Medline: 85090453), (Genbank: K02007).
- [Saueremann et al.(1990)] U. Saueremann, J. Schneider, J. Mous, U. Bruckhorst, I. Schedel, K. D. Jentsch, & G. Hunsmann. Molecular cloning and characterization of a German HIV-1 isolate. *AIDS Res Hum Retroviruses* 6:813-23, 1990. OTE: (Medline: 90303973).
- [Scarlatti et al.(1993)] G. Scarlatti, T. Leitner, E. Halapi, J. Wahlberg, P. Marchisio, M. A. Clerici-Schoeller, H. Wigzell, E. M. Fenyo, J. Albert, & M. U. et al. Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the RNA and DNA sequences of the virus populations of their mothers. *Proc Natl Acad Sci U S A* 90:1721-5, 1993. OTE: (Medline: 93189569), (Genbank: L08277 L08278 L08279 L08280 L08281 L08282 L08283 L08284 L08285 L08286 L08287 L08288 L08289 L08290 L08291 L08292 L08293 L08294 L08295 L08296 L08297 L08298 L08299 L08300 L08301 L08302 L08303 L08304 L08305 L08306).
- [Schwartz et al.(1997)] D. H. Schwartz, R. C. Castillo, S. Arango-Jaramillo, U. K. Sharma, H. F. Song, & G. Sridharan. Chemokine-independent in vitro resistance to human immunodeficiency virus (HIV-1)

Sequence Descriptions

- collected in 1976: nucleotide sequence comparison to recent isolates and generation of hybrid HIV. *AIDS Res Hum Retroviruses* 5:121-9, 1989. OTE: (Medline: 89228766).
- [Starcich et al.(1986)] B. R. Starcich, B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, & R. C. G. et al. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* 45:637-48, 1986. OTE: (Medline: 86218077), (Genbank: K03455 M12507 M12508 M17451).
- [Stevenson et al.(1990)] M. Stevenson, S. Haggerty, C. Lamonica, A. M. Mann, C. Meier, & A. Wasiak. Cloning and characterization of human immunodeficiency virus type 1 variants diminished in the ability to induce syncytium-independent cytolysis. *J Virol* 64:3792-803, 1990. OTE: (Medline: 90317877).
- [Strunnikova et al.(1995)] N. Strunnikova, S. C. Ray, R. A. Livingston, E. Rubalcaba, & R. P. Viscidi. Convergent evolution within the V3 loop domain of human immunodeficiency virus type 1 in association with disease progression. *J Virol* 69:7548-7558, 1995. OTE: (Medline: 96078998), (Genbank: U22682 U22683 U22684 U22685 U22686 U22687 U22688 U22689 U22690 U22691 U22692 U22693 U22694 U22695 U22696 U22697 U22698 U22699 U22700 U22701 U22702 U22703 U22704 U22705 U22706 U22707 U22708 U22709 U22710 U22711).
- [Su et al.(1997)] L. Su, H. Kaneshima, M. L. Bonyhadi, R. Lee, J. Auten, A. Wolf, B. Du, L. Rabin, B. H. Hahn, E. Terwilliger, & J. M. McCune. Identification of HIV-1 determinants for replication in vivo. *Virology* 227:45-52, 1997. OTE: (Medline: 97167516).
- [Takehisa et al.(1997a)] J. Takehisa, M. Osei-Kwasi, N. K. Ayisi, O. Hishida, T. Miura, T. Igarashi, J. Brandful, W. Ampofo, V. B. Netty, M. Mensah, M. Yamashita, E. Ido, & M. Hayami. Phylogenetic analysis of human immunodeficiency virus 1 in Ghana. LA - Eng. *Acta Virol* 41:51-4, 1997a. OTE: (Medline: 97343087), (Genbank: U67049 U67050 U75451 U75452 U75453 U75454 U75457 U75458 U75459 U75460 U75461).
- [Takehisa et al.(1997b)] J. Takehisa, L. Zekeng, T. Miura, E. Ido, M. Yamashita, I. Mboudjeka, L. G. Gurtler, M. Hayami, & L. Kapwa. Triple HIV-1 infection with group O and Group M of different clades in a single Cameroonian AIDS patient (letter) LA - Eng. *J Acquir Immune Defic Syndr Hum Retrovirol* 14:81-2, 1997b. OTE: (Medline: 97143206), (Genbank: U58148 U58149 U58150 U58151 U58152 U58153 U58154 U58155 U58156 U58157 U58158 U58159).
- [Takeuchi et al.(1991)] Y. Takeuchi, M. Akutsu, K. Murayama, N. Shimizu, & H. Hoshino. Host range mutant of human immunodeficiency virus type 1: modification of cell tropism by a single point mutation at the neutralization epitope in the env gene. *J Virol* 65:1710-8, 1991. OTE: (Medline: 91162716), (Genbank: M58037 M58038 M58039 M59192 M59193 M59888 M59889 M59890 M59891 M59892).
- [Tan et al.(1993)] W. Tan, R. Fredriksson, A. Bjorndal, P. Balfe, & E. M. Fenyo. Cotransfection of HIV-1 molecular clones with restricted cell tropism may yield progeny virus with altered phenotype. *AIDS Res Hum Retroviruses* 9:321-9, 1993. OTE: (Medline: 93290926).
- [Theodore et al.(1996)] T. S. Theodore, G. Englund, A. Buckler-White, C. E. Buckler, M. A. Martin, & K. W. Peden. Construction and characterization of a stable full-length macrophage-tropic HIV type 1 molecular clone that directs the production of high titers of progeny virions. LA - Eng. *AIDS Res Hum Retroviruses* 12:191-4, 1996. OTE: (Medline: 96432129).
- [Tripathy et al.(1996)] S. P. Tripathy, B. Renjifo, W. Wang, M. F. McLane, R. Bollinger, J. Rodrigues, J. Osterman, S. Tripathy, & M. Essex. Envelope glycoprotein 120 sequences of primary HIV type 1 isolates from Pune and New Delhi, India. *AIDS Res Hum Retroviruses* 12:1199-1202, 1996.
- [Trkola et al.(1996a)] A. Trkola, T. Dragic, J. Arthos, J. M. Binley, W. C. Olson, G. P. Allaway, C. Cheng-Mayer, J. Robinson, P. J. Maddon, & J. P. Moore. CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5 (see comments) LA - Eng. *Nature* 384:184-7, 1996a. OTE: (Medline: 97064177).
- [Trkola et al.(1996b)] A. Trkola, W. A. Paxton, S. P. Monard, J. A. Hoxie, M. A. Siani, D. A. Thompson, L. Wu, C. R. Mackay, R. Horuk, & J. P. Moore. Genetic subtype-independent inhibition of human

- immunodeficiency virus type 1 replication by CC and CXC chemokines. *J Virol* 72:396-404, 1996b. OTE: (Medline: 98080427).
- [Tsuchie et al.(1993)] H. Tsuchie, J. K. Maniar, N. Yoshihara, M. Inai, T. Kurimura, & T. Kitamura. Sequence analysis of V3 loop region of HIV-1 strains prevalent in India. LA - Eng. *Jpn J Med Sci Biol* 46:95-100, 1993. OTE: (Medline: 94096635), (Genbank: D13420 D13421 D13422 D13423 D13424 D13425 D13426 D13427).
- [van der Hoek et al.(1996)] L. van der Hoek, C. J. Sol, F. Snijders, J. F. Barteisman, R. Boom, & J. Goudsmit. Human immunodeficiency virus type 1 RNA populations in faeces with higher homology to intestinal populations than to blood populations. LA - Eng. *J Gen Virol* 77 (Pt 10):2415-25, 1996. OTE: (Medline: 97042272), (Genbank: Z76463 Z76464 Z76465 Z76466 Z76467 Z76468 Z76469 Z76470 Z76471 Z76472 Z76473 Z76474 Z76475 Z76476 Z76477 Z76478 Z76479 Z76480 Z76481 Z76482 Z76483 Z76484 Z76485 Z76486 Z76487 Z76488 Z76489 Z76490 Z76491 Z76492).
- [Vanden Haesevelde et al.(1994)] M. Vanden Haesevelde, J. L. Decourt, L. D. R. J., B. Vanderborcht, G. van der Groen, H. van Heurverswijn, & E. Saman. Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. *J Virol* 68:1586-96, 1994. OTE: (Medline: 94149849), (Genbank: L23119).
- [von A. et al.(1995)] B. von A., B. von B., J. Eberle, B. Biryahwaho, R. G. Downing, & L. Gurtler. New crown motif of an HIV-1 V3 loop sequence from a Ugandan AIDS patient. *AIDS Res Hum Retroviruses* 11:183-184, 1995. OTE: (Medline: 95251931), (Genbank: U15005 U15006 U15007).
- [Wain-Hobson et al.(1985)] S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole, & M. Alizon. Nucleotide sequence of the AIDS virus, LAV. *Cell* 40:9-17, 1985. OTE: (Medline: 85099333), (Genbank: K02013).
- [Wain-Hobson et al.(1991)] S. Wain-Hobson, J. P. Vartanian, M. Henry, N. Chenciner, R. Cheynier, S. Delassus, L. P. Martins, M. Sala, M. T. Nugayre, & D. G. et al. LAV revisited: origins of the early HIV-1 isolates from Institut Pasteur. *Science* 252:961-5, 1991. OTE: (Medline: 91240282), (Genbank: M64178 M64179 M64180 M64181 M64182 M64183 M64184 M64185 M64186 M64187 M64188 M64189 M64190 M64191 M64192 M64193 M64194 M64195 M64196 M64197 M64198 M64199 M64200 M64201 M64202 M64203 M64204 M64205 M64206 M64207).
- [Wang et al.(1997a)] B. Wang, Y. C. Ge, R. Jozwiak, W. Bolton, P. Palasanthiran, J. Ziegler, J. Chang, S. H. Xiang, A. L. Cunningham, & N. K. Saxena. Molecular analyses of human immunodeficiency virus type 1 V3 region quasispecies derived from plasma and peripheral blood mononuclear cells of the first long-term-nonprogressing mother and child pair. LA - Eng. *J Infect Dis* 175:1510-5, 1997a. OTE: (Medline: 97324000).
- [Wang et al.(1997b)] B. Wang, Y. C. Ge, P. Palasanthiran, J. Ziegler, W. Bolton, S. H. Xiang, D. E. Dwyer, A. L. Cunningham, & N. K. Saxena. HIV type 1 V3 loop sequences derived from peripheral blood of transmitting mothers, their infants, and nontransmitting mothers differ in their crown octapeptide motifs. LA - Eng. *AIDS Res Hum Retroviruses* 13:275-9, 1997b. OTE: (Medline: 97186421), (Genbank: U66625 U66626 U66627 U66628 U66629 U66630 U66631 U66632 U66633 U66634 U66635 U66636 U66637 U66638 U66639 U66640 U66641 U66642 U66643 U66644 U66645 U66646 U66647 U66648 U66649 U66650 U66651 U66652 U66653 U66654).
- [Wang et al.(1996a)] W. Wang, K. H. Mayer, M. Essex, & T. Lee. Sequential change of cysteine residues in hypervariable region 1 of gp120 in primary HIV-1 isolates of subtype B. *AIDS Res. Hum. Retroviruses* 12:1195-1197, 1996a.
- [Wang et al.(1996b)] W.-K. Wang, H. Brumblay, M. Essex, & T.-H. Lee. The ratio of non-synonymous nucleotide substitutions to synonymous nucleotide substitutions per site of divergent gp120 sequences: Intrasubject and intersubject comparisons. *Unpublished* 1996b.
- [Wang et al.(1995)] W. K. Wang, M. Essex, & T. H. Lee. Uncommon gp120 cysteine residues found in primary HIV-1 isolates. *AIDS Res Hum Retroviruses* 11:185-188, 1995. OTE: (Medline: 95251932),

Sequence Descriptions

- (Genbank: U16324 U16325 U16326 U16327 U16328 U16329 U16330 U16331 U16332 U16333 U16334 U16335).
- [Weniger et al.(1994)] B. G. Weniger, Y. Takebe, C. Y. Ou, & S. Yamazaki. The molecular epidemiology of HIV in Asia. *AIDS* 8:S13-28, 1994. OTE: (Medline: 95160969), (Genbank: L07442 L07443 L07444 L07445 L07446 L07447 L07448 L07449 L07450 L07451 L07452 L07453 L07454 L07455 L07456 L07457 L07458 L07459 L07460 L07461 L07462 L07463 L07464 L19238 L19239 L32084 L32085 L32086 L32087 L32088).
- [Westervelt et al.(1991)] P. Westervelt, H. E. Gendelman, & L. Ratner. Identification of a determinant within the human immunodeficiency virus 1 surface envelope glycoprotein critical for productive infection of primary monocytes. *Proc Natl Acad Sci U S A* 88:3097-101, 1991. OTE: (Medline: 91195299), (Genbank: M60157 M60158 M60159 M60160 M60161 M60162 M60163 M60164 M60165 M60472).
- [Willey et al.(1986)] R. L. Willey, R. A. Rutledge, S. Dias, T. Folks, T. Theodore, C. E. Buckler, & M. A. Martin. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. *Proc Natl Acad Sci U S A* 83:5038-42, 1986. OTE: (Medline: 86259728), (Genbank: K03346 K03347 K03455).
- [Wolfs et al.(1992)] T. F. Wolfs, G. Zwart, M. Bakker, & J. Goudsmit. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. *Virology* 189:103-10, 1992. OTE: (Medline: 92295544), (Genbank: M91819 M91820 M91821 M91822 M91823 M91824 M91825 M91826 M91819 M91820 M91821 M91822 M91823 M91824 M91825 M91826).
- [Wolfs et al.(1991)] T. F. Wolfs, G. Zwart, M. Bakker, M. Valk, C. L. Kuiken, & J. Goudsmit. Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution. *Virology* 185:195-205, 1991. OTE: (Medline: 92024075), (Genbank: M74591 M74592 M74593 M74594 M74595 M74596 M74597 M74598 M74599 M74600 M74601 M74602 M74603 M74604 M74605 M74606 M74607 M74608 M74609 M74610 M74611 M74612 M74613 M74614 M74615 M74616 M74617 M74618 M74619 M74620).
- [Wolynsky et al.(1996)] S. M. Wolynsky, B. T. M. Korber, A. U. Neumann, M. Daniels, K. J. Kunstman, A. J. Whetsell, Y. Cao, D. D. Ho, J. T. Saffrit, & R. A. Koup. Adaptive evolution of human immunodeficiency virus type 1 during the natural course of infection. *Science* 272:537-542, 1996.
- [Wrin et al.(1995)] T. Wrin, T. P. Loh, J. C. Vennari, H. Schuitemaker, & J. H. Nunberg. Adaptation to persistent growth in the H9 cell line renders a primary isolate of human immunodeficiency virus type 1 sensitive to neutralization by vaccine sera. *J Virol* 69:39-48, 1995. OTE: (Medline: 95074890), (Genbank: U15030 U15032).
- [Xin et al.(1995a)] K. Q. Xin, X. H. Ma, K. A. Crandall, H. Bukawa, Y. Ishigatsubo, S. Kawamoto, & K. Okuda. Dual infection with HIV-1 Thai subtype B and E (letter) LA - Eng. *Lancet* 346:1372-3, 1995a. OTE: (Medline: 96064971).
- [Xin et al.(1995b)] K. Q. Xin, P. Shapshak, S. Kawamoto, I. Nagano, C. B. McCoy, & K. Okuda. Highly divergent env sequences of HIV-1 B subtype with two novel V3 loop motifs detected in an AIDS patient in Miami, Florida. *AIDS Res. Hum. Retroviruses* 11:1139-41, 1995b. OTE: (Medline: 96089223).
- [York-Higgins et al.(1990)] D. York-Higgins, C. Cheng-Mayer, D. Bauer, J. A. Levy, & D. Dina. Human immunodeficiency virus type 1 cellular host range, replication, and cytopathicity are linked to the envelope region of the viral genome. *J Virol* 64:4016-20, 1990. OTE: (Medline: 90317906).
- [Youno et al.(1988)] J. Youno, S. F. Josephs, M. Reitz, D. Zagury, F. Wong-Staal, & R. C. Gallo. Nucleotide sequence analysis of the env gene of a new Zairian isolate of HIV-1. *AIDS Res Hum Retroviruses* 4:165-73, 1988. OTE: (Medline: 88281278), (Genbank: J03653).
- [Yu et al.(1995)] X. F. Yu, Z. Wang, C. Beyrer, D. D. Celentano, C. Khamboonruang, E. Allen, & K. Nelson. Phenotypic and genotypic characteristics of human immunodeficiency virus type 1 from patients with AIDS in northern Thailand. LA - Eng. *J Virol* 69:4649-55, 1995. OTE: (Medline:

- 95333239), (Genbank: M75007 U25550 U25551 U25552 U25553 U25554 U25555 U25556 U25557 U25558 U25559 U25560 U25561 U25562 U25563 U25564 U25565 U25566 U25567 U25568 U25569 U25570 U25571 U25572 U25573 U25574 U25575 U25576 U25577 U25578).
- [Zachar et al.(1996a)] V. Zachar, A. S. Goustin, V. Zacharova, H. Hager, U. Koppelhus, D. D. Wornble, X. Liu, C. Bamba, A. Nyongo, & P. Ebbesen. Genetic polymorphism of envelope V3 region of HIV type 1 subtypes A, C, and D from Nairobi, Kenya. *AIDS Res. Hum. Retroviruses* 12:75–78, 1996a.
- [Zachar et al.(1996b)] V. Zachar, V. Mayer, V. Zacharova, H. Schmidtmayerova, J. Kusanicka, M. Mokras, X. Liu, A. S. Goustin, & P. Ebbesen. Spread of HIV type 1 in Slovakia remains limited and is restricted to subtype B. *AIDS Res. Hum. Retroviruses* 12:1069–1071, 1996b.
- [Zhang et al.(1997)] L. Zhang, R. S. Diaz, D. D. Ho, J. W. Mosley, M. P. Busch, & A. Mayer. Host-specific driving force in human immunodeficiency virus type 1 evolution in vivo. LA - Eng. *J Virol* 71:2555–61, 1997. OTE: (Medline: 97184598), (Genbank: U29433 U29434 U29435 U29436 U29437 U29956 U29957 U29959 U29960 U29961 U29962 U29963 U29964 U29965 U29966 U29967 U29968 U29969 U29970 U29971 U29972 U29973 U29974 U29975 U29976 U29977 U29978 U29979 U29980 U29981).
- [Zhang et al.(1993)] L. Q. Zhang, P. MacKenzie, A. Cleland, E. C. Holmes, A. J. Brown, & P. Simmonds. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J Virol* 67:3345–56, 1993. OTE: (Medline: 93267786), (Genbank: L13488 L13489 L13490 L13491 L13492 L13493 L13494 L13495 L13496 L13497 L13498 L13499 L13500 L13501 L13502 L13503 L13504 L13505 L13506 L13507 L13508 L13509 L13510 L13511 L13512 L13513 L13514 L13515 L13516 L13517).
- [Zhu & Ho(1995)] T. Zhu & D. D. Ho. Was hiv present in 1959? *Nature* 374:503–504, 1995. OTE: (Medline: 95214770).
- [Zhu et al.(1993)] T. Zhu, H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup, & D. D. Ho. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science* 261:1179–81, 1993. OTE: (Medline: 93361976), (Genbank: L21224 L21225 L21226 L21227 L21228 L21229 L21230 L21231 L21232 L21233 L21234 L21235 L21236 L21237 L21238 L21239 L21240 L21241 L21242 L21243 L21244 L21245 L21246 L21247 L21248 L21249 L21250 L21251 L21252 L21253).
- [Zhu et al.(1996)] T. Zhu, N. Wang, A. Carr, D. S. Nam, R. Moor-Jankowski, D. A. Cooper, & D. D. Ho. Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. LA - Eng. *J Virol* 70:3098–107, 1996. OTE: (Medline: 96186740), (Genbank: U50780 U50781 U50782 U50783 U50784 U50785 U50786 U50787 U50788 U50789 U50790 U50791 U50792 U50793 U50794 U50795 U50796 U50797 U50798 U50799 U50800 U50801 U50802 U50803 U50804 U50805 U50806 U50807 U50808 U50809).
- [Zhu et al.(1995)] T. Zhu, N. Wang, A. Carr, S. Wolinsky, & D. D. Ho. Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype B in an acute seroconverter. *J Virol* 69:1324–1327, 1995. OTE: (Medline: 95115099), (Genbank: U16372 U16373 U16374 U16375 U16376 U16377 U16378 U16379 U16380 U16381 U16382 U16383 U16384 U16385 U16386 U16387 U16388).
- [Zwart et al.(1994a)] G. Zwart, N. K. Back, C. Ramautarsing, M. Valk, L. van der Hoek, & J. Goudsmit. Frequent and early HIV-1MN neutralizing capacity in sera from Dutch HIV-1 seroconverters is related to antibody reactivity to peptides from the gp120 V3 domain. *AIDS Res Hum Retroviruses* 10:245–251, 1994a. OTE: (Medline: 94289062).
- [Zwart et al.(1994b)] G. Zwart, L. van der Hoek, M. Valk, M. T. Cornelissen, E. Baan, J. Dekker, M. Koot, C. L. Kuiken, & J. Goudsmit. Antibody responses to HIV-1 envelope and gag epitopes in HIV-1 seroconverters with rapid versus slow disease progression. *Virology* 201:285–93, 1994b. OTE: (Medline: 94240812), (Genbank: U05786 U05787 U05788 U05789 U05790 U05791 U05792 U05793

Sequence Descriptions

U05794 U05795 U05796 U05797 U05798 U05799 U05800 U05801 U05802 U05803 U05804 U05805
U05806 U05807 U05808 Z29279 Z29287 Z29289 Z29292).

[Zwart et al.(1993)] G. Zwart, T. F. Wolfs, R. Bookelman, S. Hartman, M. Bakker, C. A. Boucher, C. Kuiken, & J. Goudsmit. Greater diversity of the HIV-1 V3 neutralization domain in Tanzania compared with The Netherlands: serological and genetic analysis. *AIDS* 7:467-74, 1993. OTE: (Medline: 93283054).

[Zwart et al.(1992)] G. Zwart, T. F. Wolfs, M. Valk, L. Van der Hoek, C. L. Kuiken, & J. Goudsmit. Characterization of the specificity of the human antibody response to the V3 neutralization domain of HIV-1. *AIDS Res Hum Retroviruses* 8:1897-908, 1992. OTE: (Medline: 93143999).

Mutations in Retroviral Genes Associated with Drug Resistance

Jennifer Hammond,¹ Brendan A. Larder,² Raymond E. Schinazi,³ and John W. Mellors¹

¹ University of Pittsburgh, 603 Parran Hall, Pittsburgh, PA 15261.

² Virco, UK, 162A Cambridge Science Park, Milton Road, Cambridge CB4 4GH.

³ Emory University/VAMC, 1670 Clairmont Road, Decatur, GA 30033.

Introduction

Drug resistance is the inevitable consequence of incomplete suppression of HIV replication. The rapid replication rate of HIV and its inherent genetic variation have led to the identification of many HIV variants that exhibit altered drug susceptibility. The growing number of drug resistance mutations listed in this revised table stands as a testimony to the genetic flexibility of HIV. This updated table lists 143 mutations occurring in the HIV Gag (3), Protease (44), Reverse Transcriptase (69), or Envelope (27) genes. Although the tables are quite comprehensive, the reader should be reminded that the mutations described are predominantly found in clade B virus and not in other HIV genotypes. The revised table includes for the first time drug resistance mutations that have been identified for SIV and FIV.

In the table the phrase "Enzyme resist." refers to inhibition assays done just with a mutated enzyme. Instead of introducing the mutations into a virus and testing the susceptibility of the mutant virus to a drug, researchers introduce the mutation(s) into the enzyme and determine their effect by running enzyme activity assays. These sort of assays don't take into account changes in other viral proteins (like gag) that would also help confer resistance, which is the reason for distinguishing enzyme resistance from whole virus resistance.

Acknowledgments

The authors would like to gratefully acknowledge their colleagues for assistance in assembling this table. This work was supported in part by the National Institutes of Health, the Department of Veterans Affairs, and the Georgia VA Research Center for AIDS and HIV Infections.

Contents

HIV1 RT	III-208
HIV1 Protease	III-221
HIV1 Envelope	III-233
SIV RT	III-235
FIV RT	III-235
Table of Abbreviations	III-236
Table of Compounds	III-236
References	III-239

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
M 41 L	ATG to TTG/CTG	Nucleoside RTI	AZT	?	Y	4		M41L/T215Y: 60-70-fold; M41L/D67N/K70R/T215Y: 180-fold.	(1, 2, 3)
A 62 V	GCC to GTC	Multiple Nucleoside Resistance		N	Y	Nil		A62V alone has no effect, but in combination with mutations at 75, 77, 116, 151 causes multi NRTI resistance.	(4, 5)
K 65 R	AAA to AOA	Nucleoside RTI	ddI	Y	Y	4-10	ddC; PMEA; 3TC(5)	Infrequently observed in patients receiving ddI or ddC	(6)
K 65 R	AAA to AOA	Nucleoside RTI	ddC	Y	Y	4-10			(6, 7)
K 65 R	AAA to AOA	Nucleoside RTI	I592U89	Y	N	3		K65R/L74V: 3.6-fold; K65R/M184V: 7-fold; K65R/L74V/M184V: 10.2-fold	(8)
K 65 R	AAA to AGA	Nucleoside RTI	DXG	Y	?	8	other dioxolane derivatives	Reverses AZT resistance in D67N/K70R/T215Y/K219Q background	(9)
K 65 R	AAA to AGA	Nucleoside RTI	PMEA	Y	N	10-25			(10, 11)
K 65 R	AAA to AGA	Nucleoside RTI	PMPA	Y	?	3.5			(12)
D 67 N	GAC to AAC	Nucleoside RTI	AZT	Y	Y			D67N/K70R/T215Y/K219Q: 120-fold; M41L/D67N/K70R/T215Y: 180-fold.	(1, 2, 3)
T 69 D	ACT to GAT	Nucleoside RTI	ddC	N	Y	5			(13)
K 70 E	AAA to GAA	Nucleoside RTI	PMEA	Y	Y	9	3TC (7); PFA: 2-fold hypersusceptibility		(14, 15)
K 70 R	AAA to AGA	Nucleoside RTI	AZT	Y	Y			D67N/K70R/T215Y/K219Q: 120-fold	(1, 2, 3)
L 74 I	TTA to ATA	HIV-1 Specific RTI	HBV 097	Y	?				(16)
L 74 V	TTA to GTA	Nucleoside RTI	ddI	N	Y	5-10	ddC(4)	Can reverse effect of T215Y AZT resistance mutation	(17)
L 74 V	TTA to GTA	Nucleoside RTI	I592U89	Y	N	4		K65R/L74V: 3.6-fold; K65R/L74V/M184V: 10.2-fold	(8)
L 74 V	TTA to GTA	Nucleoside RTI	DXG	Y	?	4			(9)
L 74 V	TTA to GTA	HIV-1 Specific RTI	HBV 097	Y	?				(16)
V 75 I	GTA to TTA	HIV-1 Specific RTI	HBV 097	Y	?			Compensates for negative effect of G190E mutation on RT activity	(16)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
V 75 I	GTA to ATA	Multiple Nucleoside Resistance		N	Y	Nil		V75I alone has no effect, but in combination with mutations at 62, 77, 116, 151 causes multi NRTI resistance.	(4, 5)
V 75 L	GTA to ATA	HIV-1 Specific RTI	HBV 097	Y	?				(16)
V 75 T	GTA to ACA	Nucleoside RTI	d4T	Y	Y	7	ddI; ddC; d4C; (-)-FTC	Observed with d4T selection in vitro, rarely in patients receiving d4T	(18, 19)
F 77 L	TTC to CTC	Multiple Nucleoside Resistance		N	Y	Nil		F77L alone has no effect, but in combination with mutations at 62, 75, 116, 151 causes multi NRTI resistance.	(4, 5)
W 88 G	TGG to GGG	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	Y	5	Hypersusceptibility to AZT	Observed after selection with AZT and PFA; suppresses effects of AZT mutations	(20, 21, 22)
W 88 S	TGG to TCG	Pyrophosphate Analogue RTI	Foscarnet (PFA)	N	Y	2-4	Wild-type susceptibility to AZT.	Partially suppresses effects	(20, 21, 22)
E 89 G	GAA to GGA	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	N	14		Isolated by screening RT clones for ddGTP resistance	(23)
E 89 K	GAA to GGA	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	N	> 16		Suppresses effects of AZT resistance mutations	(21, 22)
L 92 I	TTA to ATA	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	N	8		Partially suppresses effects of AZT resistance mutations	(21, 22)
A 98 G	GCA to GGA	HIV-1 Specific RTI	L-697,661	N	Y	8			(24)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
A 98 G	GCA to GGA	HIV-1 Specific RTI	Nevirapine	N	Y				(25)
L 100 I	TTA to ATA	HIV-1 Specific RTI	TIBO R82150	Y	?	> 100		suppresses effects of AZT resistance mutations	(26, 27, 28)
L 100 I	TTA to ATA	HIV-1 Specific RTI	TIBO R82913	Y	?			Found in combination with E138K	(29)
L 100 I	TTA to ATA	HIV-1 Specific RTI	L-697,661	Y	N	2			(24)
L 100 I	TTA to ATA	HIV-1 Specific RTI	Nevirapine	N	Y				(30)
L 100 I	TTA to ATA	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?	8-11		Combinations of mutations needed for high-level resistance; L100I/V108I: 1,000-fold; L100I/V179D/Y181C: 1,000-fold	(31, 32)
L 100 I	TTA to ATA	HIV-1 Specific RTI	BHAP U-88204E	Y	?				(33, 34)
L 100 I	TTA to ATA	HIV-1 Specific RTI	UC-68 (638532)	Y	?	70			(35)
L 100 I	TTA to ATA	HIV-1 Specific RTI	UC-70 (638534)	Y	?	758			(36)
L 100 I	TTA to ATA	HIV-1 Specific RTI	UC-781	Y	?	20		Activity of UC-781 versus L100I, K103N, V106A, E138K, Y181C and Y188L reduced by 2-, 7-, 1.5-, 1.5-, 5- and 150-fold, respectively, compared to wild type	(37, 38)
L 100 I	TTA to ATA	HIV-1 Specific RTI	UC-84 (615985)	Y	?	> 40, > 33			(36, 39)
K 101 E	AAA to GAA	HIV-1 Specific RTI	8-Chloro-TIBO R091767	?	Y				(40)
K 101 E	AAA to GAA	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	N	Y			K101E, Y188H, E233Y and K238T observed with U-87201E/AZT combination therapy	(41)
K 101 E	AAA to GAA	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?	1,000			(32)
K 101 E	AAA to GAA	HIV-1 Specific RTI	L-697,661	N	Y	8			(24)
K 101 E	AAA to GAA	HIV-1 Specific RTI	UC-10 (645129)	Y	?	12		K101E/Y181C: 200-fold	(36, 42, 43)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change			Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
K	101	E	AAA to GAA	HIV-1 Specific RTI	UC-38 (629243)	Y	N			K101E/G190E: > 100-fold	(36)
K	101	E	AAA to GAA	HIV-1 Specific RTI	UC-57 (647014)	Y	?			K101E/Y181C: 58-fold	(36)
K	101	E	AAA to GAA	HIV-1 Specific RTI	UC781	Y	?	7	UC040 (18); Nevirapine (15)	V108I/ Y181C: 55-fold; K101E/ V108I/ Y181C: 500-fold.	(37, 43)
K	101	I	AAA to ATA	HIV-1 Specific RTI	UC-16	Y	N	10		K101I/G141E: 10-fold	(36)
K	101	Q	AAA to CAA	HIV-1 Specific RTI	Trovirine	Y	?			Found in combination with V108I	(44, 45)
K	103	N	AAA to AAC	HIV-1 Specific RTI	8-Chloro-TIBO R091767	?	Y				(40)
K	103	N	AAA to AAC	HIV-1 Specific RTI	BHAP U-87201E (zalcitabine)	N	Y			K103N and Y181C observed with monotherapy	(41)
K	103	N	AAA to AAC	HIV-1 Specific RTI	BHAP U-90152 (delavirdine)	?	Y			K103N/Y181C seen separately and in combination in patients	(46)
K	103	N	AAA to AAC	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	Y	67		Predominant mutation in vivo	(31)
K	103	N	AAA to AAC	HIV-1 Specific RTI	L-697,593	Y	?	20		K103N/Y181C: > 1,000-fold	(47)
K	103	N	AAA to AAC	HIV-1 Specific RTI	L-697,661	Y	Y	8		K103N and Y181C most common with monotherapy	(24, 48)
K	103	N	AAA to AAC	HIV-1 Specific RTI	Lopinavir (R89439, α -APA)	Y	Y				(49)
K	103	N	AAA to AAC	HIV-1 Specific RTI	MKC442 (JEBU)	Y	?			predominant mutation in vivo	(50)
K	103	N	AAA to AAC	HIV-1 Specific RTI	Nevirapine	N	Y				(30)
K	103	N	AAA to AAC	HIV-1 Specific RTI	TIBO R82913	Y	?	> 100		K103N/Y181C: > 1,000-fold	(33)
K	103	N	AAA to AAC	HIV-1 Specific RTI	UC-10 (645129)	Y	N	5			(35)
K	103	N	AAA to AAC	HIV-1 Specific RTI	UC-81 (615727)	Y	?	40			

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
K 103 Q	AAA to CAA	HIV-1 Specific RTI	L-697,661	N	Y	8			(48)
K 103 R	AAA to AGA	HIV-1 Specific RTI	Trovirdine	Y	?		Nevirapine; 9-chloro- TIBO	K103R/V179D: 500-fold; Found in combination with V179D or Y181C	(44, 45)
K 103 R	AAA to AGA	HIV-1 Specific RTI	MKC442(1-BBU)	Y	Y				(51)
K 103 T	AAA to ACA	HIV-1 Specific RTI	BHAP U-90152 (delavirdine)	?	Y				(46)
K 103 T	AAA to ACA	HIV-1 Specific RTI	UC-42	Y	N	100			(36)
V 106 A	GTA to GCA	HIV-1 Specific RTI	BHAP U-88204B	Y	?				(34)
V 106 A	GTA to GCA	HIV-1 Specific RTI	E-BBU-dM	Y	?				(52)
V 106 A	GTA to GCA	HIV-1 Specific RTI	Nevirapine	Y	Y	~100		No effect on AZT resistance	(25, 29, 30, 33)
V 106 A	GTA to GCA	HIV-1 Specific RTI	TIBO R82913	Y	?	~100			(29)
V 106 A	GTA to GCA	HIV-1 Specific RTI	UC-69 (646989)	Y	?			V106A/V181C: 166-fold	(36)
V 106 A	GTA to GCA	HIV-1 Specific RTI	UC-82	Y	?	13		Activity of UC-82 versus L100I, K103N, V106A, E138K, Y181C and Y188L reduced by 2-, 6-, 1.5-, 2-, 4- and 200-fold, respectively, compared to wild type	(37, 38)
V 106 A	GTA to GCA	HIV-1 Specific RTI	S-2720 (Quinoxaline)	Y	?			P225H follows V106A. Also seen with L101I and Y181C. Double and triple mutants highly resistant to other NNRTI's, including MKC442	(53)
V 106 I	GTA to ATA	HIV-1 Specific RTI	HBV 097					Appears under lowered drug concentra- tion selection	(54)
V 108 I	GTA to ATA	HIV-1 Specific RTI	DMP 266 (L- 743,726)	Y	?			L100I/V108I: 1,000-fold	(31)
V 108 I	GTA to GCA	HIV-1 Specific RTI	L-697,661	Y	Y	4			(24)
V 108 I	GTA to ATA	HIV-1 Specific RTI	Loviride (R89439, α -APA)	Y	?				(49)
V 108 I	GTA to GCA	HIV-1 Specific RTI	MKC442 (1-BBU)	Y	?				(50)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
V 108 I	GTA to ATA	HIV-1 Specific RTI	Nevirapine	N	Y				(30)
V 108 I	GTT to GAT	HIV-1 Specific RTI	TIBO R82913	N	Y	> 100	R82150 (> 100)		(55)
V 108 I	GTA to ATA	HIV-1 Specific RTI	Trivinone	Y	?			Found in combination with K101Q	(44, 45)
V 108 I	GTA to ATA	HIV-1 Specific RTI	UC781	Y	?			V108I/Y181C: 55 fold, K101E/V108I/Y181C: 500 fold.	(43)
Y 115 F	TAT to TTT	Nucleoside RTI	1592U89	Y	N	2		K65R/L74V and/or Y115F with M184V: 10 fold; L74V/Y115F/M184V: 11-fold	(8)
F 116 Y	TTT to TAT	Multiple Nucleoside Resistance		N	Y	NH		F116Y alone has no effect, but in combination with mutations at 62, 75, 77, 151 causes multi NRTI resistance.	(4, 5)
P 119 S	CCC to TCC	Nucleoside RTI	F-d4A	Y	?	4		Found with V179D and/or L214F, which are possibly compensatory	(56)
E 138 K	GAG to AAG	HIV-1 Specific RTI	TSAO	Y	?	> 100		E138A (GAG to GCG) in TSAO-naïve patients confers TSAO viral resistance	(57, 58, 59)
E 138 K	GAG to AAG	HIV-1 Specific RTI	MKC442 (I-BBU)	Y	N			Obtained in the concomitant presence of low 3TC concentrations	(60)
E 138 K	GAG to AAG	HIV-1 Specific RTI	TIBO R82913	Y	?			Found in combination with L100I	(27)
E 138 K	GAG to AAG	HIV-1 Specific RTI	UC-82	Y	?	5		Activity of UC-82 versus L100I, K103N, V106A, E138K, Y181C and Y188L reduced by 2-, 6-, 1.5-, 2-, 4- and 200-fold, respectively, compared to wild type	(37, 38)
E 138 K	GAG to AAG	HIV-1 Specific RTI	UC-84 (615985)	Y	?	> 100	TSAOs		(35, 61)
T 139 I	ACA to ATA	HIV-1 Specific RTI	Calamagide A	Y	?	> 70	Not other NNRTIs		(39)
G 141 E	GGG to GAG	HIV-1 Specific RTI	UC-16	Y	N			K101I/G141E: 10-fold	(35)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	+ In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
Q 151 M	CAG to ATG	Multiple Nucleoside Resistance		N	Y	AZT: 10; ddI/ ddC: 5		Pivotal multi nucleoside RTI resistance mutation (first to occur), found in association with combinations of four other mutations: A62V/ V75I/ F77L/ R116Y/Q151M: AZT 190-fold; ddI 50-fold; ddC 20-fold; d4T > 10-fold	(4, 5, 62)
S 156 A	TCA to GCA	Pyrophosphate Analogue RTI	Foscarnet (PPA)	Y	N	4.5			(21)
Q 161 L	CAA to CTA	Pyrophosphate Analogue RTI	Foscarnet (PPA)	Y	Y	5		Q161L/H208Y: 9-fold; Q161L/H208Y suppresses effects of AZT mutations	(20)
V 179 D	GTT to GAT	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?			L100I/V179D/Y181C: 1,000-fold	(31)
V 179 D	GTT to GAT	HIV-1 Specific RTI	L-697,661	N	Y	4			(24)
V 179 D	GTT to GAT	HIV-1 Specific RTI	TIBO R82913	N	Y	20	R82150 (20)		(63)
V 179 D	GTT to GAT	HIV-1 Specific RTI	Trovirdine	Y	?			Found in combination with K103R or Y181C; V179D/Y181C: > 1,000-fold	(44, 45)
V 179 D	GTT to GAT	HIV-1 Specific RTI	UC-10 (645129)	Y	?	16			(35)
V 179 E	GTT to GAG	HIV-1 Specific RTI	L-697,661	N	Y	8			(23)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	Q-APA R18893 (loviride analogue)	Y	?				(64)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	N	Y			K103N and Y181C observed with monotherapy	(46)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	BHAP U-88204E	Y	?				(34)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
Y 181 C	TAT to TGT	HIV-1 Specific RTI	BHAP U-90152 (delavirdine)	?	Y			K103N/Y181C seen separately and in combination in vivo	(46)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	BM451.0836	Y	?				(65)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?	4		L1001/V179D/Y181C: 1,000-fold; uncommon in vivo	(31, 32)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	E-EBU	Y	?				(52)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	E-EPSeU	Y	?	> 50		Y188C confers greater resistance than Y181C	(66)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	E-BPU	Y	?	> 95		Y188C confers greater resistance than Y181C	(66)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	MKC442 (1-EBU)	?	Y				(51)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	L-697,593	Y	?	> 100		K103N/Y181C: > 1,000-fold	(47)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	L-697,661	Y	Y	> 30		K103N and Y181C most common with monotherapy	(24, 48)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	Loviride (R89439, α -APA)	?	Y				(67)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	Nevirapine	Y	Y	> 100	Other NNRTIs	Can suppress effects of AZT mutations	(25, 68, 69)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	NSC 648400 (E-BPTU)	Y	?	160	Other NNRTIs		(70)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	TIBO R82913	Y	?	> 100		K103N/Y181C: > 1,000-fold	(29)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	Trovirdine	Y	?		Nevirapine; 9-chloro-TIBO	V179D/Y181C: > 1,000-fold; Found in combination with K103R or V179D	(44, 45)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-10 (645129)	Y	?	6		K101E/Y181C: 200-fold	(36, 42)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-32 (645542)	Y	?	38			(36)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-38 (629243)	Y	?	8-149	Other NNRTIs		(36, 71)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-040	Y	?	16			(42)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-57 (647014)	Y	?			K101E/Y181C: 58-fold	(36)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In viro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-68 (638532)	Y	?	5			(36)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-69 (646989)	Y	?			V106A/Y181C: 166-fold	(36)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-80 (639475)	Y	?	18			(36)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-81 (615727)	Y	?	53			(35, 72)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-82	Y	?	5			(42)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC-84 (615985)	Y	?	> 118			(36)
Y 181 C	TAT to TGT	HIV-1 Specific RTI	UC781	Y	?	13		V108/ Y181C: 33 fold; K101E/ V108Y Y181C: 500 fold. (42)	
Y 181 C	TAT to TGT	HIV-1 Specific RTI	1737 (Tetrahydro-naphthalene derivative)	Y	?	20		Y181C also confers resistance to numerous other tetrahydronaphthalene derivatives.	(73)
Y 181 I	TGT to ATT	HIV-1 Specific RTI	BHAP U-88204B	Y	Y			Appeared after treatment of Y181C-mutated virus with BHAP; high-level resistance to BHAP, nevirapine and TBO; observed in one nevirapine-treated patient	(74)
Y 181 I	TAT to ATT	HIV-1 Specific RTI	MKC442 (I-BBU)	Y	N	1,000			(60)
Y 181 I	TGT to ATT	HIV-1 Specific RTI	Nevirapine	N	Y	High-level		Observed in one patient	(75)
M 184 I	ATG to ATA	Nucleoside RTI	3TC (lamivudine)	Y	Y			M184V and M184I can suppress effects of AZT resistance mutations	(76, 77, 78)
M 184 T		Nucleoside RTI	3TC (lamivudine)	Y	?			Reduced replication capacity and RT activity	(79, 80)
M 184 V	ATG to GTG	Nucleoside RTI	3TC (lamivudine)	Y	Y	>100	ddI; ddC; (-)-FTC	M184V and M184I can suppress effects of AZT resistance mutations; GTA seen in cell culture	(76, 77, 78)
M 184 V	ATG to GTG	Nucleoside RTI	(-)-FTC	Y	?	> 100		M184V can suppress effects of AZT mutations	(76, 77)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change			Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
M	184	V	ATG to GTG	Nucleoside RTI	I592U89	Y	N	3		K65R/L74V and/or Y115F with M184V: 10-fold; K65R/M184V: 8-fold; L74V/M184V: 9-fold resistance; L74V/Y115F/M184V: 11-fold	(8)
M	184	V	ATG to GTG	Nucleoside RTI	ddC	Y	Y	2-5			(31)
M	184	V	ATG to GTG	Nucleoside RTI	ddI	Y	Y	2-5		Rarely observed in patients receiving ddI	(31)
M	184	V	ATG to GTG	Nucleoside RTI	L-Pd0C	Y	?	> 100			(32)
Y	188	C	TAT to TGT	HIV-1 Specific RTI	B-BPS0U	Y	?	> 250		Y188C is the predominant mutation for B-BPS0U; Y188C confers greater resistance than Y181C	(56)
Y	188	C	TAT to TGT	HIV-1 Specific RTI	B-BPU	Y	?	> 250		Y188C confers greater resistance than Y181C	(66)
Y	188	C	TAT to TGT	HIV-1 Specific RTI	HEPT	Y	?				(52)
Y	188	C	TAT to TGT	HIV-1 Specific RTI	Nevirapine	N	Y				(30)
Y	188	H	TAT to CAT	HIV-1 Specific RTI	BEAP U-87201E (stavudine)	N	Y			K101E, Y188H, E233Y and K238T observed with U-87201E/AZT combination therapy	(41)
Y	188	H	TAT to CAT	HIV-1 Specific RTI	TIBO R82913	Y	?				(27)
Y	188	H/L	TAT to CAT/CTT	HIV-1 Specific RTI	Lopinavir (R89439, α -APA)	?	Y				(67)
Y	188	L	TAT to TTA	HIV-1 Specific RTI	DMP 266 (L-743,726)	Y	?	1,000			(31)
Y	188	L	TAT to TTA	HIV-1 Specific RTI	TIBO R82913	N	Y				(63)
V	189	I	GTA to ATA	HIV-1 Specific RTI	HBV 097	Y	?	2	Other NNRTIs (2-6)		(16)
G	190	A	GGA to GCA	HIV-1 Specific RTI	Lopinavir (R89439, α -APA)	?	Y				(83)
G	190	A	GGA to GCA	HIV-1 Specific RTI	Nevirapine	N	Y				(25)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
G 190 E	GGA to GAA	HIV-1 Specific RTI	AAP-BHAP (U-104489)	Y	?	>100		T139I/ G190E/ T200A/ L214F: >100. Additional mutations possibly restore the replication capacity of the G190E mutant	(84)
G 190 E	GGA to GAA	HIV-1 Specific RTI	HBV 097	Y	?		Other NNRTIs	Reduces enzymatic activity of RT and viral replication competency	(85)
G 190 E	GGA to GAA	HIV-1 Specific RTI	S-2720	Y	?				(86)
G 190 E	GGA to GAA	HIV-1 Specific RTI	UC-38 (629243)	Y	N			K101E/G190E: > 100-fold; cross resistance to: TSAO-m3T, Nev, TIBO R82913, BHAP U88204; susceptible to L697,661	(35, 43)
G 190 E	GGA to GAA	HIV-1 Specific RTI	AAP-BHAP (U-95133)	Y	?	>100		T139I/ G190Q/ T200A/ L214F: >100-fold. Additional mutations possibly restore the replication competency of the G190E mutant.	(84)
G 190 Q	GGA to CAA	HIV-1 Specific RTI	HBV 097	Y	?		Other NNRTIs	Appears exclusively in connection with V179D	(16)
G 190 T	GGA to ?	HIV-1 Specific RTI	HBV 097					Appears under lowered drug concentration selection	(54)
H 208 Y	CAT to TAT	Pyrophosphate Analogue RTI	Foscarnet (PFA)	Y	Y	2		Q161L/H208Y: 9-fold; increased susceptibility to AZT (100-fold), nevirapine (20-fold) and TIBO R82150 (30-fold); Q161L/H208Y suppresses effects of AZT mutations	(20)
H 208 Y	CAT to TAT	Multiple Nucleoside Resistance	AZT + 3TC	?	Y			Polymorphism facilitating AZT+3TC dual resistance	(87)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
L 210 W	TTG to TGG	Nucleoside RTI	AZT	Y	Y			210W/215Y: 42-fold 41L/210W/215Y: 49-fold 41L/67N/70R/210W/215Y: 366-fold Mutation arises after prolonged AZT therapy.	(88, 89, 90)
R 211 K	AGG to AAG	Multiple Nucleoside Resistance	AZT + 3TC	?	Y			Polymorphism facilitating AZT+3TC dual resistance in association with M184V and other AZT resistance mutations.	(87)
L 214 F	CTT to TTT	Multiple Nucleoside Resistance	AZT + 3TC	?	Y			Polymorphism facilitating AZT+3TC dual resistance in association with M184V and other AZT resistance mutations.	(87, 91)
T 215 F	ACC to TTC	Nucleoside RTI	AZT	?	Y			K67N/K70R/T215Y/K219Q: 120-fold	(1, 2, 3)
T 215 Y	ACC to TAC	Nucleoside RTI	AZT	Y	Y			M41L/T215Y: 60-70-fold; K67N/K70R/T215Y/K219Q: 120-fold. Effect of T215Y is reversed by a ddI mutation (L74V), NNRTI mutations (L100I; Y181C) or (-)-FTC/3TC mutations (M184L/V)	(1, 2, 3)
Y 215 C	TTC to TGC	Nucleoside RTI	ddC	N	Y	4		Arises on background of T215Y AZT resistance	(92)
K 219 E	AAA to GAA	Nucleoside RTI	AZT	Y	N				(1, 2, 3)
K 219 Q	AAA to CAA	Nucleoside RTI	AZT	?	Y			K67N/K70R/T215Y/K219Q: 120-fold	(1, 2, 3)
P 225 H	CCT to CAT	HIV-1 Specific RTI	S-2720 (Quinoxaline)	Y	?			P225H follows V106A. Also seen with L101I and Y181C. Double and triple mutants highly resistant to other NNRTI's, including MKC442.	(53)
B 233 V	GAA to GTA	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	N	Y			K101E, Y188H, E233Y and K238T observed with U-87201E/AZT combination therapy	(41)
P 236 L	CCT to CTT	HIV-1 Specific RTI	BHAP U-87201E (atevirdine)	Y	N				(93)

Mutations in HIV RT that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
P 236 L	CCT to CTT	HIV-1 Specific RTI	BHAP U-90152 (delavirdine)	Y	Y			Sensitizes RT 10-fold to nevirapine, TIBO R82913 and L-697,661	(93)
P 236 L	CCT to CTT	HIV-1 Specific RTI	HEPT	Y	?				(70)
K 238 T	AAA to ACA	HIV-1 Specific RTI	BHAP U-87201E (neviridine)	N	Y			K101E, Y188H, E233Y and K238T observed with U-87201E/AZT combi- nation therapy	(41)
G 333 D	GGC to GAC	Multiple Nucleosides	AZT+3TC	Y	Y			Facilitates dual resistance to AZT+3TC in association with M184V and stan- dard AZT resistance mutations.	(87)
G 333 E	GGC to GAG	Multiple Nucleosides	AZT + 3TC	Y	Y			Facilitates dual resistance to AZT+3TC in association with M184V and stan- dard AZT resistance mutations.	(87)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
R 8 K	CGA to AAA	Protease Inhibitor	A-77003	Y	?	10		R8K/ M46I/ G48V: 20-fold	(94, 95)
R 8 Q	CGA to CAA	Protease Inhibitor	A-77003	Y	?	10		M46I improves replication competency of R8Q mutant	(94, 96)
L 10 F	CTC to TTC	Protease Inhibitor	DMP 450	Y	?			Probably compensatory	(97, 98)
L 10 F	CTC to GGC	Protease Inhibitor	VB 11,328	Y	?			L10F/I84V: 8-fold	(99)
L 10 F	CTC to CGC	Protease Inhibitor	VX-478 (141W94)	Y	?				(100)
L 10 F	CTC to CGC	Protease Inhibitor	XM323					L10F/ V82A: 2-fold; L10F/ K45I/ I84V: 50-fold	(101)
L 10 F	CTC to CGC	Protease Inhibitor	SC-55389A	Y	?	2.8	Not SC-52151	N88S/L10F: 25-fold	(102, 103, 104)
L 10 F	CTC to TTC	Protease Inhibitor	BILA 2185 BS	Y	?		BILA 1906 BS (360)	L10F/ L23I/ V32I/ M46I/ I47V/ I54M/ A71V/ I84V: 1,500-fold. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'); p7/p1 cleavage site (Q to R (CAG to CGG) at P3'; A to V (GCT to CTT) at P2'.	(106)
L 10 I	CTC to ATC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(105)
L 10 I		Protease Inhibitor	Ro 31-8959 (saquinavir)		Y			Found in combination with G48V in vivo	(107)
L 10 R	CTC to CGC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	N	Y		XM-323 (15)	L10R/ M46I/ L63P/ V82T: 4-fold; L10R/ M46I/ L63P/ V82T/ I84V: 8-fold	(105)
L 10 V	CTC to GTC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y		A-80987 (4)		(105)
K 20 M	AAG to ATG	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y		VX-478 (8)		(105)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
K 20 R	AAG to AAA	Protease Inhibitor	ABT-538 (ritonavir)	N	Y			K20R/M35I/I54V/V82A: 41-fold	(108)
K 20 R	AAG to AAA	Protease Inhibitor	MK-639 (L- 735,524, indinavir)	?	Y		Ro-31-8959 (8);		(105)
L 23 I	CTA to ATA	Protease Inhibitor	BILA 2185 BS	Y	?		Ro 31-8959 (50); L- 735,524 (80); BILA 1906 BS (360)	L10P/ L23I/ V32I/ M46I/ I47V/ I54M/ A71V/ I84V: 1,500-fold. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'); p7/p1 cleavage site (Q to R (CAG to CCG) at P3', A to V (GCT to CTT) at P2'.	(106, 109)
L 24 I	TTA to ATA	Protease Inhibitor	MK-639 (L- 735,524, indinavir)	?	Y		SC-52151 (8)		(105)
L 24 V	TTA to GTA	Protease Inhibitor	SC-52151	Y	?	10-20	SC55389A	L24V/ G48V/ A71V/ V75I/ P81T: 1000-fold	(102, 103)
D 30 N	GAT to AAT	Protease Inhibitor	AG1343 (nelfinavir)	Y	Y			D30N/A71V: 7-fold; D30N and N88D are most common in vivo after 24 weeks of therapy; they do not cause cross-resistance to other protease inhibitors	(110, 111)
V 32 I	GTA to ATA	Protease Inhibitor	ABT-538 (ritonavir)	Y	?	40		V32I and V82I are synergistic mu- tations yielding 20-fold enzyme resistance	(108)
V 32 I	GTA to ATA	Protease Inhibitor	A-77003	Y	?	7 (enzyme resist.)		V32I appears first; progression to V32I/ M46V and V32I / M46V/ A71V/ V82A occurs even in the absence of drug	(96)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
V 32 I	GTA to ATA	Protease Inhibitor	BILA 1906 BS	Y	?			V32V/A71V: 3-fold; V32V/M46L/ A71V/I84V: 5-fold; V32V/M46L/ A71V/I84A: 520-fold. 32V/46L/71V/ 84A are functionally impaired. Asso- ciated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'))	(112, 106)
V 32 I	GTA to ATA	Protease Inhibitor	BILA 2011 (polinavir)	Y	?	1200	BILA 1906 (1400)	Other mutations found in p1/ p6 cleav- age site	(113)
V 32 I	GTA to ATA	Protease Inhibitor	KNI-272	Y	?	2		V32V/M46V/I84V: 37-fold; V32V/ L33F/K45V/F53L/A71V/I84V/ L89M: 130-fold	(114)
V 32 I	GTA to ATA	Protease Inhibitor	MK-639 (L-735, 524, indinavir)	Y	Y			V32V/M46L/V82A: 3-fold; V32V/ M46L/A71V/V82A: 14-fold	(105)
V 32 I	GTA to ATA	Protease Inhibitor	BILA 2185 BS	Y	?		BILA 1906 (360)	L10F/L23V/V32V/M46V/I47V/I54M/ A71V/I84V: 1,500-fold. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1')); p7/p8 cleavage site (Q to R (CAQ to CQG) at P3', A to V (GCT to CTT) at P2'.	(106)
L 33 F	TTA to TTC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y			M36V/I54V/A71V/V82I: 8-fold; K20R/M36V/I54V/V82A: 41-fold. In vivo, V82A/F/T/S occurs first, often followed by changes at 54, 71 and 36	(108)
M 36 I	ATG to ATA	Protease Inhibitor	ABT-538 (ritonavir)	N	Y			In vivo, V82 occurs first, often fol- lowed by changes at 54, 71 and 36	(108)
M 36 I		Protease Inhibitor	AG1343 (nelmavir)		Y				(110)
K 45 I	AAA to ATA	Protease Inhibitor	XM323					L10F/K45V/I84V: 50-fold	(95)
M 46 F	ATG to TTC	Protease Inhibitor	A-77003	Y	?	4 (enzyme resist.)		Seen with V82A	(96)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
M 46 I	ATG to ATA	Protease Inhibitor	A-77003	Y	?			No effect on susceptibility but improves replication competency of R8Q mutant; R8K/ M46I/ G48V: 20-fold	(94, 96)
M 46 I	ATG to ATA	Protease Inhibitor	ABT-538 (ritonavir)	Y	Y			M46I/ L63P/ A71V/ V82T/ I84V: 27-fold	(108)
M 46 I	ATG to ATA	Protease Inhibitor	AG1343 (nelfinavir)	Y	Y				(110)
M 46 I	ATG to ATA	Protease Inhibitor	BILA 1906 BS	Y	?		L 735,524 (60)	V32I/ A71V: 3-fold; V32I/ M46LL/ A71V/ I84V: 5-fold; V32I/ M46LL/ A71V/ I84A: 520-fold. V32I/M46L/A71V/I84A is functionally impaired. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'))	(106, 109, 112, 113)
M 46 I	ATG to ATA	Protease Inhibitor	BILA 2185 BS	Y	?		BILA 1906 (360)	L10F/ L23I/ V32I/ M46I/ I47V/ I54M/ A71V/ I84V: 1,500-fold. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1')); p7/p1 cleavage site (Q to R (CAG to CCG) at P3'; A to V (GCT to CTT) at P2'.	(106)
M 46 I	ATG to ATA	Protease Inhibitor	DMP 450	Y	?			Probably compensatory	(97, 98)
M 46 I	ATG to ATA	Protease Inhibitor	MK-639 (L-735,524, indinavir)	N	Y			M46I/ L63P/ V82T: 4-fold; L10R/ M46I/ L63P/ V82T: 4-fold; L10R/ M46I/ L63P/ V82T/ I84V: 8-fold	(105, 115)
M 46 I	ATG to ATA	Protease Inhibitor	VB 11,328	Y	?			I50V/ M46I/ I47V: 20-fold	(95, 99)
M 46 I	ATG to ATA	Protease Inhibitor	VX-478 (141W94)	Y	?	Nil			
M 46 L	ATG to TTC	Protease Inhibitor	A-77003	Y	?	2-3 (enzyme resist.)			(96)
M 46 L	ATG to TTG	Protease Inhibitor	BILA 1906 BS	Y	?			Associated p1/ p6 cleavage site mutation (L to F (CTT to TTT) at P1'	(106, 109, 112, 113)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
M 46 L	ATG to TTG	Protease Inhibitor	MK-639 (L-735,524, indinavir)	Y	Y			V32I/ M46L/ A71V/ V82A: 14-fold; V32I/ M46L/ V82A: 3-fold	(95)
M 46 L	ATG to CTG	Protease Inhibitor	XM323	Y	?			V82A/ M46L: 7-fold; V82A/ M46L/ L97V: 11-fold	(101)
M 46 V		Protease Inhibitor	A-77003	Y	?			V32I appears first; progression to V32I/ M46V and V32I/ M46V/ A71V/ V82A occurs even in the absence of drug	(95)
I 47 V	ATA to CTA	Protease Inhibitor	VB 11,328	Y	?			I50V/ M46I/ I47V: 20-fold	
I 47 V	ATA to CTA	Protease Inhibitor	VX-478 (141W94)	Y	?	NIL			
I 47 V	ATA to CTA	Protease Inhibitor	BILA 2185 BS	Y	?		BILA 1906 (360)	L10F/ L23V/ V32I/ M46V/ I47V/ I54M/ A71V/ I84V: 1,500-fold. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'); p7/p1 cleavage site (Q to R (CAG to CCG) at P3', A to V (GCT to CTT) at P2'.	(106)
G 48 V	GGG to GTG	Protease Inhibitor	A-77003	Y	?			R8K/ M46I/ G48V: 20-fold; G48V/ I82T: 100-fold	(116)
G 48 V	GGG to GTG	Protease Inhibitor	Ro 31-8959 (saquinavir)	Y	Y			Found in comb. with L10I in vivo; G48V/ I84V/ L90M: 30-fold; G48V/ L90M: >100-fold enzyme resistance; G48V/ L90M/ I54V: > 50-fold (subtype B or O)	(117, 118)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
G 48 V	GGG to GTG	Protease Inhibitor	SC-52151	Y	?		Ro 31-8959	G48V/ V82A, G48V/ L63P/ V82A or I54T: 10- to 20-fold; L24V/ G48V/ A71V/ V75I/ P81T: 1000-fold	(102, 103)
G 48 V	GGG to GTG	Protease Inhibitor	MP-167	Y	?	20	MP-134(5) SC-52151(16) Ro31-8959(5) (Fold increase in IC90s).	L10F/G48V: 20-fold	(119)
G 48 V	GGG to GTG	Protease Inhibitor	MK-639 (L-735,524, Indinavir)	?	Y				(120)
I 50 V	ATT to GTT	Protease Inhibitor	VB 11,328	Y	?	3		I50V/ M46I/ I47V: 20-fold	(95)
I 50 V	ATT to GTT	Protease Inhibitor	VX-478 (141W94)	Y	?	3			(121)
I 54 M	ATT to ATG	Protease Inhibitor	BILA 2185 BS	Y	?		BILA 1906 (360)	L10F/ L23I/ V32I/ M46I/ I47V/ I54M/ A71V/ I84V: 1,500-fold. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'); p7/p1 cleavage site (Q to R (CAG to CGG) at P3'; A to V (GCT to CTT) at P2'.	(106)
I 54 V	ATC to GTC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y			I54V/V82T: 9-fold; K20R/M36I/I54V/V82A: 41-fold; M36I/I54V/A71V/V82T: 8-fold; I54V/A71V/V82A/L90N: 7-fold; In vivo, V82A/F/T/S occurs first, followed by changes at 54, 71 and 36	(108)
I 54 V	ATC to GTC	Protease Inhibitor	MK-639 (L-735,524, Indinavir)	?	Y				(112)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
I 54 V	ATA to GTA	Protease Inhibitor	Ro 31-8959 (saquinavir)	Y				In subtype O	
I 54 V	ATC to GTC	Protease Inhibitor	Ro 31-8959 (saquinavir)	Y				In subtype B	(117, 118)
D 60 E	GAT to GAA	Protease Inhibitor	DMP 450	Y	?			Probably compensatory	(97, 98)
L 63 P	CTC to CCC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	N	Y			M46I/ L63P/ V82T: 4-fold; L10R/ M46I/ L63P/ V82T/ I84V: 8-fold; L10R/ M46I/ L63P/ V82T: 4-fold	(105)
A 71 T	GCT to ACT	Protease Inhibitor	BMS 186,318	Y	?			A71T/ V82A: 15-fold	(122, 123)
A 71 T	GCT to ACT	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(105)
A 71 V		Protease Inhibitor	A-77003	Y	?			V32I appears first; progression to V32I/ M46V and V32I/ M46V/ A71V/ V82A occurs even in the absence of drug; M46V/ L63P/ A71V/ V82P/ I84V: 27-fold	(95, 101)
A 71 V	GCT to GTT	Protease Inhibitor	ABT-538 (ritonavir)	Y	Y				(108)
A 71 V	GCT to GTT	Protease Inhibitor	AG1343 (nelfinavir)	Y	?	5		D30N/ A71V: 7-fold; M46I/ L63P/ A71V/ I84V: 30-fold	
A 71 V	GCT to GTT	Protease Inhibitor	BILA 1906 BS	Y	?			V32I/ A71V: 3-fold; V32I/ M46I/ A71V/ I84V: 5-fold; V32I/ M46I/ A71V/ I84A: 520-fold. 32I/ 46L/ 71V/ 84A are functionally impaired. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'))	(106, 109, 112, 113)
A 71 V	GCT to GTT	Protease Inhibitor	BILA 2011 (pallinavir)	Y	?		BILA 2185: 30-fold		(112)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
A 71 V	GCT to GTT	Protease Inhibitor	MK-639 (L-735,524, indinavir)	Y	Y			V32I/ M46L/ A71V/ V82A: 14-fold	(95)
A 71 V	GCT to GTT	Protease Inhibitor	SC-52151	Y	?		Not L-735,524	A71V/ V75I/ P81T: 20- to 30-fold; L24V/ G48V/ A71V/ V75I/ P81T: 1000-fold; N88D or H11V/ M46I/ P53L/ A71V/ N88D: 10- to 20-fold	(102, 103)
A 71 V	GCT to GTT	Protease Inhibitor	BILA 2185 B5	Y	?		BILA 1906 (360)	L108I/ L23V/ V32I/ M46I/ I47V/ I54M/ A71V/ I84V: 1,500-fold. Associated Gag mutations: p1/p6 cleavage site (L to P (CTT to TTT at P1'); p7/p1 cleavage site (Q to R (CAG to CGG) at P3'; A to V (GCT to CTT) at P2'.	(106)
G 73 S	GGT to GCT	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y			Emerges following a switch from saquinavir to indinavir.	(124)
V 75 I	GTA to ATA	Protease Inhibitor	SC-52151	Y	?			L24V/ G48V/ A71V/ V75I/ P81T: 1000-fold; A71V/ V75I/ P81T: 20- to 30-fold; L24V/ G48V/ A71V/ V75I/ P81T: 1000-fold	(102, 103)
V 77 I		Protease Inhibitor	AG1343 (nelfinavir)	Y	Y				
P 81 T	CCT to ACT	Protease Inhibitor	SC-52151	Y	?			A71V/ V75I/ P81T: 20- to 30-fold; L24V/ G48V/ A71V/ V75I/ P81T: 1000-fold	(102, 103)
I 82 T	ATC to ACC	Protease Inhibitor	A-77003	Y	?			G48V/ I82T: 100-fold (82T was derived from in vitro passage of 82D)	(125)
V 82 A	GTC to GCC	Protease Inhibitor	A-77003	Y	?			Rare; seen with M46P; V32I appears first; progression to V32I/ M46V and V32I/ M46V/ A71V/ V82A occurs even in the absence of drug	(95, 116, 125)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
V 82 A	GTC to GCC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y	2		In vivo, V82 occurs first, often followed by changes at I54, A71 and M36	(108)
V 82 A	GTC to GCC	Protease Inhibitor	BMS 186,318	Y	?		A-77003 (4)	A71T/ V82A: 15-fold	(122, 123)
V 82 A	GTC to GCC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	Y	Y			V32I/ M46L/ V82A: 3-fold; V32I/ M46L/ A71V/ V82A: 14-fold	(105)
V 82 A	GTC to GCC	Protease Inhibitor	P9941	Y	?	6-8			(126)
V 82 A	GTC to GCC	Protease Inhibitor	SC-52151	Y	?			G48V/ V82A, G48V/ L63P/ V82A or I54T: 10- to 20-fold	
V 82 A		Protease Inhibitor	SKF108922	Y	?				
V 82 A	GTC to GCC	Protease Inhibitor	XM323	Y	?			V82A/ M46L: 7-fold; V82A/ M46L/ L97V: 11-fold; L10F/ V82A: 2-fold; ; V82A/ L97V: 3-fold	(101)
V 82 A	GTC to GCC	Protease Inhibitor	Ra 31-8959 (saquinavir)	?	Y			Follows G48V during saquinavir therapy or after a switch to nelfinavir or indinavir.	(127, 128, 129)
V 82 F	GTC to TTC	Protease Inhibitor	ABT-538 (ritonavir)	Y	Y			V82F/ I84V: 8- to 10-fold; M46I/ L63P/ A71V/ V82F/ I84V: 27-fold	(108)
V 82 F	GTC to TTC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(99)
V 82 F	GTC to TTC	Protease Inhibitor	XM323	Y	?			V82F/ I84V: 92-fold	(101)
V 82 I	GTC to ATC	Protease Inhibitor	A-77003	Y	?			No resistance alone but V32I and V82I are synergistic mutations yielding 20-fold enzyme resistance (82T was derived from in vitro passage of 82I)	(96)
V 82 I	GTC to ATC	Protease Inhibitor	XM323	Y	?	<2			(101)
V 82 S	GTC to TCC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y	6		In vivo, V82 occurs first, often followed by changes at I54, A71 and M36	(108)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
V 82 T	GTC to ACC	Protease Inhibitor	ABT-538 (ritonavir)	N	Y	3		In vivo, V82 occurs first, often followed by changes at 154, A71 and M36; V82T has reduced replication efficacy in natural background	(108)
V 82 T	GTC to ACC	Protease Inhibitor	MK-639 (L-735,524, indinavir)	N	Y			M46I/ L63P/ V82T: 4-fold; L10R/ M46I/ L63P/ V82T: 4-fold; L10R/ M46I/ L63P/ V82T/ I84V: 8-fold	(105)
V 82 T		Protease Inhibitor	SKP108842	Y	†				(130)
V 82 T		Protease Inhibitor	SKP108922	Y	†				
I 84 A	ATA to GCA	Protease Inhibitor	BILA 1906 BS	Y	†		BILA 2185 BS (200)	V32I/ A71V: 3-fold; V32I/ M46I, L/ A71V/ I84V: 5-fold; V32I/ M46I, L/ A71V/ I84A: 520-fold. 32I/ 46L/ 71V/ 84A are functionally impaired. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'))	(106, 109, 112, 113)
I 84 A	ATG to ATA	Protease Inhibitor	BILA 2013 (palinavir)	Y	†		Ro 31-8959 (400);	I84A is the most common mutation	(112)
I 84 V	ATA to GTA	Protease Inhibitor	ABT-538 (ritonavir)	Y	Y			M46I/ L63P/ A71V/ V82P/ I84V: 27-fold; V82P/ I84V: 8- to 10-fold; M46I/ L63P/ A71V/ V82P/ I84V: 27-fold	(108)
I 84 V	ATA to GTA	Protease Inhibitor	AG1343 (nelfinavir)		†			M46I/ L63P/ A71V/ I84V: 30-fold	(110)
I 84 V	ATA to GTA	Protease Inhibitor	BILA 1906 BS	Y	†		BILA 2185 BS(200)	V32I/ A71V: 3-fold; V32I/ M46I, L / A71V/ I84V: 5-fold; V32I/ M46I, L/ A71V/ I84A: 520-fold. 32I/ 46L/ 71V/ 84A are functionally impaired. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'))	(106, 109, 112, 113)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
I 84 V	ATA to GTA	Protease Inhibitor	BILA 2185 BS	Y	?		BILA 1906 BS(360)	L10F/ L23I/ V32I/ M46I/ M7V/ I54M/ A71V/ I84V: 1,500-fold. Associated Gag mutations: p1/p6 cleavage site (L to F (CTT to TTT at P1'); p7/p1 cleavage site (Q to R (CAG to CGG) at P3', A to V (GCT to CTT) at P2'.	(106)
I 84 V	ATA to GTA	Protease Inhibitor	DMP 450	Y	?				(97, 98)
I 84 V	ATA to CTA	Protease Inhibitor	MK-639 (L-735,524, Indinavir)	N	Y			G48V/ I84V/ L90M: 30-fold; L10R/ M46I/ L63P/ V82T/ I84V: 8-fold	(105)
I 84 V	ATA to GTA	Protease Inhibitor	Ro 31-8959 (saquinavir)	Y	?				(95)
I 84 V	ATA to GTA	Protease Inhibitor	RPI-312	Y	?	.5			(131)
I 84 V		Protease Inhibitor	SKP108842	Y	?				
I 84 V	ATA to GTA	Protease Inhibitor	VB 11,328	Y	?			L10F/ I84V: 8-fold	
I 84 V	ATA to GTA	Protease Inhibitor	VX-478 (141W94)	Y	?				
I 84 V	ATA to GTA	Protease Inhibitor	XM323	Y	?	12	P9941; not A-77003 or Ro 31-8959	V82F/ I84V: 92-fold; L10F/ K45M/ I84V: 50-fold	(95, 101)
I 84 V	ATA to GTA	Protease Inhibitor	MP-134	Y	?	10	MP-167(5) ABT-538(10) MK-639(8) SC-52151(8) Ro31-8959(2) (IC90 data)		(132)
N 88 D		Protease Inhibitor	AG1343 (nelfinavir)	Y	Y			D30N and N88D are most common in vivo after 24 weeks of therapy; they do not cause cross-resistance to other protease inhibitors	(110)
N 88 D	AAT to CAT	Protease Inhibitor	SC-52151	Y	?			N88D compensatory, no resistance alone	(132)

Mutations in Protease that confer drug resistance, ordered by position.

Amino Acid Change			Codea Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
N	88	S	AAT to AGT	Protease Inhibitor	SC-55389A	Y	?	20	L735,524 (3); not SC-52151	N88S/L10P: 25	(104)
L	90	M	TTG to ATG	Protease Inhibitor	ABT-538 (nelfinavir)	N	Y			82A/ 54V/ I/ 71V/ 90L/ M: 7-fold	(108)
L	90	M	TTG to ATG	Protease Inhibitor	AG1343 (nelfinavir)	N	Y			Rare in patients	(110)
L	90	M	TTG to ATG	Protease Inhibitor	MK-639 (L-735,524, indinavir)	?	Y				(103)
L	90	M	TTG to ATG	Protease Inhibitor	Ra 31-8959 (saquinavir)	Y	Y			G48V/ L90M: >100-fold enzyme resistance; double mutant rare in vivo; L90M most common in vivo; G48V/ I84V/ L90M: 30-fold	(117)
L	97	V	TTA to GTA	Protease Inhibitor	XM323	Y	?			No resistance alone; V82A/ L97V: 3-fold; V82A/ M46L/ L97V: 11-fold	(101)

Mutations in Envelope that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
R 22 A	AGG to AGA	Fusion/ Binding Inhibitor	RPR103611	Y	?				(132)
I 84 S	ATC to AGC	Fusion/ Binding Inhibitor	RPR103611	Y	?				(132)
S 113 N	AGT to AAT	Fusion/ Binding Inhibitor	Dextran sulphate (DS)	Y	?			S113N/ S134N/ K269E/ Q278E/ N293D/ N323S/ R387I: 250-fold; 113 is in the VI loop region	(132, 133)
S 134 N	AGC to AAC	Fusion/ Binding Inhibitor	Dextran sulphate (DS)	Y	?			V2 loop region; S113N/ S134N/ K269E/ Q278E/ N293D/ N323S/ R387I: 250-fold	(133, 134)
F 145 L	TTC to FTA	Fusion/ Binding Inhibitor	JM-3100	Y	?			Combination of mutations: 2- to 100- fold	(135, 136)
N 188 K	AAT to AAA	Fusion/ Binding Inhibitor	Siamycin I	Y	?			N188K/ G332E/ N351D/ A550T/ N633D/ L762S: 9-fold	(137)
I 228 V	ATA to GTA	Fusion/ Binding Inhibitor	JM-2763	Y	?			Combination of mutations	
K 269 E	AAA to GAA	Fusion/ Binding Inhibitor	Dextran sulphate (DS)	Y	?			V3 loop region; S113N/ S134N/ K269E/ Q278E/ N293D/ N323S/ R387I: 250-fold	(133, 134)
N 270 S	AAT to AGT	Fusion/ Binding Inhibitor	JM-3100	Y	?				
R 272 T	AGA to ACA	Fusion/ Binding Inhibitor	JM-3100	Y	?				
S 274 R	AGT to AGA	Fusion/ Binding Inhibitor	JM-2763	Y	?			Combination of mutations: 95- to 792- fold	(135, 136)
S 274 R	AGT to AGA	Fusion/ Binding Inhibitor	JM-3100	Y	?	DS (> 7 to 6,667)			
Q 278 H	CAG to CAT	Fusion/ Binding Inhibitor	Dextran sulphate (DS)	Y	?			V3 loop region; S113N/ S134N/ K269E/ Q278E/ N293D/ N323S/ R387I: 250-fold	(133, 134)
Q 278 H	CAG to CAT	Fusion/ Binding Inhibitor	JM-2763	Y	?				
Q 278 H	CAG to CAC	Fusion/ Binding Inhibitor	JM-3100	Y	?				
I 288 V	ATA to GTA	Fusion/ Binding Inhibitor	JM-3100	Y	?				
N 293 D	AAT to GAT	Fusion/ Binding Inhibitor	Dextran sulphate (DS)	Y	?			V3 loop region; S113N/ S134N/ K269E/ Q278E/ N293D/ N323S/ R387I: 250-fold	(133, 134)

Mutations in Envelope and Drug Resistance

Mutations in Envelope that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
N 293 H	AAT to CAT	Fusion/ Binding Inhibitor	JM-3100	Y	?				
A 297 T	GCA to ACA	Fusion/ Binding Inhibitor	JM-2763	Y	?				
A 297 T	GCA to ACA	Fusion/ Binding Inhibitor	JM-3100	Y	?				
N 323 S	AAT to AGT	Fusion/ Binding Inhibitor	Dextran sulphate (DS)	Y	?			C3 region; S113N/ S134N/ K269E/ Q278E/ N293D/ N323S/ R387I: 250- fold	(133, 134)
G 332 E	GGA to GAA	Fusion/ Binding Inhibitor	Siamycin I	Y	?			N188K/ G332E/ N351D/ A550T/ N633D/ L762S: 9-fold	(137)
N 351 D	AAT to GAT	Fusion/ Binding Inhibitor	Siamycin I	Y	?			N188K/ G332E/ N351D/ A550T/ N633D/ L762S: 9-fold	(137)
P 385 L	CCA to CTA	Fusion/ Binding Inhibitor	JM-2763	Y	?				
P 385 L	CCA to CTA	Fusion/ Binding Inhibitor	JM-3100	Y	?				
R 387 I	AGA to ACA	Fusion/ Binding Inhibitor	Dextran sulphate (DS)	Y	?			CD4 binding region; S113N/ S134N/ K269E/ Q278E/ N293D/ N323S/ R387I: 250-fold	(133, 134)
Q 410 E	CAA to GAA	Fusion/ Binding Inhibitor	JM-3100	Y	?				
S 433 P	TCC to CCC	Fusion/ Binding Inhibitor	JM-3100	Y	?				
V 457 I	GTA to ATA	Fusion/ Binding Inhibitor	JM-3100	Y	?				
A 550 T	GCC to ACC	Fusion/ Binding Inhibitor	Siamycin I	Y	?			N188K/ G332E/ N351D/ A550T/ N633D/ L762S: 9-fold	(137)
N 633 D	AAT to GAT	Fusion/ Binding Inhibitor	Siamycin I	Y	?			N188K/ G332E/ N351D/ A550T/ N633D/ L762S: 9-fold	(137)
L 762 S	TTG to TCG	Fusion/ Binding Inhibitor	Siamycin I	Y	?			N188K/ G332E/ N351D/ A550T/ N633D/ L762S: 9-fold	(137)

Mutations in SIV Reverse Transcriptase that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
K 65 R	AAA to AGA	SIV Nucleoside RT Inhibitor	PMPA	?	Y	5	3TC (80); ddI; ddC; d4T; PMEA	K65R appears first, followed by N69S and I118V. Observed changes at N69S and I118V do not result in increased resistance.	(138, 139, 140)
Q 151 M	CAG to ATG	SIV Nucleoside RT Inhibitor	AZT	?	Y	>100	ddI; ddC; d4T; 3TC		(141)
M 184 V	ATG to GTG	SIV Nucleoside RT Inhibitor	(-)-FTC	Y	?				(82)

Mutations in FIV Reverse Transcriptase that confer drug resistance, ordered by position.

Amino Acid Change	Codon Change	Class of Drug	Compound	In vitro	In vivo	-Fold -resistance	Cross-resistance (-fold)	Comments	Refs
D 3 H	GAT to CAT	FIV Nucleoside RT Inhibitor	ddC	Y	?	4	ddI; PFA		(142, 143)
V 47 I	GTA to ATA	FIV Nucleoside RT Inhibitor	d4T	Y	?	4-6	PFA (>50); AZT; ddI; PMEA		(144)
P 156 S	CCA to TCA	FIV Nucleoside RT Inhibitor	3TC	Y	?	?	AZT (4), AZT + 3TC (6)		(145)
M 183 Y	ATG to ACG	FIV Nucleoside RT Inhibitor	(-)-FTC	Y	?	10	ddC	Corresponds to 184 in HIV; M183V recombinant displays 10-fold resistance to 3TC or (-)-FTC.	(146)

Abbreviations

Amino acids

A	alanine
C	cysteine
D	aspartate
E	glutamate
F	phenylalanine
G	glycine
H	histidine
I	isoleucine
K	lysine
L	leucine
M	methionine
N	asparagine
P	proline
Q	glutamine
R	arginine
S	serine
T	threonine
V	valine
W	tryptophan
Y	tyrosine

Compounds

1592U89	(1 <i>S</i> ,4 <i>R</i>)-4-[2-amino-6-cyclopropyl-amino)-9 <i>H</i> -purin-9-yl]-2-cyclopentene-1-methanol succinate (a carbovir analogue, Glaxo Wellcome)
3TC	(-)- β -L-2',3'-dideoxy-3'-thiacytidine (Glaxo Wellcome)
1737	Tetrahydronaphthalene lignan derivative
α -APA R18893	α -nitro-anilino-phenylacetamide
A-77003, A-75925 and A-80987	C2 symmetry-based protease inhibitors (Abbott Laboratories)
AAP-BHAP	bisheteroaryl piperazine analogue (Pharmacia & Upjohn)
ABT-538	C2 symmetry-based protease inhibitor (Abbott Laboratories)
AZdU	3'-azido-2',3'-dideoxyuridine
AZT	3'-azido-3'-deoxythymidine (Glaxo Wellcome)
AZT-p-ddI	3'-azido-3'-deoxythymidyl-(5',5')-2',3'-dideoxyinosinic acid (Ivax)
BHAP	bisheteroaryl piperazine
BILA 1906	<i>N</i> -{1 <i>S</i> -[[[3-[2 <i>S</i> -(1,1-dimethylethyl)amino]carbonyl-4 <i>R</i> -]3-pyridinylmethyl]thio]-1-piperidinyl]-2 <i>R</i> -hydroxy-1 <i>S</i> -(phenylmethyl)propyl]amino]carbonyl]-2-methylpropyl}-2-quinolinecarboxamide (Bio-Mega/Boehringer Ingelheim)
BILA 2185	<i>N</i> -(1,1-dimethylethyl)-1-[2 <i>S</i> -[[2-2,6-dimethylphenoxy)-1-oxoethyl]amino]-2 <i>R</i> -hydroxy-4-phenylbutyl]4 <i>R</i> -pyridinylthio)-2-piperidine-carboxamide (Bio-Mega/Boehringer Ingelheim)
BM+51.0836	thiazolo-isindolinone derivative
BMS 186,318	aminodiol derivative HTV-1 protease inhibitor (Bristol-Myers Squibb)

Abbreviations (cont)

Compounds (cont)

d4API	9-[2,5-dihydro-5-(phosphonomethoxy)-2-furanyl]adenine (Gilead Sciences)
d4C	2',3'-didehydro-2',3'-dideoxycytidine
d4T	2',3'-didehydro-3'-deoxythymidine (Bristol-Myers Squibb)
ddC	2',3'-dideoxycytidine (Roche)
ddI	2',3'-dideoxyinosine (Bristol-Myers Squibb)
DMP 266	a 1,4-dihydro-2 <i>H</i> -3,1-benzoxazin-2-one
DMP 450	[4 <i>R</i> -(4- α ,5- α ,6- β ,7- β)]-hexahydro-5,6-bis(hydroxy)-1,3-bis(3-amino)phenyl[methyl]-4,7-bis(phenylmethyl)-2 <i>H</i> -1,3-diazepin-2-one-bismesylate (Avid Therapeutics)
DXG	(-)- β -D-dioxolane-guanosine
EBU-dM	5-ethyl-1-ethoxymethyl-6-(3,5-dimethylbenzyl)uracil
E-EBU	5-ethyl-1-ethoxymethyl-6-benzyluracil
DS	dextran sulphate
E-EPSeU	1-(ethoxymethyl)-(6-phenylselenyl)-5-ethyluracil
E-EPU	1-(ethoxymethyl)-(6-phenylthio)-5-ethyluracil
F-ddA	2'-fluoro-2',3'-dideoxyadenosine
(-)-FTC	(-)- β -L-2',3'-dideoxy-5-fluoro-3'-thiacytidine (Triangle Pharmaceuticals)
HBV 097	(<i>S</i>)-4-isopropoxycarbonyl-6-methoxy-3-(methylthio-methyl)-3,4-dihydroquinoxalin-2(1 <i>H</i>)-thione
HEPT	1-[(2-hydroxyethoxy)methyl]6-(phenylthio)thymine
JM2763	1,1'-(1,3-propanediyl)-bis-1,4,8,11-tetraazacyclo-tetradecane (Johnson Matthey)
JM3100	1,1'-[1,4-phenylenebis-(methylene)]bis-(1,4,8,11-tetraazacyclotetradecane) octahydrochloride dihydrate (Johnson Matthey)
KNI-272	(2 <i>S</i> ,3 <i>S</i>)-3-amino-2-hydroxy-4-phenylbutyric acid-containing tripeptide
L-697,593	5-ethyl-6-methyl-3-(2-phthalimido-ethyl)pyridin-2(1 <i>H</i>)-one
L-697,661	3-[(4,7-dichloro-1,3-benzoxazol-2-yl)methyl]amino-5-ethyl-6-methylpyridin-2(1 <i>H</i>)-one
L-FDDC	(-)- β -L-5-fluoro-2',3'-dideoxy-cytidine
L-FDOC	(-)- β -L-5-fluoro-dioxolane cytosine
MK-639	hydroxy-aminopentane amide HIV-1 protease inhibitor (Merck & Co)
MKC442	6-benzyl-1-ethoxymethyl-5-isopropyluracil (I-EBU, Triangle Pharmaceuticals/Mitsubishi)
MP-134	C2 symmetry-based protease inhibitor
MP-167	C2 symmetry-based protease inhibitor
nevirapine	11-cyclopropyl-5,11-dihydro-4-methyl-6 <i>H</i> -dipyridol[3,2- <i>b</i> :2',3'- <i>e</i>]diazepin-6-one (Boehringer Ingelheim)
NNRTI	non-nucleoside reverse transcriptase inhibitor
NSC648400	1-benzoyloxymethyl-5-ethyl-6-(α -pyridylthio)uracil (E-BPTU)
P9941	[2-pyridylacetyl-PhePheAla-y(CHOH)] ₂ (Dupont Merck)
PFA	phosphonoformate (foscarnet, Astra)
PMBA	9-(2-phosphonylmethoxyethyl)adenine (Gilead Sciences)
PMPA	(<i>R</i>)-9-(2-phosphonyl-methoxypropyl)adenine (Gilead Sciences)
Ro 31-8959	hydroxyethylamine derivative HIV-1 protease inhibitor (Roche)

Abbreviations (cont)

Compounds (cont)

RPI-312	1-[(3 <i>S</i>)-3-(<i>n</i> -alpha-benzoyloxycarbonyl)-l-asparginy]-amino-2-hydroxy-4-phenyl-butyl]- <i>n</i> -tert-butyl-l-proline amide (peptidyl protease inhibitor)
RPR103611	
RT	reverse transcriptase
S-2720	6-chloro-3,3-dimethyl-4-(isopropenyl-oxycarbonyl)-3,4-dihydro-quinoxalin-2(1 <i>H</i>)-thione
SC-52151	hydroxyethylurea isostere protease inhibitor (Searle)
SC-55389A	hydroxyethyl-urea isostere protease inhibitor (Searle)
TIBO R82150	(+)-(5 <i>S</i>)-4,5,6,7-tetrahydro-5-methyl-6-(3-methyl-2-butenyl)-imidazo[4,5,1- <i>jk</i>][1,4]-benzodiazepin-2(1 <i>H</i>)-thione (Janssen)
TIBO 82913	(+)-(5 <i>S</i>)-4,5,6,7-tetrahydro-9-chloro-5-methyl-6-(3-methyl-2-butenyl)-imidazo-[4,5,1- <i>jk</i>]-[1,4]benzo-diazepin-2(1 <i>H</i>)-thione (Janssen)
TSAO-m ³ T	[2',5'-bis- <i>O</i> -(tert-butyl-dimethylsilyl)-3'-spiro-5'-(4'-amino-1',2'-oxathiole-2',2'-dioxide)]-β-D-pentofuranosyl-N ³ -methylthymine
U-90152	1-[3-[(1-methylethyl)-amino]-2-pyridinyl]-4-[[5-[(methylsulphonyl)-amino]-1 <i>H</i> -indol-2yl]carbonyl]-piperazine
U-95133	(Alkylamino)piperidine bis(heteroaryl)piperazine analog
U-104489	(Alkylamino)piperidine bis(heteroaryl)piperazine analog
UC-040	thiocarboxanilide derivative (Uniroyal Chemical Co)
UC	thiocarboxanilide derivatives (Uniroyal Chemical Co)
UC-781	<i>N</i> -[4-chloro-3-(3-methyl-2-butenyloxy)phenyl]-2-methyl-3-furan-carbothioamide
UC-82	<i>N</i> -[4-chloro-3-(3-methyl-2-butenyloxy)phenyl]-2-methyl-3-thiophene-carbothioamide
VB 11,328	hydroxyethyl-sulphonamide protease inhibitor (Vertex Pharmaceuticals)
VX-478	hydroxyethylsulphonamide protease inhibitor (Vertex Pharmaceuticals)
XM 323	cyclic urea protease inhibitor (Dupont Merck)

References

- [1] Larder BA and Kemp SD. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* 1989; 246:1155-1158.
- [2] Larder BA, Coates KE and Kemp SD. Zidovudine-resistant human immunodeficiency virus selected by passage in cell culture. *Journal of Virology* 1991; 65(10):5232-5236.
- [3] Kellam P, Boucher CA and Larder BA. Fifth mutation in human immunodeficiency virus type 1 reverse transcriptase contributes to the development of high-level resistance to zidovudine. *Proceedings of the National Academy of Sciences, USA* 1992; 89(5):1934-1938.
- [4] Iversen AK, Shafer RW, Wehrly K, Winters MA, Mullins JI, Chesebro B and Merigan TC. Multidrug-resistant human immunodeficiency virus type 1 strains resulting from combination antiretroviral therapy. *Journal of Virology* 1996; 70(2):1086-1090.
- [5] Shirasaka T, Kavlick MF, Ueno T, Gao WY, Kojima E, Alcaide ML, Choekijchai S, Roy BM, Arnold E, Yarchoan R and Mitsuya H. Emergence of human immunodeficiency virus type 1 variants with resistance to multiple dideoxynucleosides in patients receiving therapy with dideoxynucleosides. *Proceedings of the National Academy of Sciences, USA* 1995; 92:1-5.
- [6] Zhang D, Caliendo AM, Eron JJ, DeVore KM, Kaplan JC, Hirsch MS and D'Aquila RT. Resistance to 2',3'-dideoxycytidine conferred by a mutation in codon 65 of the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrobial Agents and Chemotherapy* 1994; 38(2):282-287.
- [7] Gu Z, Gao Q, Fang H, Salomon H, Parniak MA, Goldberg E, Cameron JM and Wainberg MA. Identification of a mutation at codon 65 in the IKKK motif of reverse transcriptase that encodes human immunodeficiency virus resistance to 2',3'-dideoxycytidine and 2',3'-dideoxy-3'-thiacytidine. *Antimicrobial Agents and Chemotherapy* 1994; 38(2):275-281.
- [8] Tisdale M, Alnadaf T, Cousens D. Combination of mutations in human immunodeficiency virus type 1 reverse transcriptase required for resistance to the carbocyclic nucleoside 1592U89. *Antimicrobial Agents and Chemotherapy* 1997; 41(5):1094-1098.
- [9] Mellors JW, Bazmi H, Chu CK and Schinazi RF. K65R mutation in HIV-1 reverse transcriptase causes resistance to (-)- β -D-dioxolane-guanine and reverses AZT resistance. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3-6 July, 1996. Abstract 7.
- [10] Gu Z, Salomon H, Cherrington JM, Mulato AS, Chen MS, Yarchoan R, Folli A, Sogocio KM and Wainberg MA. K65R mutation of human immunodeficiency virus type 1 reverse transcriptase encodes cross-resistance to 9-(2-phosphonylmethoxyethyl)adenine. *Antimicrobial Agents and Chemotherapy* 1995; 39(8):1888-1891.
- [11] Folli A, Sogocio KM, Anderson B, Kavlick M, Saville MW, Wainberg MA, Gu X, Cherrington J, Mitsuya H and Yarchoan R. In vitro selection and molecular characterization of human immunodeficiency virus type 1 with reduced sensitivity to 9-[2-(phosphonomethoxy-ethyl)adenine (PMEA). *Antiviral Research* 1996; 32(2): 91-98.
- [12] Cherrington JM, Chandok R, Mulato AS, Lamy PD, Mitsuya H, Wainberg M. In vitro selection and characterization of HIV-1 variants with reduced susceptibility to PMPA. 6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA: Abstract 26.
- [13] Fitzgibbon JE, Howell RM, Haberzettl CA, Sperber SJ, Gockle DJ and Dubin DT. Human immunodeficiency virus type 1 pol gene mutations which cause decreased susceptibility to 2',3'-dideoxycytidine. *Antimicrobial Agents and Chemotherapy* 1992; 36:153-157.
- [14] Cherrington J, Mulato AS, Fuller MD and Chen MS. A novel mutation (K70E) in HIV-1 reverse transcriptase confers decreased susceptibility to 9-[2-(phosphonomethoxy-ethyl)adenine (PMEA) in vitro. *Antimicrobial Agents and Chemotherapy* 1996; 40(9): 2212-2216.
- [15] Mulato AS, Lamy PL, Li W, Miller MD, Cherrington JM. Genotypic characterization of HIV-1 variants isolated from AIDS patients treated with adefovir dipivoxil (bis-POM PMEA). 6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA: Abstract 24.

References

- [16] Kleim JP, Rosner M, Winkler I, Paessens A, Kirsch R, Hsiou Y, Arnold E and Riess G. Selective pressure of a quinoxaline nonnucleoside inhibitor of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) on HIV-1 replication results in the emergence of nucleoside RT-inhibitor-specific (RT Leu-74 to Val or Ile and Val-75 to Leu or Ile) HIV-1 mutants. *Proceedings of the National Academy of Sciences, USA* 1996; 93(1):34-38.
- [17] St Clair MH, Martin JL, Tudor-Williams G, Bach MC, Vavro CL, King DM, Kellam P, Kemp SD and Larder BA. Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. *Science* 1991; 253:1557-1559.
- [18] Lacey SF and Larder BA. Novel mutation (V75T) in human immunodeficiency virus type 1 reverse transcriptase confers resistance to 2',3'-dideoxy-2',3'-dideoxythymidine in cell culture. *Antimicrobial Agents and Chemotherapy* 1994; 38(6):1428-1432.
- [19] Schinazi RF, Stuyver L, Wyseur A, Lloyd RM Jr, Hough L, Rombout A, Rossau R, and Rimland D. Proviral and plasma virus genotyping using a line probe assay in nucleoside treated HIV infected Veterans Affairs patients. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3-6 July, 1996. Abstract 65.
- [20] Mellors J, Bazmi H, Schinazi RF, Roy B, Hsiou Y, Arnold E, Weir J and Mayers D. Novel mutations in the reverse transcriptase of human immunodeficiency virus type 1 reduce susceptibility to foscarnet in laboratory and clinical isolates. *Antimicrobial Agents and Chemotherapy* 1995; 39(5):1087-1092.
- [21] Tachedjian G, Hooker DJ, Gurusinge AD, Bazmi H, Deacon NJ, Mellors J, Birch C and Mills J. Characterisation of foscarnet-resistant strains of human immunodeficiency virus type 1. *Virology* 1995; 212(1):58-68.
- [22] Tachedjian G, Mellors J, Bazmi H, Birch C and Mills J. Zidovudine resistance is suppressed by mutations conferring resistance of human immunodeficiency virus type 1 to foscarnet. *Journal of Virology* 1996; 70:7171-7181.
- [23] Prasad VR, Lowy I, de los Santos T, Chiang L and Goff SP. Isolation and characterization of a dideoxyguanosine triphosphate-resistant mutant of human immunodeficiency virus reverse transcriptase. *Proceedings of the National Academy of Sciences, USA* 1991; 88(24):11363-11367.
- [24] Byrnes VW, Sardana VV, Schleif WA, Condra JH, Waterbury JA, Wolfgang JA, Long WJ, Schneider CL, Schlabach AJ, Wolanski BS, Graham DJ, Gotlib L, Rhodes A, Titus DL, Roth E, Blahy OM, Quintero JC, Staszewski S and Emini BA. Comprehensive mutant enzyme and viral variant assessment of human immunodeficiency virus type 1 reverse transcriptase resistance to nonnucleoside inhibitors. *Antimicrobial Agents and Chemotherapy* 1993; 37(8):1576-1579.
- [25] Richman DD, Havlir D, Corbeil J, Looney D, Ignacio C, Spector SA, Sullivan J, Cheeseman S, Barringer K, Pauletti D, Shih CK, Myers M and Griffin J. Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy. *Journal of Virology* 1994; 68(3):1660-1666.
- [26] Mellors JW, Im GJ, Tramontano E, Winkler SR, Medina DJ, Deutschman GE, Bazmi HZ, Piras G, Gonzalez CJ and Cheng YC. A single conservative amino acid substitution in the reverse transcriptase of human immunodeficiency virus-1 reverse transcriptase confers resistance to TIBO R82150. *Molecular Pharmacology* 1993; 43(1):11-16.
- [27] Balzarini J, Karlsson A, Perez-Perez MJ, Vrang L, Walbers J, Zhang H, Oberg B, Vandamme AM, Camarasa MJ and De Clercq E. HIV-1 specific reverse transcriptase inhibitors show differential activity against HIV-1 mutant strains containing different amino acid substitutions in the reverse transcriptase. *Virology* 1993; 192:246-253.
- [28] Byrnes V, Blahy O, Condra J, Gotlib L, Graham D, Long W, Quintero J, Rhodes A, Roth E, Sardana V, Schlabach A, Schleif W, Schneider C, Titus D, Wolanski B, Wolfgang J and Emini E. Phenotypic susceptibility of human immunodeficiency virus type 1 RT containing substitutions which engender resistance to nucleoside and non-nucleoside inhibitors. Third Workshop on Viral Resistance, 1993, Gaithersburg, MD, USA.

- [29] Larder, BA. 3'-Azido-3'-deoxythymidine resistance suppressed by a mutation conferring human immunodeficiency virus type 1 resistance to nonnucleoside reverse transcriptase inhibitors. *Antimicrobial Agents and Chemotherapy* 1992; 36(12):2664-2669.
- [30] Richman DD. Resistance of clinical isolates of human immunodeficiency virus to antiretroviral agents. *Antimicrobial Agents and Chemotherapy* 1993; 37(6):1207-1213.
- [31] Winslow DL, Garber S, Reid C, Scarnati H, Baker D, Rayner, MM, Anton ED. Selection conditions affect the evolution of specific mutations in the reverse transcriptase gene associated with resistance to DMP 266. *AIDS* 1996; 10(11):1205-1209.
- [32] Young SD, Britcher SF, Tran LO, Payne LS, Lumma WC, Lyle TA, Anderson JR, Huff PS, Olsen DB, Carroll SS, Pettibone DJ, O'Brien JA, Ball RG, Balani SK, Lin JH, Chen IW, Schleif WA, Sardana VV, Long WJ, Byrenes VW and Emini EA. L-743,726 (DMP-266): a novel, highly potent nonnucleoside inhibitor of the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrobial Agents and Chemotherapy* 1995; 39(12):2602-2609.
- [33] Balzarini J, Karlsson A, Perez-Perez MJ, Camarasa MJ, Tarpley WG and De Clercq E. Treatment of human immunodeficiency virus type 1 (HIV-1)-infected cells by combinations of HIV-1-specific inhibitors results in a different resistance pattern than does treatment with single-drug therapy. *Journal of Virology* 1993; 67(9):5353-5359.
- [34] Vasudevachari MB, Battista C, Lane HC, Psallidopoulos MC, Zhao B, Cook J, Palmer JR, Romero DL, Tarpley WG and Salzman NP. Prevention of the spread of HIV-1 infection with nonnucleoside reverse transcriptase inhibitors. *Virology* 1992; 198(1):269-2977.
- [35] Balzarini J, Perez-Perez MJ, Velazquez S, San-Felix A, Camarasa MJ, De Clercq E and Karlsson A. Suppression of the breakthrough of human immunodeficiency virus type 1 (HIV-1) in cell culture by thiocarboxanilide derivatives when used individually or in combination with other HIV-1-specific inhibitors (ie, TSAO derivatives). *Proceedings of the National Academy of Sciences, USA* 1995; 92:5470-5474.
- [36] Buckheit RW, Kinjerski TL, Fliakas-Boltz V, Russell JD, Stup TL, Pallansch LA, Brouwer WG, Dao DC, Harrison WA, Schultz RJ, Bader JP and Yang SS. Structure-activity and cross-resistance evaluations of a series of human immunodeficiency virus type 1-specific compounds related to oxathiin carboxanilide. *Antimicrobial Agents and Chemotherapy* 1995; 39(12):2718-2727.
- [37] Balzarini J, Pelemans H, Aquaro S, Perno CF, Witvrouw M, Schols D, De Clercq E and Karlsson A. Highly favourable antiviral activity and resistance profile of the novel thiocarboxanilide pentenyloxy ether derivatives UC-781 and UC-82 as inhibitors of human immunodeficiency virus type 1 (HIV-1) replication. *Molecular Pharmacology* 1996; 50(2): 394-410.
- [38] Balzarini J, Brouwer WG, Dao DC, Osika EM and De Clercq E. Identification of novel thiocarboxanilide derivatives that suppress a variety of drug-resistant mutant human immunodeficiency virus type 1 strains at a potency similar to that for wild-type virus. *Antimicrobial Agents and Chemotherapy* 1996; 40(6):1454-1466.
- [39] Buckheit RW Jr, Fliakas-Boltz V, Decker WD, Roberson JL, Stup TL, Pyle CA, White EL, McMahon JB, Currens MJ, Boyd MR and Bader JP. Comparative anti-HIV evaluation of diverse HIV-1-specific reverse transcriptase inhibitor-resistant virus isolates demonstrates the existence of distinct phenotypic subgroups. *Antiviral Research* 1995; 26:117-132.
- [40] Moeremans M, De Raeymaeker M, Van den Broeck R, Stoffels P, De Brabander M, De Cree J, Hertogs K, Pauwels R, Staszewski S and Andries K. Virological analysis of HIV-1 isolates in patients treated with the non-nucleoside reverse transcriptase inhibitor RO91767, 8-chloro-TIBO. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6-9 July, 1995. Abstract 33.

References

- [41] Demeter L, Resnick L, Nawaz T, Timpone JG Jr, Batts D and Reichman RC. Phenotypic and genotypic analysis of atevirdine (ATV) susceptibility of HIV-1 isolates obtained from patients receiving ATV monotherapy in a phase I clinical trial (ACTG 187): comparison to patients receiving combination therapy with ATV and zidovudine. Third Workshop on Viral Resistance, 1993, Gaithersburg, MD, USA.
- [42] Balzarini J, Brouwer WG, Pelauer EE, De Clercq E and Karlsson A. Activity of various thiocarboxanilide derivatives against wild-type and several mutant human immunodeficiency virus type 1 strains. *Antiviral Research* 1995; 27:219-236.
- [43] Buckheit RW Jr, Snow JH, Gliakas-Boltz V, Kinjerski TL, Russell JD, Rallanssch LA, Brouwer WG, Yang SS. Highly potent oxathiin carboxanilide derivatives with efficacy against NNRTI resistant isolates. *Antimicrob Agents Chemother* 1997; 41(4):831-837.
- [44] Zhang H, Vrang L, Backbro K, Lindz P, Sahlberg C, Unge T and Oberg B. Inhibition of human immunodeficiency virus type 1 wild-type and mutant reverse transcriptase by the phenyl ethyl thiozoyl thiourea derivatives trovirdine and MSG-127. *Antiviral Research* 1995; 28:331-342.
- [45] Vrang L, Rydergard C, Ahgren C, Engelhardt P, Hogberg M, Johansson NG, Kangasmetza J, Lind P, Noreen R, Sahlberg C, Zhou XX, Karlsson A, Lopez C, Morin Jr JM, Ternansky RJ, Bell FW, Jordan CL, Kinnick MD, Palkowitz JA, Parrish CA, Franc P, Vasileff RT, West SJ and Oberg B. Comparative rates of in vitro resistance development of HIV-1 to non-nucleoside analog RT inhibitors. *Antiviral Research* 1993; 28(Supplement 1):77.
- [46] Demeter LM, Shafer RW, Para M, Morse G, Freimuth W, Merigan TC and Reichman RC. Delavirdine (DLV) susceptibility of HIV-1 isolates obtained from patients receiving DLV monotherapy (ACTG 260). *J Acquir Immune Defic Syndrom Hum Retrovir* 1995; 10(S11).
- [47] Nunberg JH, Schleif WA, Boots EJ, O'Brien JA, Quintero JC, Hoffman JM, Emmini EA and Goldman ME. Viral resistance to human immunodeficiency virus type 1-specific pyridinone reverse transcriptase inhibitors. *Journal of Virology* 1991; 65(9):4887-4892.
- [48] Saag MS, Emmini EA, Laskin OL, Douglas J, Lapidus WI, Schleif WA, Whitley RJ, Hildebrand C, Byrnes VW, Kappes JC, Anderson KW, Massari FE and Shaw GM. A short-term clinical evaluation of L-697,661, a non-nucleoside inhibitor of HIV-1 reverse transcriptase L-697,661 Working Group. *New England Journal of Medicine* 1993; 329(15):1065-1072.
- [49] Staszewski S, Miller V, Kober A, Colebunders R, Vandercam B, Delescluse J, Chumek N, Van Wanzeele F, De Brabander M, De Cree J, Moeremans M, Andries K, Boucher C, Stoffels P and Janssen PAJ. Evaluation of the efficacy and tolerance of RO18893, RO89439 (loviride) and placebo in asymptomatic HIV-1-infected patients. *Antiviral Therapy* 1996; 1:42-50.
- [50] Seki M, Sadakata Y, Yuasa S and Baba M. Isolation and characterization of human immunodeficiency virus type-1 mutants resistant to the non-nucleoside reverse transcriptase inhibitor MKC-442. *Antiviral Chemistry and Chemotherapy* 1995; 6(2):73-79.
- [51] Borroto-Esoda K, Noel DS, Moxham CP, Furman PA. Preliminary genotypic analysis of HIV-1 in plasma from volunteers receiving repeated multiple doses of MKC-442. 6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA: Abstract 22.
- [52] Balzarini J, Karlsson A and De Clercq E. Human immunodeficiency virus type 1 drug-resistance patterns with different 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine derivatives. *Molecular Pharmacology* 1993; 44(4):694-701.
- [53] Balzarini J, Pelemans H, Esnouf R, Dunkler A, Parniak MA, Bandamme A-M, Karlsson A, De Clercq E, Kleim J-P. Significance of the 225 Pro(His mutation in HIV-1 reverse transcriptase. 6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA: Abstract 21.
- [54] Kleim JP, Winkler I, Rosner M, Kirsch R, Rubsamen-Waigmann H, Paessens A and Reiss G. In vitro selection for different mutational patterns in the HIV-1 reverse transcriptase using high and low selective pressure of the nonnucleoside reverse transcriptase inhibitor HBY 097. *Virology* 1997; 231:112-118.

- [55] Vandamme A-M. Polymerase chain reaction (PCR) as a diagnostic tool in HIV infection. *Verhandelingen van de Koninklijke Academie voor Geneeskunde van België* 1994; 56(3):231-265.
- [56] Tanaka M, Srinivas RV, Ueno T, Kavlick MF, Hui FD, Fridland A, Driscoll JS, Mitsuya H. In vitro induction of human immunodeficiency virus type 1 variants resistant to 2'-b-fluoro-2',3'-dideoxyadenosine. *Antimicrob Agents Chemother* 1997; 41(6):1313-1318.
- [57] Balzarini J, Velazquez S, Sanfelix A, Karlsson A, Perez-Perez MJ, Camarasa MJ and De Clercq E. Human immunodeficiency virus type-1 specific purine analogues show a resistance spectrum that is different from that of the human immunodeficiency virus type-1-specific non-nucleoside analogues. *Molecular Pharmacology* 1993; 43(1):109-114.
- [58] Balzarini J, Karlsson A, Vandamme AM, Perez-Perez MJ, Zhang H, Vrang L, Oberg B, Backbro K, Unga T and San-Felix A. Human immunodeficiency virus type 1 (HIV-1) strains selected for resistance against the HIV-1-specific [2',5'-bis-O-(tert-butyldimethylsilyl)-3'-spiro-5'-(4'-amino-1',2'-oxathiole-2',2'-dioxide)]- β -D-pentofuranosyl (TSAO) nucleoside analogues retain sensitivity to HIV-1-specific nonnucleoside inhibitors. *Proceedings of the National Academy of Sciences, USA* 1993; 90(15):6952-6956.
- [59] Vandamme A-M, Schmit JC, Balzarini J, Van Laethem K, Witvrouw M, Hermans P, Sprecher S, Martinez-Picado J, Clotet B, Peetermans W, Desmyter J and De Clercq E. Presence of TSAO-resistant virus strains in non-experienced patients. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3-6 July, 1996. Abstract 47.
- [60] Balzarini J, Pelemans H, Perez-Perez MJ, San-Felix A, Camarasa MJ, De Clercq E and Karlsson A. Marked inhibitory activity of non-nucleoside reverse transcriptase inhibitors against human immunodeficiency virus type 1 when combined with (-)2',3'-dideoxy-3'-thiacytidine. *Molecular Pharmacology* 1996; 49(5):882-890.
- [61] Balzarini J, Jonckheere H, Harrison WA, Dao DC, Anne J, De Clercq E and Karlsson A. Oxathiin carboxanilide derivatives: a class of non-nucleoside HIV-1-specific reverse transcriptase inhibitors (NNRTIs) that are active against mutant HIV-1 strains resistant to other NNRTIs. *Antiviral Chemistry and Chemotherapy* 1995; 6:169-178.
- [62] Schmit JC, Vanderlinden I, Ruiz L, Clotet B, Hermans P, Sprecher S, Arendt V, Peetermans W, Harrer T, Vaira D, Desmyter J, De Clercq E and Vandamme AM. Prevalence of multi-drug resistance to dideoxynucleoside (ddN) analogues in patients on ddN combination therapy. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3-6 July, 1996. Abstract 39.
- [63] Vandamme A-M, Debyser Z, Pauwels R, De Vreese K, Goubau P, Youle M, Gazzard B, Stoffels PA, Cauwenbergh GF, Anne J, Andries K, Janssen PAJ, Desmyter J and De Clercq E. Characterization of HIV-1 strains isolated from patients treated with TBO R82913. *AIDS Research and Human Retroviruses* 1994; 10(1):39-46.
- [64] de Bethune M-P, Pauwels R, Andries K, Vandamme AM, Peeters M, Colebunders R, Stoffels P, De Clercq E and Desmyter J. AZT resistance reversal by the non-nucleoside reverse transcriptase inhibitor α -APA R18893 in a symptomatic HIV-infected individual. Second HIV Drug Resistance Workshop, Noordwijk, The Netherlands, 3-5 June, 1993. Abstract.
- [65] Maass G, Immendoerfer U, Koenig B, Leser U, Mueller B, Goody R and Pfaff E. Viral resistance to the thiazolo-iso-indolinones, a new class of nonnucleoside inhibitors of human immunodeficiency virus type 1 reverse transcriptase. *Antimicrobial Agents and Chemotherapy* 1993; 37(12):2612-2617.
- [66] Nguyen MH, Schinazi RF, Shi C, Goudgaon NM, McKenna PM and Mellors JW. Resistance of human immunodeficiency virus type 1 to acyclic 6-phenylselenenyl- and 6-phenylthiopyrimidines. *Antimicrobial Agents and Chemotherapy* 1994; 38(10):2409-2414.
- [67] Staszewski S, Miller V, Rehmet S, Stark T, De Cree J, De Brabander M, Peeters M, Andries K, Moeremans M, De Raeymaeker M, Pearce G, Van Den Broeck RM, Verbiest W and Stoffels P. Virological and immunological analysis of a triple combination pilot study with loviride, lamivudine and zidovudine in HIV-1-infected patients. *AIDS* 1996; 10(5):F1-F7.

References

- [68] Richman D, Shih CK, Lowy I, Rose J, Prodanovich P, Goff S and Griffin J. Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture. *Proceedings of the National Academy of Sciences, USA* 1991; 88(24):11241-11245.
- [69] Mellors JW, Dutchman GE, Im GJ, Tramontano E, Winkler SR and Cheng YC. In vitro selection and molecular characterization of human immunodeficiency virus-1 resistant to non-nucleoside inhibitors of reverse transcriptase. *Molecular Pharmacology* 1992; 41(3):446-451.
- [70] Buckheit RW Jr, Fliakas-Boltz V, Yeagy-Bargo S, Weislow O, Mayers DL, Boyer PL, Hughes SH, Pan BC, Chu SH and Bader JP. Resistance to 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine derivatives is generated by mutations at multiple sites in the HIV-1 reverse transcriptase. *Virology* 1995; 210(1):186-193.
- [71] Kinjerski TL, Pallansch LA and Buckheit RW Jr. Isolation and characterization of HIV-1 isolates resistant to oxathiin carboxanilide derivatives: Evaluation of variables in the selection process. *Antiviral Chemistry and Chemotherapy* 1996; (in press).
- [72] Yang SS, Pattabiraman N, Gussio R, Pallansch L, Buckheit RW Jr and Bader JP. Cross-resistance analysis and molecular modelling of non-nucleoside reverse transcriptase inhibitors targeting drug-resistance Leukaemia 1996; (in press).
- [73] Hara H, Fujihashi T, Sakata T, Kaji A, Kaji H. Tetrahydronaphthalene lignan compounds as potent anti-HIV type 1 agents. *AIDS Research Human Retroviruses* 1997; 13(8):695-705.
- [74] Balzarini J, Karlsson A, Sardana VV, Emmini EA, Camarasa MJ and De Clercq E. Human immunodeficiency virus 1 (HIV-1)-specific reverse transcriptase (RT) inhibitors may suppress the replication of specific drug-resistant (E138K)RT HIV-1 mutants or select for highly resistant (Y181C to C181D)RT HIV-1 mutants. *Proceedings of the National Academy of Sciences, USA* 1994; 91(14):6599-6603.
- [75] Shaw G, Wei X, Johnson V, Taylor M, Decker J, Kilby M, Lifson J, Hahn B and Saag M. Nucleotide sequence analysis of HIV-1 RNA and DNA from plasma and PBMCs of patients treated with ZDV, ddI and nevirapine: rapid turnover and resistance development in vivo. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2-5 August, 1994. Abstract 71.
- [76] Schinazi RF, Lloyd RM Jr, Nguyen M-H, Cannon DL, McMillan A, Ilksoy N, Chu CK, Liotta DC, Bazmi HZ and Mellors JW. Characterization of human immunodeficiency viruses resistant to oxathiolane-cytosine nucleosides. *Antimicrobial Agents and Chemotherapy* 1993; 37(4):875-881.
- [77] Tisdale M, Kemp SD, Parry NR and Larder BA. Rapid in vitro selection of human immunodeficiency virus type 1 resistant to 3'-thiacytidine inhibitors due to a mutation in the YMDD region of reverse transcriptase. *Proceedings of the National Academy of Sciences, USA* 1993; 90:5653-5656.
- [78] Gao Q, Gu Z, Parniak MA, Cameron J, Cammack N, Boucher C and Wainberg MA. The same mutation that encodes low-level human immunodeficiency virus type 1 resistance to 2',3'-dideoxyinosine and 2',3'-dideoxycytidine confers high-level resistance to the (-) enantiomer of 2',3'-dideoxy-3'-thiacytidine. *Antimicrobial Agents and Chemotherapy* 1993; 37(6):1390-1392.
- [79] Larder BA, Kemp SD and Harrigan PR. Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy. *Science* 1995; 269:696-699.
- [80] Keulen W, van Wijk A, Boucher C and Berkhout B. Initial appearance of 184Ile variant in 3TC-treated patients can be explained by the mutation bias of the HIV-1 RT enzyme. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3-6 July, 1996. Abstract 95.
- [81] Gu Z, Gao H, Li X, Parniak MA and Wainberg MA. Novel mutation in the human immunodeficiency virus type 1 reverse transcriptase gene that encodes cross-resistance to 2',3'-dideoxyinosine and 2',3'-dideoxycytidine. *Journal of Virology* 1992; 66(12):7128-7135.
- [82] Schinazi RF, Lloyd RM Jr, McMillan A, Gosselin G, Imbach JL and Sommadossi J-P. Development of HIV-1 and SIV resistant to β -L-2',3'-dideoxycytidine analogues. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6-9 July, 1995. Abstract 10.

- [83] Moeremans M, De Raeymaeker M, Van den Broeck R, Stoffels P and Andries K. Genotypic analysis of HIV-1 isolates from patients receiving loviride alone or in combination with nucleoside reverse transcriptase inhibitor. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6-9 July, 1995. Abstract 34.
- [84] Olmsted RA, Slade DE, Kopta LA, Poppe SM, Poel TJ, Newport SW, Rank KB, Biles C, Morge RA, Deuweke TJ, Yagi Y, Romero DL, Thomas RC, Sharma SK, Tarpley WG. (Alkylamino)piperidine bis(heteroaryl)piperazine analogs are potent, broad-spectrum nonnucleoside reverse transcriptase inhibitors of drug-resistant isolates of human immunodeficiency virus type 1 (HIV-1) and select for drug-resistant variants of HIV-1_{111B} with reduced replication phenotypes. *Journal of Virology* 1996; 70(6):3698-3705.
- [85] Klein JP, Bender R, Kirsch R, Meichsner C, Paessens A, Rosner M, Rubsamen-Waigmann H, Kaiser R, Wichers M, Schneeweis KE, Winkler I and Riess G. Preclinical evaluation of HBV 097, a new nonnucleoside reverse transcriptase inhibitor of human immunodeficiency virus type 1 replication. *Antimicrobial Agents and Chemotherapy* 1995; 39(10):2253-2257.
- [86] Klein J-P, Bender R, Billhardt UM, Meichsner C, Riess G, Rosner M, Winkler I and Paessens A. Activity of a novel quinoxaline derivative against human immunodeficiency virus type 1 reverse transcriptase and viral replication. *Antimicrobial Agents and Chemotherapy* 1993; 37(8):1659-1664.
- [87] Kemp SD, Shi C, Bloor S, Harrigan PR, Mellors JW, Larder BA. A novel polymorphism at codon 333 of human immunodeficiency virus type 1 reverse transcriptase can facilitate dual resistance to AZT and 3TC. *J Virol* 1998; In Press.
- [88] Gurusinge AD, Land SA, Birch C, McGavin C, Hooker DJ, Tachedjian G, Doherty R and Deacon NJ. Reverse transcriptase mutations in sequential HIV-1 isolates in a patient with AIDS. *Journal of Medical Virology* 1995; 46(3):238-243.
- [89] Harrigan PR, Kinghorn I, Bloor S, Kemp SD, Najera I, Kohli A, and Larder BA. Significance of amino acid variation at human immunodeficiency virus type 1 reverse transcriptase residue 210 for zidovudine susceptibility. *Journal of Virology* 1996; 70(9): 5930-5934.
- [90] Hooker DJ, Tachedjian G, Solomon AE, Gurusinge AD, Land S, Anderson JL, Roy BM, Arnold E, Deacon NJ. An in vivo mutation from leucine to tryptophan at position 210 in human immunodeficiency virus type 1 reverse transcriptase contributes to high-level resistance to 3'-azido-3'-deoxythymidine. *Journal of Virology* 1996; 71(11):8010-8018.
- [91] Stuyver L, Wyseur A, Rombout A, Louwagie J, Scarcez T, Verhofstede C, Rimland D, Schinazi RF, Rossau R. Line probe assay for rapid detection of drug-selected mutations in the human immunodeficiency virus type 1 reverse transcriptase gene. *Antimicrob Agents Chemother* 1997; 41(2):284-291.
- [92] Slade DE, Vavro CL, Stapelton JT, Swack N and St Clair MH. A cysteine at codon 215 of HIV RT confers resistance to ddC. Second HIV Drug Resistance Workshop, Noordwijk, The Netherlands, 3-5 June, 1993. Abstract.
- [93] Dueweke TJ, Pushkarskaya T, Poppe SM, Swaney SM, Zhao Q, Chen SY, Stevenson M and Tarpley WG. A mutation in reverse transcriptase of bis(heteroaryl)piperazine-resistant human immunodeficiency virus type 1 that confers increased sensitivity to other nonnucleoside inhibitors. *Proceedings of the National Academy of Sciences, USA* 1993; 90(10):4713-4717.
- [94] Ho DD, Toyoshima T, Mo H, Kempf DJ, Norbeck D, Chen CM, Wideburg NE, Burt SK, Erickson JW and Singh MK. Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C2-symmetric protease inhibitor. *Journal of Virology* 1994; 68(3):2016-2020.
- [95] Tisdale M, Myers R, Parry NR, Oliver N, Machera B and Blair E. Comprehensive analysis of HIV-1 variants individually selected for resistance to six HIV protease inhibitors. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2-5 August, 1994. Abstract 14.

References

- [96] Kaplan AH, Michael SF, Wehbie RS, Knigge MF, Paul DA, Everitt L, Kempf DJ, Norbeck DW, Erickson JW and Swanstrom R. Selection of multiple human immunodeficiency virus type 1 variants that encode viral proteases with decreased sensitivity to an inhibitor of the viral protease. *Proceedings of the National Academy of Sciences, USA* 1994; 91:5597-5601.
- [97] Otto MJ, Reid CD, King RW, Garber S, Baker DB, Anton E and Winslow DL. Exposure of chronically infected PBMCs to DMP 450 can completely suppress virus replication or select resistant variants depending upon the dose of compound. Second National Conference on Human Retroviruses and Related Infections, Washington, DC, USA, 29 January-2 February, 1995. Abstract 464.
- [98] Winslow DL, Garber S, Reid C, Anton E and Otto MJ. DMP 450, a new cyclic urea inhibitor of HIV protease with potent in vitro antiviral activity. Eighth International Conference on Antiviral Research, Santa Fe, NM, USA, 23-28 April, 1995. Abstract 22.
- [99] Partaledis JA, Yamaguchi K and Byrn RA. In vitro selection and characterization of HIV-1 viral isolates with reduced sensitivity to inhibitors of HIV protease. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2-5 August, 1994. Abstract 8.
- [100] Tisdale M, Myers R, Najera I, Kohli A, Kemp S and Larder BA. Analysis of resistance interactions with 141W94 (VX-478) and other HIV-1 protease inhibitors. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3-6 July, 1996. Abstract 27.
- [101] King RW, Garber S, Winslow DL, Reid C, Bacheler LT, Anton E and Otto MJ. Multiple mutations in the human immunodeficiency virus protease gene are responsible for decreased susceptibility to protease inhibitors. *Antiviral Chemistry and Chemotherapy* 1995; 6(2):80-88.
- [102] Potts KE, Smidt ML, Stallings WC, Clare M, Pillay D, Richman DD and Bryant ML. In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) variants with decreased sensitivity to hydroxyethylurea isostere containing protease inhibitors. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2-5 August, 1994. Abstract 4.
- [103] Pillay D, Smidt ML, Potts KE, Bryant ML and Richman DD. In vitro selection of protease inhibitors resistant human immunodeficiency virus type 1 (HIV-1) strains. 34th Interscience Conference on Antimicrobial Agents and Chemotherapy, Orlando, FL, USA, 2-5 October, 1996. Abstract 7.
- [104] Smidt ML, Potts KE, Tucker SP, Blystone L, Stiebel TR Jr, Stallings WC, McDonald JJ, Pillay D, Richman DD, Bryant ML. A mutation in human immunodeficiency virus type 1 protease at position 88, located outside the active site, confers resistance to the hydroxyethylurea inhibitor SC-55389A. *Antimicrob Agents Chemother* 1997; 41(3):515-522.
- [105] Condra JH, Holder DJ, Schleif WA, Blahy OM, Danovich RM, Gabryelski LJ, Graham DJ, Laird D, Quintero JC, Rhodes A, Robbins HL, Roth E, Shivaprakash M, Yang T, Chodakewitz JA, Deutsch PJ, Leavitt RY, Massari FE, Mellors JW, Squires KE, Steigbigel RT, Teppler H and Emini EA. Genetic correlates of in vivo viral resistance to the HIV-1 protease indinavir. *Journal of Virology* 1996; 70(12): 8270-8276.
- [106] Croteau G, Doyon L, Thibeault D, McKercher G, Pilote L, Lamarre D. Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. *Journal of Virology* 1997; 71(2):1089-1096.
- [107] Schapiro JM, Winters MA, Viera M, Jacobsen H, Mous J and Merigan TC. Resistance mutations in patients receiving saquinavir: simultaneous appearance in lymph nodes, peripheral blood mononuclears (PBM) and plasma. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3-6 July, 1996. Abstract 28.
- [108] Molla A, Korneyeva M, Gao Q, Vasavanonda S, Schipper PJ, Mo HM, Markowitz M, Chernyavskiy T, Niu P, Lyons N, Hsu A, Granneman R, Ho DD, Boucher CAB, Leonard JM, Norbeck DW and Kempf DJ. Ordered accumulation for mutation in HIV protease confers resistance to ritonavir. *Nature Medicine* 1996; 2(7):760-766.

- [109] Doyon I, Croteau G, Thibeault D, Poulin F, Pilote L and Lamarre D. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *Journal of Virology* 1996; 70:3763-3769.
- [110] Patick AK, Mo H, Markowitz M, Appelt K, Wu B, Musick L, Kalish V, Kaldor S, Reich S, Ho D and Webber S. Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrobial Agents and Chemotherapy* 1996; 40(2):292-297; 40(6):1575 (erratum).
- [111] Patick AK, Kuritzkes D, Johnson VA, Shugarts D, Bakhtiari M, Potts KE, Farnsworth A, Anderson R, Koel JL, Hazelwood JD, Nail CD, Duran M, Markowitz M, Ho D, Richman D. Genotypic and phenotypic analyses of HIV-1 variants isolated from patients treated with nelfinavir and other HIV-1 protease inhibitors. *6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA*: Abstract 18.
- [112] Lamarre D, Croteau G, Pilote L, Rousseau P and Doyon L. Molecular characterization of HIV-1 variants resistant to specific viral protease inhibitors. *Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2-5 August, 1994*. Abstract 10.
- [113] Lamarre D, Doyon L, Croteau G, Pilote L and Thibeault D. Molecular basis of HIV-1 resistance to protease inhibitors Structural flexibility of the protease and second-site compensatory mutations in cleavage sites. *Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6-9 July, 1995*. Abstract 62.
- [114] Gulnik SV, Suvorov LI, Liu B, Yu B, Anderson B, Mitsuya H and Erickson JW. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* 1995; 34(29):9282-9287.
- [115] Condra JH and Schleif WA. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* 1995; 374:569-571.
- [116] Borman AM, Paulous S and Clavel F. Continued accumulation of protease inhibitor resistance mutations in culture in the absence of the drug. *Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6-9 July, 1995*. Abstract 93.
- [117] Jacobsen H, Brun-Vezinet F, Duncan I, Hanggi M, Ott M, Vella S, Weber J and Mous J. Genotypic characterization of HIV-1 from patients after prolonged treatment with proteinase inhibitor saquinavir. *Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2-5 August, 1994*. Abstract 16.
- [118] Eberle J, Bechowsky B, Rose D, Hauser U, Von Der Helm K, Gurtler L and Nitschki H. Resistance of HIV type 1 to proteinase inhibitor Ro 31-8959. *AIDS Research and Human Retroviruses* 1995; 11(6):671-676.
- [119] Mo H, Markowitz M, Majer P, Burt SK, Gulnik SV, Suvorov LI, Erickson JW, Ho DD. Design, synthesis, and resistance patterns of MP-134 and MP-167, two novel inhibitors of HIV type 1 protease. *AIDS Research and Human Retroviruses* 1996; 12(1):55-61.
- [120] Vasudevachari MB, Zhang Y-M, Imamichi H, Imamich T, Falloon J, Salzman NP. Emergence of protease inhibitor resistance mutations in human immunodeficiency virus type 1 isolates from patients and rapid screening procedure for their detection. *Antimicrob Agents Chemother* 1996; 40(11):2535-2541.
- [121] Rao BG, Dwyer MD, Thomson JA, Baker CT, Deininger DD, Murcko MA, Tung RD, Navia MA and Kim EE. Structural and modelling analysis of the basis of viral resistance to VX-478. *Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3-6 July, 1996*. Abstract 22.
- [122] Patick AK, Rose R, Greytak J, Bechtol CM, Hermsmeier MA, Chen PT, Barrish JC, Zahler R, Colonno RJ and Lin PE. Characterization of a human immunodeficiency virus type 1 variant with reduced sensitivity to an aminodiol protease inhibitor. *Journal of Virology* 1995; 69(4):2148-2152.

References

- [123] Rose B, Greytak J, Bechtold C, Alam M, Terry B, Gong YF, DeVore K, Patrick A, Colono R and Lin PF. Combination therapy with two protease inhibitors as an approach to antiviral therapy. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2-5 August, 1994. Abstract 17.
- [124] Dulioust A, Paulous S, Guillemot L, Boue F, Galanaud P, Clavel F. Selection of saquinavir-resistant mutants by indinavir following a switch from saquinavir. *6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA*: Abstract 16.
- [125] Swanstrom R, Smith T, Petit S, Irlbeck D, Shao W, Wehbie R, Sawhney R, Everitt L and Erickson L. Multiple sequence changes within HIV-1 protease confer reduced sensitivity to a symmetric protease inhibitor. Third International Workshop on HIV Drug Resistance, Kauai, Hawaii, USA, 2-5 August, 1994. Abstract 6.
- [126] Otto MJ, Garber S, Winslow DL, Reid CD, Aldrich P, Jadhav PK, Patterson CB, Hodge CN and Cheng YS. In vitro isolation and identification of human immunodeficiency virus (HIV) variants with reduced sensitivity to C-2 symmetrical inhibitors of HIV type 1 protease. *Proceedings of the National Academy of Sciences, USA* 1993; 90(16):7543-7.
- [127] Winters MA, Schapiro JM, Lawrence J, Merigan TC. Genotypic and phenotypic analysis of the protease gene in HIV-1-infected patients that failed long-term saquinavir therapy and switched to other protease inhibitors. *6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA*: Abstract 17.
- [128] Eastman PS, Duncan IB, Gee C, Race E. Acquisition of genotypic mutations associated with reduced susceptibility to protease inhibitors during saquinavir monotherapy. *6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA*: Abstract 30.
- [129] Schapiro JM, Winters M, Lawrence J, Norris J, Merigan TC. Clinical and genotypic cross-resistance between the protease inhibitors saquinavir and indinavir. *6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA*: Abstract 87.
- [130] Shao W, Smith T and Swanstrom R. Selection and analysis of HIV-1 variants with increased resistance to SKF108842 and SKF108922, two protease inhibitors. Fourth International Workshop on HIV Drug Resistance, Sardinia, Italy, 6-9 July, 1995. Abstract 65.
- [131] el-Farrash MA, Kuroda MJ, Kitazaki T, Masuda T, Kato K, Hatanaka M and Harada S. Generation and characterization of a human immunodeficiency virus type 1 (HIV-1) mutant resistant to an HIV-1 protease inhibitor. *Journal of Virology* 1994; 68(1):233-9.
- [132] Labrosse B, Pleskoff O, Sol N, Jones C, Henin Y, Alizon M. Antiviral and resistance studies of RPR103611, an inhibitor of HIV replication. *6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA*: Abstract 33.
- [133] Este JA, Schols D, De Vreese K, Van Laethem K, Vandamme AM, Desmyter J and De Clercq E. Development of resistance of human immunodeficiency virus type 1 to dextran sulfate associated with the emergence of specific mutations in the envelope gp120 glycoprotein. *Molecular Pharmacology* 1997; 52(1): 98-104.
- [134] Este JA, Van Laethem K, Vandamme AM, Desmyter J and De Clercq E. Resistant phenotype of human immunodeficiency virus type 1 to dextran sulfate is conferred by specific amino acid substitutions in the gp120 molecule. Fifth International Workshop on HIV Drug Resistance, Whistler, Canada, 3-6 July, 1996. Abstract 80.
- [135] De Vreese K, Reyman D, Griffin P, Steinkasserer A, Werner G, Bridger GJ, Este J, James W, Henson GW, Desmyter J, Anne J and De Clercq E. The bicyclams, a new class of potent human immunodeficiency virus inhibitors, block viral entry after binding. *Antiviral Research* 1996; 29:209-219.
- [136] De Vreese K, Kofler-Mongold V, Leutgeb C, Weber V, Vermeire K, Schacht S, Anne J, De Clercq E, Datema R and Werner G. The molecular target of bicyclams, potent inhibitors of human immunodeficiency virus replication. *Journal of Virology* 1996; 70(2):689-696.

- [137] Lin P, Samanta H, Bechtold CM, Deminie CA, Patick AK, Alam M, Riccardi K, Rose RE, White RJ and Colonna RJ. Characterization of siamycin 1, a human immunodeficiency virus fusion inhibitor. *Antimicrobial Agents and Chemotherapy* 1995; 40:133-138.
- [138] Van Rompay KKA, Cherrington JM, Marthas ML, Barardi CJ, Mulato AS, Spinner A, Tarara RP, Canfield DR, Telm S, Bischofberger N, Pedersen NC. PMPA therapy of established SIV infection of infant rhesus macaques. *Antimicrob Agents Chemother* 1996; 40:2586-2591.
- [139] Cherrington JM, Van Rompay KKA, Mulato AS, Marthas ML, Barardi CJ, Telm S, Bischofberger N, Pedersen NC. Phenotypic and genotypic characterization of simian immunodeficiency viruses (SIV) with reduced susceptibility to PMPA isolated after PMPA therapy. *Fifth International Workshop on HIV Drug Resistance, 1996, Whistler, Canada: Abstract 75.*
- [140] Van Rompay K, Cherrington J, Marthas M, Agatep E, Dehqanzada Z, Lamy P, Barardi C, Bischofberger N, Pedersen N. Therapeutic efficacy of PMPA treatment for infant macaques infected with PMPA-resistant simian immunodeficiency virus. *6th International Workshop on HIV Drug Resistance, 25-28 June 1997; St. Petersburg, USA: Abstract 117.*
- [141] Van Rompay KKA, Breenier JL, Marthas ML, Otsyula MG, Tarar RP, Miller CJ, Pedersen NC. A zidovudine resistant simian immunodeficiency virus mutant with a Q151M mutation in reverse transcriptase causes AIDS in newborn macaques. *Antimicrob Agents Chemother* 1997; 41:278-283.
- [142] Medlin HK, Zhu Y, Remington KM, Phillips TR, North TW. Selection and characterization of a mutant of feline immunodeficiency virus resistant to 2',3'-dideoxycytidine. *Antimicrob Agents Chemother* 1996; 40:953-957.
- [143] Zhu Y-Q, Remington KM, North TW. Mutants of feline immunodeficiency resistance to 2',3'-dideoxy-2',3'-didehydrothymidine. *Antimicrob Agents Chemother* 1996; 40:1983-1987.
- [144] Smith RA, Remington KM, Lloyd RM Jr., Schinazi RF, North TW. A novel point mutation at position 156 of reverse transcriptase from feline immunodeficiency virus confers resistance to the combination of (-)- β -L-2',3'-dideoxy-3'-thiacytidine and 3'-azido-3'-deoxythymidine. *J Virol* 1997; submitted for publication.
- [145] Smith R, Remington K, Lloyd R, Schinazi R, North T. Mutants of feline immunodeficiency virus resistant to FTC and 3TC. *Third International Feline Retrovirus Research Symposium, 1996, Ft. Collins, CO: Abstract 41.*
- [146] Smith RA, Remington K, Lloyd RM Jr, Schinazi RF, North TW. A novel met to thr mutation in the YMDD motif of reverse transcriptase from feline immunodeficiency virus confers resistance to oxathiolane nucleosides. *J Virol* 1997; 71:2357-2362.

Drug Resistance Mutations Superimposed on the Structures of HIV-1 Protease and Reverse Transcriptase

Holly Tao

*Theoretical Biology and Biophysics (T10), Theoretical Division, Los Alamos National Laboratory
Los Alamos, New Mexico 87545*

The 3-D structure of HIV-1 protease and reverse transcriptase and the mutable positions that confer drug resistance, as summarized by Mellors, et al., [1] are presented here in a companion piece to the Hammond et al. drug resistance table for 1997 (pages xx-xx, this volume) [2]. These figures are intended to provide a quick reference indicating the location of resistance mutations on the folded protein, and facilitate the comparison of different active compounds.

The figure depicting the protease protein was based on the structure found in the Protein Data Bank (PDB) database (<http://molbio.info.nih.gov/cgi-bin/molDraw?1HVK>) under accession number 1hvk. This crystal structure was originally determined by Hosur, et al., [3], and is based on an X-ray crystal structure of HIV-1 protease complexed with the inhibitor A76928 (S,S) at a resolution of 1.8 Å. A total of 26 mutations found to confer resistance to any of the compounds directed against protease as summarized by Mellors et al., [1], are indicated on the folded structure. Single mutations at positions 10, 20, 24, 32, 36, 46, 47, 48, 50, 54, 63, 71, 82, 84, 88, 90 conferred resistance to more than one drug. Only one monomer is labeled in the figure, as they are identical in the dimer. The active sites of HIV-1 protease are also indicated; in particular Asp 25, in a deep cleft formed in the protease dimer, is shown in red [4].

The figure depicting the HIV-1 reverse transcriptase (RT) protein was based on the structure found in PDB (<http://molbio.info.nih.gov/cgi-bin/molDraw?1HNV>) under accession number 1hmv. This crystal structure was originally determined by Ding et al., [5], and is based on an X-ray crystal structure of HIV-1 RT complexed with the inhibitor R 86183 at a resolution of 3.0 Å. A total of 35 mutations found to confer resistance to any of the compounds directed against RT [1] are indicated on the folded structure. Single mutations at positions 65, 70, 74, 75, 88, 89, 98, 100, 101, 103, 106, 108, 138, 179, 181, 184, 188, 190, 215, 219, 236, conferred resistance to more than one drug. Two subunits of HIV-1 RT are labeled in the figure, as they aren't identical in the dimer. The active sites of HIV-1 RT are also indicated; in particular Asp 110, Asp 186, and Asp 186, in a deep cleft formed in the p66 POL, are shown in red [6].

References

- [1] Mellors, J.W.; Schinazi, R.F.; Larder, B.A.; Mutations in Retroviral Genes Associated with Drug Resistance, in Myers, et al. Eds., *Human Retroviruses and AIDS 1996*, III-206, Los Alamos National Laboratory, Los Alamos, NM.
- [2] Hammond, J.; Larder, B.A.; Schinazi, R.F.; Mellors, J.W., Mutations in Retroviral Genes Associated with Drug Resistance, in Korber, et al. Eds., *Human Retroviruses and AIDS 1997*, III-xx, Los Alamos National Laboratory, Los Alamos, NM.
- [3] Hosur, M.V.; Bhat, T.N.; Kempf, D.; Baldwin, E.T.; Liu, B.; Gulnik, S.; Widebury, N.E.; Norbeck, D.W.; Appelt, K.; Erickson, J.W., *J. Am. Chem. Soc.*, (1994) 116:847.
- [4] Meek, T.D.; Dayton, B.D.; Metcalf, B.W.; Dreyer, G.B.; Strickler, J.E.; Gorniak, J.G.; Rosenberg, M.; Moore, L.; Magaard, V.W.; Bebouck, C., *Proc. Natl. Acad. Sci. USA*, (1989) 86:1841.
- [5] Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P.A.; Hughes, S.H.; Arnold, E., *Nat. Struct. Biol.*, (1995) 2:407.
- [6] Kohlstaedt, L.A.; Wang, J.; Friedman, J.M.; Rice, P.A.; Steitz, T.A., *Science*, (1992) 256:1783.

The mutations in HIV-1 protease that confer drug resistance are indicated in the ball-and-stick representations, helices in green, strands in pink and turn and coil in yellow.

Image courtesy of Dr. David A. Asch and Dr. Robert W. Johnson, University of Illinois.

IV

Related Sequences

The following table is a summary of cellular proteins known or hypothesized to interact with or be affected by HIV and related viruses. We thank Dr. Kuan-Teh Jeang, NIAID, for his thoughtful review of this table and we apologize to those whose work or original publications were not cited because of space and time limitations.

Name	Description	Reference No.
AP-1	Transcription factor that appears to be involved in HIV gene expression; binding sites in LTR U3	1,2
ATN/TFIIS	Transcription elongation factor that acts synergistically with Tat	3
β -COP	Protein from non-clathrin coated vesicles reported to interact with Nef	1,4
B23	Cell protein that binds the nuclear-localizing domain of Rev	1,5
Bel-1	Human foamy virus protein that can transactivate HIV LTR	6
c-Ets-1	Transcription factor said to be indispensable for HIV-1 enhancer function; see also Elf-1, LBP-1	7
c-Fos	Transcription factor that activates HIV provirus expression; see also NFAT and c-Jun	1,8
c-Jun	Transcription factor that activates HIV provirus expression; see also NFAT and c-Fos	1,8
c-myc	Transactivator of HIV through binding to high-affinity site in LTR	9
c-rel	Strong activator of basal HIV gene expression	10
CCR5	Regarded to be the major coreceptor for macrophage-tropic viruses	10a,10b
CD4	Primary receptor which interacts with and is sequestered by Nef	1,11
CKII	Casein kinase II, apparently inhibitory through phosphorylation of cell proteins involved in HIV transactivation	12
CXCR4	Regarded to be the major coreceptor for T-cell-tropic viruses	13a
COUP	Steroid/thyroid receptor family protein that interacts with LTR U3 (may compete with AP-1 for binding)	1,13
CyPA	Gag p24 binds cyclophilin A; cyclophilin B is also involved with Gag	1,14
DNA-PK	Binds Tat and possibly increases thereby the phosphorylation of SP-1	14a
E1A	Activator and repressor of HIV transcription	15
E2F-1	Cell cycle regulating protein that suppresses HIV LTR-directed transcription	16
EF-1A, SRB, PTB	Possible cofactors of Tat and TRP-185	17
eIF-5A	Initiation factor reported to be a cofactor for Rev	1,18

Cellular Proteins

Elf-1	Bts-related transcription factor required for HIV-2 transactivation; see also c-Bts-1, LBP-1	19
FBI-1	Protein that binds LTR elements and appears to be involved in abortive transcription complexes	20
FLK-1/KDR	Hypothesized to mediate Tat-induced angiogenesis in vascular endothelial cells	20a
GATA-3	Cellular zinc-finger transcriptional factor with binding sites in the LTR U3 (transcriptional activator of HIV)	21
GM-CSF	Granulocyte/macrophage colony stimulating factor suppresses HIV replication	22
HHR23A,D21235	Nucleotide excision repair protein that interacts with Vpr	23
HIP116	Transcription factor related to SNF2/SWI2 that binds HIV promoter; may act as ATPase	24
HIVEN86A	Tax-inducible protein that becomes an HIV enhancer-binding protein	25
hnRNP-49	RNP fraction (49 kDa) that binds RRE	1,26
HuChr12	Cellular protein (83 kDa) encoded by human chromosome 12 that targets TAR RNA	1,27
IE1, IE2	HCMV phosphoproteins that transactivate HIV synergistically with Tat	28
II10	Inhibitor of HIV replication in maturing monocytes; enhances viral production in the presence of TNF- α	29
ILF	Interleukin binding factor that binds to HIV LTR and also interleukin 2 promoter	30
integrin	Alpha V Beta 5 integrin proposed as receptor for Tat	1,31
IRF	Interferon regulatory factor (family) that may bind to HIV	32
Ku	Lupus autoantigen binds TAR RNA with high affinity; Ku normally involved in transcriptional elongation and activation.	1,33
La	La autoantigen binds to TAR RNA and alleviates translational repression	1,34
lamin B	Member of a complex of nuclear envelope proteins that is reported to interact with Rev and Tat	35
LBP-1	DNA binding protein that can repress HIV transcription when bound to a site which overlaps TATA (inhibition of TFIID implied); see also Elf-1.	1,36
LEF-1	T-cell specific transcription factor that binds LTR U3	1
MBP1	Inhibits core promoter of HIV-1 LTR	37
MIP-1	Alpha and beta forms of this CC-beta chemokine inhibit infection by macrophage-tropic viruses	37a
MSS1	Cell protein similar (42%) to TBP-1 that binds to and modulates/suppresses Tat	1,38
NAK	A cellular serine kinase that associates with Nef	39
NF-IL6	LTR-mediated activator of HIV gene expression	40
NFAT	Nuclear factor from activated T-cells that interacts with LTR U3; see also Fos and Jun	1,41
NFkB	Primary HIV transcription factor activated by Tat	1,42

NFRRE	Nuclear factor (56 kDa) that binds RRE (Rev responsive element)	1,43
NRF	See USF43	1
Oct-2	Cellular transcription factor that interacts with and enhances Tat activity	44
p140	HeLa nuclear protein that binds specifically to TAR RNA	45
p32	Human homolog of the murine protein YL2 that interacts with Rev	1,46
p36	36 kDa protein that binds Tat and enhances transactivation	1,47
p56lck,Hck	Kinases which appear to be involved with Nef function; see also PAK	1,48
p68	Cellular protein that appears to modulate TAR activity	1,49
PAF	Platelet-activating factor, possible mediator of HIV gene expression	50
PAK	Member of the p21-activated kinase family that interacts with Nef; see also p56lck	1,51
pets factor	Cellular protein that binds the 'pets' site in the HIV-2 enhancer	52
PITALRE	Human transcription elongation factor; P-TEFb, its homolog in <i>Drosophila</i> , appears to be required for Tat transactivation <i>in vitro</i>	52a
PKC	Protein kinase C, inhibited by env gp41 peptide	53
PKR	RNA-activated kinase that binds to and phosphorylates Tat; also binds TAR RNA stem; formerly P1/dsI	1,54
PRDII-BF1	Transcription factor that binds to NFkB and Tar DNA elements and activates	1,55
prothymosin	Prothymosin alpha reported to bind Rev	1,56
Raf-1	Oncoprotein (serine/threonine kinase) that mediates HIV activation through binding to core enhancer sites	57
RANTES	CC-beta type chemokine that inhibits infection by macrophage-tropic viruses	37a
RBP9-27	RRE binding protein that inhibits REV-dependent posttranscriptional processing.	58
RIP	Vpr interactive protein (41 kDa) (not to be confused with RIP/RAB that interacts with Rev)	59
RIP/RAB	Human Rev-interacting protein, or Rev activation domain binding protein, a 58 kDa nucleoporin (not to be confused with RIP that interacts with Vpr)	1,60
RNA polII	Binds TAR RNA as well as HIV-1 and HIV-2 Tat proteins	1,61
SBP	TAR stem binding protein	1,62
SC-35	Nuclear ribonucleoprotein RNA processing factor found in complex with Tat in the nucleus.	63
Sp1	Transcription factor that binds to LTR; interacts with Vpr and Tat; cooperates with NFkB (Sp3 form can be negative)	1,64
Stat-50	Transcriptional activator that down-regulates HIV LTR expression.(IFN-regulated)	65
T3R	Thyroid hormone receptor interacts with Tat; under certain conditions binds to LTR Sp-1 motifs	1,66

Cellular Proteins

TAK/TFIIH	Tat associated kinase, a cofactor in Tat activation; hyperphosphorylates the CTD of RNA polII	1,67
TAP	Tat-associated protein, which appears to be a cofactor in Tat transactivation	68
TARBP-b	T-cell protein that binds to the bulge of TAR RNA	1,69
Tat-SF1	Cellular protein (140 kDa) that supports Tat-elongation of transcription; complex appears to also involve a kinase.	70
TBP	TATA binding protein that also binds to Tat and activates; see also TFIIID	1,71
TDP-43	TAR DNA binding protein that represses LTR-directed transcription	72
TFI	Transcription factor (39 kDa) related to USF that binds to LTR U3	73
TFE3	Cellular factor that binds sites also bound by USF (NRF)	1
TFIID, TFIIIB	Transcription factors which interact with Tat to support initiation; see also TBP	1,74
TFIIIC	Transcription factor activated by Tat	75
TGF- β	Cytokine (transforming growth factor beta) induced by gp160	76
Tip60	Tat interactive protein (60 kDa), may be a cofactor in Tat transactivation	1,77
TMF	TATA modulatory factor that inhibits LTR activation	78
TNF- α	Inhibits HIV-1 transcription; inhibits HIV entry into macrophages	1,79
TNF- β	Growth regulatory factor transactivated by Tat	80
TRBP	TAR binding protein, said to be inhibitor of PKR; also binds RRE	1,81
TRP1/TRP185	TAR-binding protein involved in regulation of HIV gene expression	1,82
TRP2	Cell protein that binds to TAR RNA bulge	1,82
ubiquitin	A complex between ubiquitin and Gag p6 has been reported	83
USF43	Transcription factor that recognizes HIV-1 promoter; also called NRF	1,84
YY1	Transcriptional factor that represses HIV replication	1,85

References

- [1] Jeang, Part IV, Human Retroviruses and AIDS 1996
- [2] Van Lint et al., J. Virol. 65:7066-7072 (1991)
- [3] Kato et al., Genes Dev. 6:655-666 (1992)
- [4] Benichou et al., J. Biol. Chem. 269:30073-30076 (1994)
- [5] Miyazaki et al., Biochimie 78:1081-1086 (1996); Fankhauser et al., Mol. Cell Biol. 11:2567-2575 (1991)
- [6] Keller et al., J. Virol. 66:3946-3949 (1992)
- [7] Holzmeister et al., Biochem. Biophys. Res. Commun. 197:1229-1233 (1993)

- [8] Roebuck et al., *AIDS* 10:819-826 (1996)
- [9] Dasgupta et al., *Proc. Nat. Acad. Sci. USA* 87:8090-8094 (1990)
- [10] Muchardt et al., *J. Virol.* 66:244-250 (1992)
- [10a] Alkhatib et al., *Science* 272:1955-1958 (1996)
- [10b] Dragic et al., *Nature* 381:667-673 (1996)
- [11] Rossi et al., *Virology* 217:397-403 (1996)
- [12] Critchfield et al., *Proc. Nat. Acad. Sci. USA* 94:6110-6115 (1997)
- [13] Cooney et al., *J. Virol.* 65:2853-2860 (1991)
- [13a] Feng et al., *Science* 272:872-877 (1996)
- [14] Luban et al., *Cell* 73:1067-1078 (1993); Ou and Henderson, Part III, *Human Retroviruses and AIDS* 1995.
- [14a] Chun et al., *J. Virol.* 72:2615-2629 (1998)
- [15] Parker et al., *J. Virol.* 71:2004-2012 (1997)
- [16] Kundu et al., *J. Virol.* 69:6940-6946 (1995)
- [17] Wu-Baer et al., *J. Biol. Chem.* 271:4201-4208 (1996)
- [18] Ruhl et al., *J. Cell Biol.* 123:1309-1320 (1993); Bevec et al., *PNAS* 91:10829-10833 (1994); Bevec et al., *Science* 271:1858-1860 (1996)
- [19] Leiden et al., *J. Virol.* 66:5890-5897 (1992)
- [20] Pessler et al., *Mol. Cell Biol.* 17:3786-3798 (1997)
- [20a] Albim et al., *Nat. Medicine*. 2:1371-1375 (1996)
- [21] Yang and Engel, *NAR* 21:2831-2836 (1993)
- [22] Matsuda et al., *ARHR* 11:1031-1038 (1995)
- [23] Withers-Ward et al., *J. Virol.* 71:9732-9742 (1997)
- [24] Sheridan et al., *J. Biol. Chem.* 270:4575-4587 (1995)
- [25] Bohnlein et al., *J. Virol.* 63:1578-1586 (1989)
- [26] Xu et al., *J. Biomed. Sci.* 3:82-91 (1996)
- [27] Alonso et al., *J. Virol.* 66:4617-4621 (1992); Hart et al., *J. Virol.* 69:6593-6599 (1995)
- [28] Dal Monte et al., *AIDS* 11:297-303 (1997)
- [29] Naif et al., *ARHR* 12:1237-1245 (1996); Finnegan et al., *J. Immunol.* 156:841-851 (1996)
- [30] Li et al., *Proc Nat Acad Sci USA* 88:7739-7743 (1991); Li et al., *Genomics* 13:665-671 (1992)
- [31] Vogel et al., *J. Cell Biol.* 121:461-468 (1993)
- [32] Thornton et al., *Proc. Natl. Acad. Sci. USA* 93:383-387 (1996)
- [33] Kaczmarek and Kahn, *BBRC* 196:935-942 (1993)
- [34] Svitkin et al., *J. Virol.* 68:7001-7007 (1994); Chang et al., *J. Virol.* 68:7008-7020 (1994)
- [35] Nikolakaki et al., *J. Biol. Chem.* 271:8365-8372 (1996)
- [36] Kato et al., *Science* 251:1476-1479 (1991); Yoon et al., *Mol. Cell Biol.* 14:1776-1785 (1994); Mitra et al., *Virology* 214:512-521 (1995); Parada et al., *J. Biol. Chem.* 270:2274-2283 (1995)

- [37] Ray and Srinivas, *J. Cell Biochem.* 64:565-572 (1997)
- [37a] Cocchi et al., *Science*. 270:1811-1815 (1995)
- [38] Shibuya et al., *Nature* 357:700-702 (1992); Dubiel et al., *FEBS Lett* 323:276-278 (1993)
- [39] Lu et al., *Curr. Biol.* 6:1677-1684 (1996)
- [40] Tesmer and Bina, *J. Mol. Biol.* 262:327-335 (1996)
- [41] McCaffrey et al., *Science* 262:750-754 (1993)
- [42] Nabel and Baltimore, *Nature* 326:711-713 (1987); Demarchi et al., *J. Virol.* 70:4427-4437 (1996)
- [43] Vaishnav et al., *New Biol.* 3:142-150 (1991)
- [44] Liu et al., *AIDS* 10:1323-1329 (1996)
- [45] Rothblum et al., *J. Virol.* 69:5156-5163 (1995)
- [46] Luo et al., *J. Virol.* 68:3850-3856 (1994); Tange et al., *J. Biol. Chem.* 271:10066-10072 (1996)
- [47] Desai et al., *Proc. Nat. Acad. Sci USA* 88:8875-8879 (1991)
- [48] Salghetti et al., *Proc. Nat. Acad. Sci USA* 92:349-353 (1995); Lee et al., *EMBO J.* 14:5006-5015 (1995); Bodeus et al., *J. Gen. Virol.* 76:1337-1344 (1995); Colette et al., *J. Biol. Chem.* 271:6333-6341 (1996); Saksela et al., *EMBO J.* 14:484-491 (1995)
- [49] Marciniak et al., *Proc. Nat. Acad. Sci. USA* 87:3624-3628 (1990)
- [50] Squinto et al., *J. Mol. Neurosci.* 2:79-84 (1990)
- [51] Nunn and Marsh, *J. Virol.* 70:6157-6161 (1996)
- [52] Fu and Markovitz, *J. Biol. Chem.* 271:19599-19605 (1996); Hilfinger et al., *J. Virol.* 67:4448-4453 (1993)
- [52a] Zhun et al., *Genes Dev.* 11:2622-2632 (1997)
- [53] Ward et al., *Cancer Lett.* 88:37-40 (1995)
- [54] Brand et al., *J. Biol. Chem.* 272:8388-8395 (1997); McMillan et al., *Virology* 213:413-424 (1995)
- [55] Seeler et al., *J. Virol.* 68:1002-1009 (1994)
- [56] Kubota et al., *Eur. J. Biochem.* 233:48-54 (1995)
- [57] Bruder et al., *NAR* 21:5229-5234 (1993)
- [58] Constantoulakis et al., *Science* 259:1314-1318 (1993)
- [59] Refaeli et al., *Proc. Nat. Acad. Sci USA* 92:3621-3625 (1995)
- [60] Fritz et al., *Nature* 376:530-533 (1995); Bogerd et al., *Cell* 82:485-494 (1995); Stutz et al., *Cell* 82:495-506 (1995)
- [61] Mavankal et al., *Proc. Natl. Acad. Sci. USA* 93:2089-2094 (1996)
- [62] Rounseville and Kumar, *J. Virol.* 66:1688-1694 (1992)
- [63] Berthold and Makedarelli, *J. Virol.* 70:4667-4682 (1996)
- [64] Wang et al., *J. Biol. Chem.* 255:64-25569 (1995); Majello et al., *NAR* 22:4914-4921 (1994)
- [65] Tissot and Mechti, *J. Biol. Chem.* 270:14891-14898 (1995)
- [66] Desai-Yajnik et al., *J. Virol.* 69:5103-5112 (1995)
- [67] Herrmann and Rice, *J. Virol.* 69:1612-1620 (1995); Garcia-Martinez et al., *EMBO J.* 16:2836-2850 (1997)

- [68] Yu et al., *J. Virol.* 69:3007-3016 (1995); Yu et al., *J. Virol.* 69:3017-3023 (1995)
- [69] Reddy et al., *ARHR* 11:663-669 (1995)
- [70] Zhou and Sharp, *Science* 274:605-610 (1996)
- [71] Veschambre et al., *J. Mol. Biol.* 250:169-180 (1995)
- [72] Ou et al., *J. Virol.* 69:3584-3596 (1995)
- [73] Maskawa et al., *Nucl. Acids Res.* 19:4689-4694 (1991)
- [74] Veschambre et al., *J. Gen. Virol.* 78:2235-2245 (1997); Kashanchi et al., *Nature* 367:295-299 (1994)
- [75] Jang et al., *JAIDS* 5:1142-1147 (1992)
- [76] Hu et al., *Clin. Immunol. Immunopathol.* 80:283-289 (1996)
- [77] Kamine et al., *Virology* 216:357-366 (1996)
- [78] Garcia et al., *Proc. Natl. Acad. Sci. USA* 89:9372-9376 (1992)
- [79] Herbein et al., *J. Virol.* 70:7388-7397 (1996)
- [80] Yang et al., *Virus Res.* 50:195-203 (1997); Buonaguro et al., *J. Virol.* 68:2677-2682 (1994)
- [81] Park et al., *Proc. Nat. Acad. Sci. USA* 91:4713-4717 (1994)
- [82] Wu et al., *Genes Dev.* 5:2128-2140 (1991); Wu-Baer et al., *EMBO J.* 14:5995-6009 (1995)
- [83] Ott and Henderson, Part III, *Human Retroviruses and AIDS* 1995.
- [84] Du et al., *EMBO J.* 12:501-511 (1993)
- [85] Margolis et al., *J. Virol.* 68:905-910 (1994)