21,838

COO-155--107

APPLICATION FOR RENEWAL OF SUPPORT

of

AEC CONTRACT AT(11-1)-1552

AREA PROGRAM IN POPULATION GENETICS

by

James V. Neel
Henry Gershowitz
Arthur D. Bloom
Charles F. Sing

MASTER

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes 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.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Department of Human Genetics, University of Michigan Medical School, Ann Arbor

April 17, 1973

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, makes 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 in electronic image products. Images are produced from the best available original document.

TABLE OF CONTENTS

I. SCIENTIFIC PERSONNEL, 1972-1973	2
II. PROJECT ABSTRACT	2
III. FOREIGN TRAVEL	3
IV. INTRODUCTION	3
V. PROGRESS REPORT	4
V.1. CONTINUING STUDIES OF THE AMERICAN INDIAN	4
V.1.1. Yanomama Demography (Neel, Weiss)	4
V.1.2. Validation of Dendrograms (Spielman, Neel)	6
V.1.3. A Comparison of Linguistic and Genetic Differentiation (Spielman, Migliazza)	8
V.1.4. Further Studies on Salt Balance in the Yanomama (Oliver, Migliazza)	10
V.1.5. Migration Matrices for a Tribal Population (Ward, Neel)	11
V.1.6. A Simulation of the Fate of a Mutant Gene of Neutral Selec- tive Value in a Primitive Population (Li)	12
V.1.7. An Indirect Estimate of Mutation Rates Employing Biochemical Variants (Neel)	15
V.1.8. Demographic Models for Primitive Human Populations (Weiss)	18
V.2. ORGANIZATION OF THE PROGRAM IN JAPAN (Neel)	19
V.3. SUPPORTING LABORATORY ACTIVITIES	22
V.3.1. The Activities of the Biochemical Screening Laboratory (Tanis, Ferrell)	22
V.3.2. The Activities of the Serological Screening Laboratory (Gershowitz)	23
V.4. MATHEMATICAL AND STATISTICAL ASPECTS OF POPULATION GENETICS	25
V.4.1. Estimation of Parameters which Measure Human Population Structure (Rothman, Templeton, Sing)	26
V.4.2. Development of Statistical Procedures to Estimate Parameters Which Define Selection Effects in Parthenogenetic Populations (Rothman, Templeton)	27
V.4.3. An Application of the Analysis of Variance of Allele Fre- quencies to a Human Population (Sing)	29
V.4.4. An Analysis of the Role Genes Play in Determining Choles- terol Variation in the Tecumseh, Michigan Population (Sing)	32
V.4.5. Linkage in the Tecumseh, Michigan Population (Sing)	35

TABLE OF CONTENTS, CONT'D.

V.4.6.	The Effect of Density-Dependent Population Regulation on the Age-Distribution of Single-Sex Populations (Smouse, Weiss)	35
V.4.7.	Computer Use (Griffith, Sing)	37
V.5.	SUPPORTING LABORATORY AND OTHER INVESTIGATIONS	40
V.5.1.	The Genetic Unit of Selection (Sing, Templeton)	40
V.5.2.	Genetic, Chemical and Serological Studies of Human Immunoglobulins (Gershowitz)	53
V.5.2.1.	Further Subdivision of the Inv System	53
V.5.2.2.	Idiotypic Specificity of an Anti-Rh Antibody	54
V.5.2.3.	Gm Antigens in Lymphocyte Culture Media (Kohler, Gershowitz)	56
V.5.3.	Population Studies of Microbial Ecosystems (Smouse)	59
V.5.4.	Admixture Studies and the Detection of Natural Selection (Ward, Adams)	62
V.5.5.	Continuing Participation in the Cytogenetic Studies of the ABCC (Bloom)	62
VI.	PROPOSED PROGRAM	63
VI.1.	THE CONTINUING PROGRAM ON THE AMERICAN INDIAN	63
VI.1.1.	Further Field Work (Neel)	63
VI.1.2.	Analysis of Data at Hand on the American Indian	65
VI.1.2.1.	Implications of Frequency of Occurrence of Rare Variants in Indian Villages (Li, Neel, Smouse)	65
VI.1.2.2.	Extensions and Applications of Matrix Theory (Ward)	65
VI.1.2.3.	Genetic Relationship Between Individuals Within Villages	67
VI.1.2.4.	Partition of Genetic Variation Within and Among Yanomama Villages (Spielman, Sing, Neel)	68
VI.1.2.4.1.	"Conventional" Cockerham Analysis	68
VI.1.2.4.2.	The Allele Correlations and Lineal Effect	69
VI.1.3.	Further Monte Carlo Simulation Based on Yanomama Population Structure (Li)	70
VI.2.	ORGANIZATION OF THE PROGRAM IN JAPAN (Neel)	72
VI.2.1.	A Hopeful Projection	72
VI.2.2.	Contingency Planning Should the Program Currently Proposed for Hiroshima and Nagasaki not Materialize	73
VI.3.	SUPPORTING LABORATORY ACTIVITIES	75
VI.3.1.	The Development of New Systems (Tanis, Ferrell)	75
VI.3.2.	The Alpha Helix Expedition	76

VI.4.	MATHEMATICAL AND STATISTICAL ASPECTS OF POPULATION GENETICS (Sing et al.)	76
VI.4.1.	Application of the Likelihood Principle to Problems of Estimation in Population Genetics (Rothman, Sing)	78
VI.4.2.	Genetic Statistics for Mutation Problems (Jensen)	79
VI.4.3.	Application of Yule's Law to Fixation of Alleles in Populations with Complex Demographic Structure (Rothman, Templeton)	80
VI.4.4.	Density-Dependent Selection in Age Structured Populations (Smouse, Weiss)	80
VI.4.5.	Consistent Two-Sex Models for Human Population Demography (Weiss, Smouse)	81
VI.4.6.	The Analysis of the Role Genes Play in Determining Normal Variation in Predictors of Risk for Coronary Heart Disease in the Tecumseh, Michigan Population (Sing)	81
VI.4.7.	Applications of Statistical Procedures to Estimate Selec- tion Effects in Parthenogenetic Populations to Human Populations (Rothman, Templeton)	86
VI.4.8.	Proposed Updating of the Computing Facility (Griffith, Sing)	87
VI.5.	SPECIAL STUDIES	
VI.5.1.	The Genetic Unit of Selection (Sing, Templeton)	88
VI.5.2.	Genetic, Chemical and Serological Studies of Human Immuno- globulins (Gershowitz)	93
VI.5.2.1.	The Search for Subclass and Allotypic Variation Among Non-IgG Immunoglobulins	93
VI.5.2.2.	A New Allotypic Specificity of the Human Kappa Light Chain	94
VI.5.2.3.	Further Investigation Into the Myeloma-Like Nature of an Anti-Rh Antibody	96
VI.5.2.4.	The Nature of Gm Genes (Kohler, Gershowitz)	98
VI.5.3.	An Additional Approach to the Estimation of Mutation Rates in a Small Genetic Isolate (Ward, Neel)	100
VI.5.4.	The Continuing Participation of Bloom in the Cytogenetic Program of the ABCC	103
VI.6.	CORE SUPPORT	104
VI.7.	REFERENCES	105
VII.	FEDERAL SUPPORT FOR DEPARTMENTAL PROGRAMS	108
VIII.	STATEMENT OF TOTAL PROJECT COST AND UNEXPENDED FUNDS STATEMENT	110
	EQUIPMENT REPORT	111

TABLE OF CONTENTS, CONT'D.

SECTIONAL BUDGETS	112
XI. SUMMARY BUDGET	125
XII. PUBLICATIONS SINCE LAST ANNUAL REPORT	127
XIII. PAPERS CURRENTLY IN PRESS	129
XIV. APPENDICES	
A. APPLICATION TO NSF TO UTILIZE ALPHA HELIX	131
B. ADDITIONS TO COMPUTER SYSTEM	141
C. TRIP REPORT, DR. ARTHUR BLOOM	149
D. SUPPLIES FOR PILOT STUDY IN JAPAN	154

I. SCIENTIFIC PERSONNEL, 1972-1973

NAME	POSITION	DEGREE	% OF TIME ON PROJECT
Arthur D. Bloom	Associate Professor	M.D.	15
Robert Ferrell	Research Associate	Ph.D.	100
Henry Gershowitz	Professor	Ph.D.	50
Ronald Griffith	Research Associate	B.S.	100
Louis Jensen	Assistant Professor	Ph.D.	25
Francis Li	Research Associate	Ph.D.	100
Ernest Migliazza	Research Associate	Ph.D.	25
James V. Neel	Professor	M.D., Ph.D.	50
William J. Oliver	Professor	M.D.	5
Edward D. Rothman	Assistant Professor	Ph.D.	25
Charles F. Sing	Associate Professor	Ph.D.	50
Peter Smouse	Assistant Professor	Ph.D.	25
Richard Spielman	Research Associate	Ph.D.	100
Robert Tanis	Research Associate	Ph.D.	50
Richard Ward	Research Associate	Ph.D.	100
Kenneth Weiss	Research Associate	Ph.D.	100

II. PROJECT ABSTRACT

As in past years, this continues to be a broad program in population genetics, with particular emphasis on a multidisciplinary study of some of the remaining relatively unacculturated Indians of South America and on an effort to develop a "second look" at the children of survivors of the atomic bombings of Hiroshima and Nagasaki, this consisting of a search for an increase in mutations which influence protein structure and function. In addition, it will be apparent that this year's program puts increased emphasis on certain theoretical aspects of population genetics. Finally, the program continues to encompass a number of smaller projects in experimental population genetics and to support Bloom's continuing participation in the cytogenetic studies of the Atomic Bomb Casualty Commission.

III. FOREIGN TRAVEL

Although mentioned at the appropriate places in the Budget, we have found it wise to flag foreign travel very specifically, not with reference to our fiscal year (1 Nov. - 31 Oct.), but to that of AEC (1 July - 30 June). Between these latter dates, for 1973-1974, we project the following:

<u>Individual</u>	<u>Destination</u>	<u>Dates (Approximate)</u>	<u>Cost</u>
Dr. Kishimoto, wife and child	Hiroshima- Ann Arbor- Hiroshima	1 Aug. 1973 - 30 June 1974	\$ 2,888
Dr. Neel	Ann Arbor- Hiroshima- Ann Arbor	1 Sept. 1973 - 31 Oct. 1973	1,155
Dr. Tanis	Ann Arbor- Hiroshima- Ann Arbor	1 Sept. 1973 - 31 Oct. 1973	1,155
Dr. Ferrell	Ann Arbor- Hiroshima- Ann Arbor	1 Feb. 1974 - 30 March 1974	1,155
Dr. Bloom	Ann Arbor- Hiroshima- Ann Arbor	1 Feb. 1974 - 28 Feb. 1974	1,155
Per diem (\$20/day)			<u>3,600</u>
			\$11,108

IV. INTRODUCTION

The principal projection of this program for 1972-1973 was to continue to phase out the large project on the population genetics of the American Indian while at the same time developing a major genetic study at the Atomic Bomb Casualty Commission. For reasons beyond our control, that latter program has failed to develop on schedule, and as of this writing it is impossible to set a precise timetable. Fortunately, the Indian data are continuing to yield much of interest to the population geneticist. At this point, despite the uncertainties regarding the program at the ABCC, we are still planning to initiate large-scale studies in Japan in 1974, and are projecting the necessary preparatory work during 1973-1974. If the Japanese program does not material-

ize, the demonstration (see Sec. V.1.7.) that primitive Indians may have higher mutation rates than civilized Caucasians suggests a number of exciting lines of inquiry, involving experimental animals as well as man, which we feel can profitably occupy our attention in subsequent years.

We will present both the Progress Report and the Proposed Program in five sections, recognizing that there is considerable overlap between the sections. These will be: 1) Studies on the American Indian (supported largely by the NSF contribution to this budget); 2) Organization of the Program in Japan; 3) Laboratory activities to support 1) and 2); 4) Activities of the Statistical Laboratory (exclusive of 1); and 5) Special Studies, these latter being a group of smaller undertakings with respect to various aspects of population genetics.

V. PROGRESS REPORT

V.1. CONTINUING STUDIES OF THE AMERICAN INDIAN

In this section we will briefly describe 8 facets of our studies of primitive populations which we feel to have come to fruition this year. The general themes of those studies are 1) to contribute to an understanding of human population structure, and 2) to contribute to an understanding of man's interaction with his environment at this cultural level. This year most of the progress reported pertains to 1).

V.1.1. Yanomama Demography (Neel, Weiss). - Over the past five years there has been a major effort to collect the data which would permit the definitive treatment of "primitive demography". This is the year of the analysis. The availability of a series of model stable populations based on classical demographic theory has allowed us to analyze the census data taken over the past several seasons in detail. The study rests on the census count from 29 villages by age and sex. These data have been rechecked many times by Chagnon, and are probably the best such data ever gathered. Ageing errors, universal in anthropological demography, have been corrected, and we have an age pyramid of 5-year age classes, with each sex being represented by over 1000 individuals.

We assume conditions of the stable population to hold. The age-specific birth and death rates are assumed to have remained nearly the same

over the past several decades, such that deviations from a stable age pyramid are the result of stochastic and sampling errors, or of sporadic events which merely overlie the relatively constant basic vital rates. We assume further that proportionately the amount of migration to or from the 29 villages surveyed is approximately equal. These assumptions are reasonable for a group such as the Yanomama.

The growth rate of the Yanomama can only be estimated for the past century. Evidence from E. Miglizza indicates that areal expansion, and expansion in number of villages, have progressed at between .5% and 1% per annum, and a rate as high as 2% can not be excluded. We attempt to fit the age pyramid to stable population models with growth rates between 0 and 2%.

It is necessary to know the age-specific fertility rates in order to construct a stable population model. We used three sources of data to determine these rates: 1. Uterine palpations made on fertile-age females, detecting pregnancy in the last 5-6 lunar months of gestation; 2. Urinalyses to detect urinary chorionic gonadotrophins, detecting pregnancy after approximately the first month of the gestational period; 3. Missionary reports concerning term births in some villages. By appropriate corrections, we arrive at a consistent set of fertility rates. An average fertile-age female bears a term birth every three years, approximately. We use these data to develop a relative age-specific fertility function to be used in the model construction.

We have fitted the Yanomama growth rate, fertility, and age pyramid data to over 500 model stable populations, and were able to determine a consistent, well-fitting, two-sex, female-dominant population model (cf. V.1.8.). The growth rate with this model is .85% per year, and the age-specific fertility rates are nearly identical with those observed. Given the uncertain age estimates and the small numbers, the age pyramid fits the prediction closely. The observed age-specific sex ratios are predicted as well. Our model thus includes a fit to, and hence estimation of, the infanticide rates among the Yanomama, which is on the order of 25% for females. We find that over the entire life span, only as infants do females have higher mortality rates than males. This is the opposite of what is generally observed among primitives, and is probably due to differential female infanticide and male mortality in warfare. Previous estimates of juvenile mortality have been too

low, owing to underreporting of births by Yanomama women. Juvenile mortality (including that due to infanticide) is on the order of 50%, and the average number of births to all women reaching maturity is about 5 1/2, with completed family size for women reaching 45 around 8 births; these latter figures are 1-2 births per woman higher than previously estimated.

Detailed data from 5 Yanomama villages are being used to derive descriptive statistics such as variance in family size, variance in fertility rates, mating patterns, paternity exclusions, child spacing, and infanticide. The models and parameter values can then be used to test the effects of changes in the vital rates among such populations as the Yanomama. The models can also be applied to such genetic and epidemiological topics as the survival of mutations, effects of diseases, effective population size, Crow's index of selection, and so forth. The model will improve on previously available estimates of primitive population structure.

V.1.2. Validation of Dendrograms (Spielman, Neel). - The correspondence between different kinds of population differentiation or "distances" has been extensively studied, beginning in the previous grant year. Now that dermatoglyphic differences among Yanomama villages are also available, Spielman has undertaken some extensions and improvements in the methods for comparing different kinds of data. Two rather different, but equally plausible definitions of "correspondence" of sets of data are proposed: one for similarity of implicit cluster-structure, and one for approximation to geometric congruence. In the technique originally developed and described in last year's report, the criterion of total path length is used to define and identify networks or dendrograms which are good representations of the cluster-structure implicit in a set of multivariate data for 7 Yanomama villages. The novel aspect of the comparison begins when we examine not one best but some collection of best networks having the shortest net length. For 7 populations, there are 945 possible networks of the kind used; it is convenient to examine the best 50 (about 5%) of these. If two sets of data correspond, it follows that nets which are good representations for one set (say the marker gene data) should be good representations for the other (e.g. the anthropometric). Thus we ask: of the best 50 networks for one set of data, how many are also found among the best 50 for the second set?

In order to assess the statistical significance of this "best-50"

statistic, its distribution under the null hypothesis of "no correspondence" is needed. For example, the anthropometric and marker gene data for the 7 villages used in Spielman's thesis yield 16 nets in common among the best 50; but can this many be expected by chance alone with high frequency? It has become clear that the levels of significance originally attached to given values of the "best-50" statistic were too optimistic, and a major effort has gone into obtaining by simulation the distribution of values to be expected by chance alone. This distribution has been obtained and provides a test of "cluster-structure correspondence" between different kinds of data on the same groups.

In addition, an attempt has been made to obtain corroborating evidence from a very different second approach which tests another notion of "correspondence". Using a technique developed by Schönemann and Carroll (1970), it is possible to fit one matrix of village coordinates (say the anthropometric) to another (e.g. genetic markers) by rigid rotation, uniform dilation-contraction, and a shift of the origin, and to measure the degree of departure from perfect geometric congruence using a least-squares deviation (called normalized symmetric error or N.S.E.). The simulation results were also used for a test of significance of correspondence defined as congruence, as described below.

In the computer simulation, 7 "populations" are placed in a unit hypercube of 6 dimensions using a uniform random number generator to choose a position on each coordinate. The 945 possible networks for these 7 simulated populations are evaluated and ranked according to net length. From two hundred such sets of 7 populations, 100 random pairs were formed to yield a sample distribution of the best-50 statistic for correspondence resulting from chance alone. Among these 100 trials, 14 or more nets in common were found among the best 50 five times, 16 or more were found twice, and 21 (the maximum observed) once. If these figures are taken as significance levels for the cluster-structure comparisons, the marker gene-map position comparison is judged highly significant (30 nets in common among best 50) and the anthropometrics-marker gene comparison is significant by conventional criteria (16 in common, $P \leq .02$).

The same simulation was used analogously to generate a distribution for the statistic (N.S.E.) which tests the congruence of two sets of data. For this definition, however, only the anthropometrics-map position comparison

is found to be significant ($P < .01$); no other comparison even approaches significance.

Neel, Rothhammer, and Lingoos then extended this treatment to the Yanomama dermatoglyphic data. For a group of 7 villages (partially overlapping with those treated by Spielman), distance matrices and dendrograms were generated for four sets of characteristics: genetic markers, anthropometrics, dermatoglyphics, and geographic distances. There are six pair-wise comparisons for correspondence of the four distance matrices yielded by these data. When these correspondences as measured by the Schönemann and Carroll N.S.E. were compared with the correspondence expected by chance, as determined by Spielman, in only one of the six comparisons did the correspondence exceed the 2 percent level of significance. The number of times nets found among the best 50 (least string criteria) characterizing one set of data are also found among the best 50 yielded by another set of data as shown in Table 1. By the criteria developed above, 5 of the 6 comparisons are significant at the 2 percent level, and two of these 6 at the 1 percent level.

Two conclusions seem inescapable: 1) the kind of correspondence in cluster structure defined by the network comparisons can be detected in the virtual absence of geometric congruence (tested by matrix-fitting), and 2) there exists substantial correspondence of the cluster-structure kind among various kinds of biological data, and of some of these with map positions.

V.1.3. A Comparison of Linguistic and Genetic Differentiation (Spielman, Migliazza). - In order to compare some form of cultural differentiation with the remarkable biological divergence shown by Yanomama subdivisions, we have quantified linguistic divergence in 7 major language or dialect areas. Two major kinds of linguistic data have been used as a basis for classification. The first is lexical differentiation, measured by the classical percentage of cognates between two languages in a given word list (750 items, in our case, with the Swadesh 200-word list as a subset). The second kind of linguistic data is the variation in the set of 38 grammatical rules accounting for syntactic and phonological differences.

Since the previous progress report, these data have been organized into generalized distances, and the distances used to generate cluster-structures or networks for comparison with geographic and genetic relationships among villages. A multivariate measure has been derived for linguistic differentia-

Table 1. The number of dermatoglyphic, anthropometric, genetic marker and geographic topologies in common among the best 50 of the 945 possible on the basis of 7 subsets of data.

Contrasted set	Reference set		
	Dermatoglyphic	Anthropometric	Marker genes
Anthropometric	18		
Marker genes	16	20	
Geographic	10	30	26

tion which is analogous to the geometric metaphor of generalized distance. Using the lexical data, for instance, we define as the square of the distance between two groups being compared, the total number of non-cognates in the word list. This definition is geometrically equivalent to placing the groups at two corners of a multidimensional "cube" with edges of length unity representing one cognate difference. The distance between two groups is therefore the appropriate diagonal of a cube with up to 750 dimensions. Exactly the same principle applies for the distances based on differences in grammatical rules. For comparison with the two kinds of linguistic distance, genetic distances were calculated (method of Cavalli-Sforza and Edwards) for representative villages in the 7 language areas.

The various kinds of networks generated from these distances were compared using the techniques developed by Spielman (see Sec. V.1.2.), and a manuscript describing the results has been prepared. By those techniques, it is clear that the degree of correspondence between genetic and linguistic differentiation for these groups far exceeds that expected by chance. However, some aspects of the statistical comparison developed by Spielman have undergone substantial revision in the last year, and the completion of the manuscript was delayed pending final developments of the method. With these achieved, the treatment of linguistic differentiation should be completed in the current grant year.

V.1.4. Further Studies on Salt Balance in the Yanomama (Oliver, Migliazza). - In a previous report we have provided data demonstrating that the Yanomama are maintaining salt balance with a urinary output on the average of less than 1 meq. per 24 hours, an almost unbelievably low value. At the time these urines were collected, blood specimens were collected for the determination of the levels of renin and aldosterone, the two hormones primarily involved in the regulation of salt balance. For technical reasons we were not satisfied with the results of the renin determinations. During the summer of 1972, Oliver and Migliazza returned to Brazil, for additional collections. The determinations of 24-hour urinary sodium and potassium excretion are complete, and confirm the earlier findings. It is anticipated that the determinations of serum renin and urinary aldosterone will be completed during this grant period, and that a report will be prepared for publication during this same period.

V.1.5. Migration Matrices for a Tribal Population (Ward, Neel). - Of the commonly recognized factors influencing the genetic microdifferentiation of subpopulations (selection, migration and drift), migration looms as the one factor that is particularly amenable to detailed analysis by reference to data from human populations. The study of migration as it affects the rate of genetic divergence is particularly appropriate for the Yanomama for two reasons:

1) The Yanomama represent one of the most completely studied non-urbanized populations. As such, any conclusions regarding the role of migration in the development of genetic microdifferentiation can be offered as tentative generalities applying to the stage of human evolution characterized by tribal societies with a simple culture.

2) The Yanomama, like other tribal populations, are characterized by a high degree of genetic microdifferentiation even though the migration between villages is appreciable. Investigation of the manner in which migration occurs, particularly in the context of the fission-fusion process, is thus likely to indicate how the observed extent of genetic heterogeneity is maintained in the face of such apparently high migration rates.

The present study is concerned with two distinct aspects of migration, migration between villages within the tribe and migration across tribal boundaries.

a) Migration between villages. In this section we were principally concerned with documenting the fission-fusion process as it affects the genetic structure of two historically defined clusters of Yanomama villages; the Shamatarí and the Namoweitarí (see Ward, 1972). In addition two other well documented instances of fusion could be examined in the light of their genetic consequences. For the Shamatarí and Namoweitarí villages we had recourse to historical data gathered by Chagnon and were thus able to reconstruct the actual evolution of the village clusters from, in each case, a single village, in terms of the estimated gene frequencies at each stage. This information yields a description of the genetic effect due to repeated fissions to form village clusters. Also by comparison with the estimates of ancestral village frequencies obtained by reference to a previous analysis of genetic relationships between villages, the deviations due to migration (fusion) events can be estimated. The observed estimates of genetic relationship were then compared with preliminary estimates of the expected genetic relationships, defined in terms

of Malécot's coefficient of relationship. Although the latter estimates were necessarily based on incomplete genealogical information, the data gave some indication of a higher endogamy rate for the Shamatari group as opposed to the Namoweitari group.

Finally a more general analysis of the effect of inter-village migrations on genetic diversity was attempted using the migration approach of Bodmer and Cavalli-Sforza. Since their model does not allow for changing migration rates through time nor for the effect of village fissions and fusions an attempt was made at generating an extension to their model. We investigated two types of matrix manipulation which imposed regular cycles of endogamy and exogamy (which at the limit is fission and fusion, respectively) on the colonies. Although no exact solutions could be found, the preliminary simulation results indicated that the extensive cycles of fission and fusion resulted in a greater rate of approach to genetic equilibrium. However these models can only be regarded as a first step in the extension of migration matrices since they do not yield a process readily interpretable in terms of the actual behavior of genetic subpopulations.

b) Migration between tribes. In this section, attention was focused on the distribution of two alleles, D_i^A and AP^A , thought to have entered the Yanomama from an adjacent tribe. While the existence of two such marker genes should in theory allow the rate of gene flow to be calculated, the distribution of the AP^A allele is sufficiently "spotty" within the tribe to raise the possibility that the acid phosphatase A allele had not entirely been lost from the Yanomama population before additional alleles were introduced from neighboring tribes. Alternatively the distribution of the AP^A allele may reflect the fact that when inter-tribal gene flow occurs, the diffusion of the introduced allele into the recipient population may be wholly unlike the classical process, as a consequence of the inter-village relationships described in the preceding section.

V.1.6. A Simulation of the Fate of a Mutant Gene of Neutral Selective Value in a Primitive Population (Li). - In last year's proposal we outlined four questions to be approached by Monte Carlo-type computer simulation. Thus far we have worked on the first of these problems, with rather surprising results. Using the Yanomama simulation program developed originally by MacCluer, Li has been studying the fate of a mutant gene introduced into this simulation population of four Yanomama villages, and relating it to certain

theoretical expectations. Specifically, Kimura and Ohta (1969 a, b) have derived an explicit formula by diffusion equations for predicting the average extinction time of a single neutral mutant in a finite population, namely,

$$\bar{t}_0 = 2 \left(\frac{N_e}{N} \right) \log_e (2N) \quad (1)$$

where $\bar{t}_0$ is the mean extinction time, N is the size and N_e is "effective" population size. In an organism with overlapping generations, such as man, N must be based on the number of individuals in the "reproductive generation", in order to render the definition comparable with "annuals". "Effective" population size has been variously defined and calculated; for a randomly-mating population one definition is "the population size that would give the same amount of drift in a reproducing population as that in the ... model having N constant in time, non-overlapping generations, and all individuals and each sex expected to make equal contributions to the progeny" (Cavalli-Sforza and Bodmer, 1971). This particular N_e is sometimes termed the "variance effective number", and it is this N_e which Kimura and Ohta (1969b) employ in the above formula. It is assumed that the mutant allele is represented only once at the moment of its appearance, with initial frequency $p = \frac{1}{2N}$; the $\bar{t}_0$ value, of course, excludes those mutants going to fixation.

For this population of four villages, N_e has been computed by two different methods:

1) The first estimate of N_e is obtained from the variance of the gene frequency obtained in a series of simulations, namely, $N_e = \frac{pq}{2V_p}$, where V_p is the variance of gene frequency obtained by dropping a pair of alleles with initial frequencies $p = q = 0.5$ 100 times through a pedigree constructed from a 400 year simulation of this population (MacCluer and Neel, unpublished). Two different pedigrees were used. During the course of the simulation the two populations increased in size from the initial value of 451 to 465 and 551 persons, respectively. The two estimates were 180 and 210; we shall use the average of 195.

2) The second of the estimates results from a formulation by Nei and Imaizumi (1966), as independently extended by Cavalli-Sforza and Bodmer (1971) and Crow and Kimura (1972). These formulations attempt to take into consideration the occurrence of overlapping generations in human populations but assume a stable age distribution and that age specific birth and mortality rates are constant. The formulations of Cavalli-Sforza and Bodmer (1971) and Crow

and Kimura (1972) yield similar results in this population; we shall use the latter's, namely

$$N_e = N_0 \bar{l} \tau \quad (2)$$

where N_0 is the number of births in one census unit (say, 1 year), $\bar{l}$ is the weighted probability of surviving into the reproductive period, and τ is the average age at reproduction. Since $\bar{l}$ and τ are based on age-specific vital rates, they (and N_e) differ for males and females, and so we must compute N_e as twice the harmonic mean of male and female effective sizes, according to the formula

$$N_e = \frac{4N_e^f N_e^m}{N_e^f + N_e^m} \quad (3)$$

Mutant loss may occur in almost any generation of the simulation. Rather than derive a weighted estimate of N_e based on all the simulation runs, we have for present purposes simply computed N_e on the basis of the population present at the completion of the simulation in the two 400-year simulations referred to earlier. This should result in satisfactory comparability between estimates 1) and 2). The two estimates are 204 and 213, and we shall use the average, of 209. We note in passing that these two estimates of N_e are derived from the population at the completion of the simulation, whereas the estimate of N_e to be derived from $\bar{l}_0$ is based largely on the population towards the beginning of the simulation. Since for these two runs population size changed relatively little during the simulation, this cannot be an important source of discrepancy.

For the computation of $\bar{l}_0$, we will set N at 200, twice the number of women in the reproductive age. This is done to extract from this population, containing representatives of four overlapping generations, an estimate of N comparable to that employed for an annual species. Because of polygyny and adultery, we cannot in the strict sense speak of couples, but the 100 women correspond roughly to 100 sets of reproductive partners, with some men having children by several wives, and vice-versa. In equating N to this value rather than the census value, we are, we believe, using N in the sense employed

by Kimura and colleagues in the derivation of equation (1). The values of $\bar{t}_0$ predicted from these two estimates of N_e (cf. 195 and 209) and a value for N of 200 are approximately 12.

The simulation value of $\bar{t}_0$, by contrast, is 3.3 generations, approximately one-quarter that predicted by the Kimura formulation. If we consider $\bar{t}_0$ for those mutations whose possessor survives to adult life the first generation, it is 4.7. This discrepancy is undoubtedly due to many factors, but at this writing, two stand out:

1) In the Yanomama, pre-reproductive mortality is approximately 50%, whereas the Kimura model assumes a replacement type population with no pre-reproductive mortality and a Poisson-type distribution of number of children.

2) There is in the model preferential village endogamy, which impedes the spread of a mutant from one village to another.

Our conclusion is that the N_e of reference in mutant survival is not the "variance effective N_e ", but in this population, a substantially smaller figure. Conversely, average mutant survival, because of this smaller N_e , is substantially less than one would predict using standard formulations.

V.1.7. An Indirect Estimate of Mutation Rates Employing Biochemical Variants (Neel). - In the preceding section, we described studies on the average mutant survival time in the Yanomama. In Section VI.2. we shall describe the frequency with which biochemical variants of 15 different serum proteins and erythrocyte enzymes have been detected by electrophoresis in the Indian tribes studied to date. Kimura and Ohta (1969) have treated the problem of mutant loss in populations of finite size, and, on the assumption of a stable population size and an equilibrium situation, developed a relationship between the mutation rate to alleles with neutral phenotypic effects (μ) and the frequency and mean extinction time for these mutants, namely,

$$\mu = \frac{I}{2N} \cdot \frac{1}{\bar{t}_0}$$

where

I = the average number of mutant alleles per locus among the loci sampled

N = the number of independent genomes sampled, and

$\bar{t}_0$ = the average mutant survival time for those mutants not going to fixation.

The final formula has an appealing simplicity: on the assumption of equilibrium, the mutation rate is equal to the frequency of variants per locus ($\frac{I}{2N}$) times the fraction lost each generation ($\frac{1}{t_0}$).

In order to use this formula we must estimate I and N . In the 15-40 age group, there were 9 rare protein variants plus the Tf_{D_1} and PGM_1^2 polymorphisms. Fifteen proteins were examined, but two of these (hemoglobin and lactic dehydrogenase) are polymers of two different polypeptides, so the products of 17 cistrons were under scrutiny. Our best estimate of I is thus $11/17 = 0.65$. We have estimated 1853 persons in the adult generation of our total Indian sample. N as defined by Kimura and Ohta implies independence of the genomes tested, but our sample includes some siblings, parents and offspring, and others of more remote relationship. Thus, the number of independent genomes is somewhat less than 1853. Any attempt to correct sample size for this fact is complicated by the relatively short t_0 , which implies that relatives are less non-independent tests of mutation than the coefficient of relationship would indicate. We shall attempt no correction at this time (resulting in an under-estimate of μ). $I/2N$ thus becomes 0.00017. As noted earlier, the mean value for t_0 when one considers the adult population is 4.7 generations. Substitution of this value in the equation leads to an estimate for μ of about 3.6×10^{-5} /locus/generation. The fact that the estimate is based upon an expanding population should bias the estimate upwards, but probably not greatly.

It should be noted that albumin variants comprise almost half the total on which this estimate is based. Only further work will determine whether 1) genetic variants are especially common at the locus encoding for albumin, or 2) albumin variants are detected with higher-than-average efficiency in the present systems, or 3) this is a sampling quirk. At this stage of the problem, we can only speak of average rates, but in time it should be possible to consider locus-specific rates.

Use of the Indian data in this context also assumes selective neutrality on the part of the variants encountered. Since at the gene frequencies concerned, the variants were usually encountered in the heterozygous state (131 of 135 occurrences), the assumption basically reduces to neutrality as a heterozygote. To the extent the variants are deleterious, the assumption of neutrality is conservative with respect to the calculation of mutation rates. If, on the other hand, as many as one-third of the variants encountered were

maintained as balanced polymorphisms, the estimate of mutation rate would be reduced but, because of the random loss of even favorable mutants, by an amount difficult to calculate but less than one-third.

In the case of the hemoglobin molecule, it can be shown that less than one-third of all possible amino acid substitutions will result in a charge change. Assuming this is true for all the polypeptides under study, a rough correction of the estimate to include the "silent" mutations is achieved by multiplying by a factor of 3, resulting in an estimate of $\sim 10^{-4}$ /gene/generation. Boyer (1972) has recently shown that with respect to primate hemoglobin, the ratio of "silent" (non-electrophoretically detectable) substitutions (ascertained by sequence analysis) is substantially higher than the predicted 2:1. This suggests heavier selection against the electrophoretically detectable than the "silent" mutations. If as a general rule there is negative selection against the electrophoretically detectable mutants, this is another reason to regard the present estimate as conservative.

On the basis of a study of the frequency of origin through mutation of approximately a dozen rare phenotypes in civilized Caucasian populations, the rate of mutation in man is in round figures often given as 10^{-5} /gene/generation. Reasons have been advanced why this may be an over-estimate, and some prefer a figure of 10^{-6} (Stevenson & Kerr, 1967; Cavalli-Sforza & Bodmer, 1971). Thus, the estimate generated by these data must be regarded as relatively high, especially when it is realized that even with the attempted allowance for mutation resulting in silent amino acid substitutions, not all mutations are detected by this approach. For instance, mutation resulting in duplication, deletion, or inactivation of a genetic locus would probably go undetected. It is clear the present estimate is destined to undergo many refinements. But already it should have served a useful purpose, since to reconcile the present data on variant frequency with a mutation rate of 10^{-6} requires for Indian population a $\bar{t}_0$ of 175, clearly an excessive value.

We believe that for the present, caution should be observed in applying this rate either to civilized populations or other ethnic groups. There are no data really comparable to the present (defined groups surveyed for so many proteins) for Caucasian populations. In surveys of single or several proteins, private variants (i.e. exclusive of polymorphisms) have thus far been detected with a frequency of approximately 1/900. A more important datum from the standpoint of estimating mutation rate is the ratio, $I/2N$, since the

variants encountered in surveys of Caucasians have usually been different, $1/2N \approx 0.0005$. (This frequency is biased upward by the hemoglobins where, contrary to the other systems, it is routine to fingerprint and sequence electrophoretically similar mutants, thus permitting a decision as to whether different amino acid substitutions are involved.) However, we argue that if these Indian populations were "shuffled together" in the way that Caucasian populations have been in the process of detribalization, then in a sampling of, say, Indian adults, the number of different variants per 1000 tested would increase, although the frequency of any given variant would decrease. In other words, the ratio of $1/2N$ would be higher if instead of sampling villages as we have, our sample consisted of unrelated Indians scattered the length of South America. We suggest that under these sampling conditions, $1/N$ would be at least as large as in Caucasians, and quite possibly greater. This would increase an already high estimate of mutation rate.

There is as yet no simulation of t_0 for a civilized population. Since a major reason for the short t_0 value observed in the Yanomama is the relatively high infant and childhood mortality in this group, we surmise that t_0 in a civilized group would be greater. If $1/N$ for Indians is equal to or greater than the value in civilized Caucasoids, and if t_0 is shorter in Indians, the implication is that mutation rates are no less and may actually be somewhat higher in primitive Indian than in civilized Caucasian populations. If the apparent difference proves valid, the present type of data will not permit a decision as to whether the basis for the difference is ethnic or due to environmental factors. It may be difficult to reach a decision, since ethnicity and degree of exposure to urbanization-industrialization tend to be confounded, but one can visualize developments, such as the demonstration of similar rates in other Mongoloid groups, which would have to be considered suggestive of the proper interpretation.

V.1.8. Demographic Models for Primitive Human Populations (Weiss). - Weiss in 1972 submitted a thesis entitled "Model life tables for pre-industrial populations" in which an initial effort at creating generalized demographic models for a wide range of primitive human populations was made. Considerable effort has gone into the development of that initial effort during the current grant year, resulting in a monograph, "Demographic models for anthropology", to be published in April 1973 as Memoir 27, Society for American Archeology. This

work has improved and expanded the original thesis and has added a complete discussion of means of adapting the models to various changing demographic situations, nonzero growth rates, and two-sex population models.

The work is based on a worldwide survey of existing data from primitives and from skeletal and historical sources through medieval European times. It relies heavily on experience gained this year in the analysis of Yanomama demography. Model mortality and fertility schedules are derived from published data, and a series of model stable life tables is constructed following classical demographic theory in a way not before attempted. Included are discussions of problems in handling available anthropological field data, largely as learned from the Yanomama. Methods are derived for adapting the models to test the future effects of changes in the vital rates, and for computing consistent two-sex population models from typical data.

The work introduces a revised estimate of the history of human life expectancy and a measure which reflects the relative importance of mortality and fertility factors in controlling population growth. Methods for fitting crude and fragmentary data are stressed, so that more of the available anthropological data may be brought to bear on studies of human population biology. The resulting monograph provides the only comprehensive work applying demographic theory to anthropological populations and generalizing about their demography. The techniques for handling field data were written especially to be applied to the demographic analysis of the Yanomama currently being done by Neel and Weiss (Sect. V.1.1.).

V.2. ORGANIZATION OF THE PROGRAM IN JAPAN

A separate proposal to AEC has dealt with the initiation of a large-scale search for evidences of increased mutation in the children of survivors of the atomic bombings, the F_1 , utilizing biochemical methods of the type employed in the study of the American Indian. Unfortunately, the "Japanese Side" has thus far limited its approval to a Pilot Study, to be based on blood specimens reaching the Atomic Bomb Casualty Commission in connection with studies already in progress.

In July-August, 1972, Tanis and Neel were at ABCC to initiate this Pilot Study. This necessitated not only equipping a laboratory for electrophoretic studies and training a staff, but also developing all of the necessary forms

and clerical procedures. The serum proteins and erythrocyte enzymes currently under study are: 6-phosphogluconate dehydrogenase (6-PGD), phosphoglucomutase 1 and 2 (PGM₁ and PGM₂), lactic dehydrogenase (LDH), malic dehydrogenase (MDH), carbonic anhydrase I and II (CA I, CA II), adenosine deaminase (ADA), peptidase A (Pep A), peptidase B (Pep B), phosphohexose isomerase (PHI), isocitric dehydrogenase (ICD), Hemoglobin A (Hb A) and A₂ (Hb A₂), transferrin (Tf), ceruloplasmin (Crpl) and albumin (Aib). The sole basis for inclusion in this study was that the protein (enzyme) in question yielded stable, sharply defined "bands" in the appropriate system, as demonstrated by experience at the University of Michigan. Systems with which we have experience but which, in our hands, did not meet these criteria were: haptoglobin, Group-specific component (Gc), acid phosphatase, galactose-1-phosphate uridyl transferase, glucose-6-phosphate dehydrogenase, and catalase. The previous occurrence of variants in the hands of others was not a basis for selection. Two of these proteins (LDH, hemoglobin) are composed of two subunits, so that the survey involved the product of 17 different cistrons.

Ferrell then spent January-February 1973 at ABCC, reviewing the progress to date. All but one of the systems introduced last fall are yielding satisfactory results. To date there are approximately 750 determinations for each of 17 systems, for a total of 12,750 system determinations. Eleven variants have been encountered, namely, 3 of CA I, 1 of Hb A, 2 of Pep B, 2 of PHI, and 3 of Alb. This frequency, of ~ 1:1000, compares favorably with the results in the literature. Twelve CD type transferrins and 3 CB types have also been detected. However, these may be normal polymorphisms present in Japanese populations and have not been classified as variants. In addition, there are well recognized polymorphisms segregating at three of these loci (PGM₁, ADA, and 6-PGD).

At the time of the last Application, it was projected that the full-scale study would be initiated in the fall of 1973. At that time, we planned to acquire during the current year a second-generation GEMSAEC-type fast auto-analyzer and to begin to adapt it for quantitative studies in Japan of the same enzymes being studied for qualitative variants by electrophoresis. With the current uncertainties about the future of the program at the ABCC, the decision was made to delay purchase of the equipment until a full program in Japan is assured. The time to have been allocated to these developments has instead gone into the exploration of new electrophoretic systems and extending the

Indian series. However to acquire a "nuts and bolts" feeling for the GENSAEC analyzer we (Neel, Tanis, and Ferrell) have visited Dr. Charles Scott's laboratory at the Oak Ridge National Laboratory this past January. At that time we had an opportunity to experiment with the miniature, NASA version, of the fast analyzer series. In addition to routine runs, a red cell hemolysate enzyme (Glucose-6-Phosphate dehydrogenase) was analyzed; the preliminary results were encouraging and suggest that our proposed applications appear well within the capabilities of the instrument. Furthermore, the development of computer programs to handle "isozyme stripping" (i.e., the determination of the K_m values of isozyme mixtures) is now well along. It is our intention to return to ORNL later this year, when a permanent unit will be available for testing, to experiment with additional assay procedures. Ideally we would hope to take to Japan a battery of at least six assay procedures which would be immediately applicable; other procedures would be added at a later date when feasibility studies, carried out in this country, indicated they were ready for general use.

The Pilot Study was to have given some better insight into the frequency of variants which might require family studies, and the amount of cooperation to be expected from the F_1 and, in case of the need for family studies, from the families of the F_1 . We do now begin to have, on the basis of the frequency of "rare variants" (~ 1 in 1000 determinations or 1 in 80 persons) a better estimate of the frequency with which family studies will be necessary. On the other hand, approximately 90% of the specimens becoming available are not from F_1 ; ABCC has contacted so few F_1 during the period since the Pilot Study was initiated (September 1972 - March 1973) that there is little basis for a judgment on their cooperation and, when necessary, that of their families.

It is anticipated that late in the current grant year (September - October, 1973), Drs. Neel and Tanis will proceed to Hiroshima and Nagasaki, for a general review of the program which will include training of new personnel. It is anticipated that by then, a definite decision will have been reached concerning the expanded program in genetics. If the decision is favorable, it will be necessary at that time to work out the joint U.S.-Japan governance of the program and recruit for the expansion. If the decision is unfavorable, at least for the foreseeable future, a decision must be reached concerning the management and future of the Pilot Study.

V.3. SUPPORTING LABORATORY ACTIVITIES

V.3.1. The Activities of the Biochemical Screening Laboratory (Tania, Ferrell). - The Screening Laboratory has continued the effort of typing the red cell enzymes and serum proteins of South American Indian blood samples. This work has involved both newly collected bloods and stored specimens. Table 2 summarizes the number of blood samples processed in our laboratories during the past year. Old systems include: AP, PGM₁, PGM₂, LDH, MDH, AK, 6-PGD, and ADA. New systems include: Hb, Pep A, Pep B, PHI, and ICD. Serum proteins include: Albumin, Gc, Tf, Hp, and Crpl.

Table 2
A Summary of the Search for Protein and En-
zyme Variants During 1972-1973.

Procedure	Number of stored bloods	Number of new bloods
Old systems	0	675
New systems	2591	601
Serum proteins	0	598

This work has resulted in some 24,000 system determinations which have substantially extended our knowledge concerning the frequency of polymorphic systems. It has also resulted in the identification of 5 rare variants distributed among the following systems: Pep A, PHI, ICD, Crpl and Albumin. Much of this recent work has been included in a manuscript which summarizes the variants encountered in 15 proteins for the Yanomama Indians and five neighboring tribes.

In addition to the routine screening of South American Indian samples, the laboratory has performed the complete series of electrophoretic analyses on a group of 140 samples from the Maroon population of Western Jamaica. This unique population is of African origin, but has remained relatively isolated for the past 200 years. From our small sample we were able to detect the AS and AC hemoglobin polymorphisms as well as several "private" variants. We hope to continue this study and look, also, at the Maroon population of Eastern Jamaica.

An additional responsibility of the Screening Laboratory has involved providing assistance to the Biochemical Genetics Study being initiated at the ABCC in Hiroshima, Japan. Tanis and Ferrell have set up a laboratory there, trained professional and technical personnel, provided that group with a detailed manual of screening techniques, as currently used in our laboratory, and reviewed their progress at the end of six months.

Finally, this Laboratory has spent a certain proportion of time experimenting with acrylamide gel electrophoresis as an alternative method to our routine starch gel procedures. In one case, ceruloplasmin, superior band patterns are now easily obtainable using acrylamide. However, to date several of the red cell enzymes appear to run as well, or better, on starch relative to acrylamide gel. In addition to acrylamide and starch we are also investigating techniques using cellulose acetate strips. Already this has proven applicable to screening stored South American Indian samples for the red cell enzyme carbonic anhydrase B (CA I).

V.3.2. The Activities of the Serological Screening Laboratory (Gershowitz).

a. Blood typings. - As part of the study of South American Indians, the laboratory performed blood typing for the ABO, MNS, Rh, K, P, Fy, Jk, Le and Di systems on bloods collected during 1972. These consisted of 185 individuals of 7 Makiritare villages, 98 individuals of a Macushi village, 62 individuals of an Arawak village and 137 individuals from two additional Yanomama villages.

As part of the continuing effort of the laboratory to cooperate with collaborators in the study of unusual populations, we have undertaken to assist Dr. J. Ahern of the University of the West Indies, Jamaica, in a study of the Maroons referred to in the preceding section. The 140 samples referred to above have also been typed for the abovementioned systems.

b. Gm and Inv typings. - The sera of all the individuals described above have been typed for the Gm and Inv systems, as have sera from 90 Apache, 189 Navajo and 202 Zuni from southwest U.S., and 222 Ojibwa from Canada.

With regard to the results of Gm tests among American aboriginals, we had previously described the fact that we had been unable to detect any evidence for the existence of the Gm(b) specificity in the South American Indians we had tested. We can now report a clear distinction between North

and South American Indians in that several different Gm(b)-containing genes are found among the former but none can be found among the latter.

Among the more than 4000 South American Indians tested from among the Xavante, Makiritare, Macushi, Moro, Arawak and Cayapo of Brazil and Venezuela, only two phenogroups Gm^{ag} and Gm^{axg} have been found to be segregating. Among Orientals on the Asian mainland the genes normally found are Gm^{ag} , Gm^{axg} , Gm^{afb} , Gm^{ab^0s} , and Gm^{ab^0st} . Gm^{afb} , we have demonstrated, is at a very high frequency in Southeast Asia and falls markedly, in a clinal fashion, as one proceeds northerly. Conversely, the Gm^{ab^0} genes increase from south to north Asia.

It seems now that the Gm^{ab^0} genes are fairly well distributed in North American Indians, in contradistinction to their apparently total absence from South American Indians. We have studied two Ojibwa populations from Canada, and Apaches, Navajo and Zuni from the southwestern U.S. and have found appreciable frequencies of the Gm^{ab^0} genes in all populations.

It seems unlikely that Gm^{afb} exists in aboriginal populations of the Americas, but the presence of Caucasian admixture in each of these Indian populations (as detected by the appearance of the Gm phenotype Gm(fb) known not to occur in Indians but occurring with high frequency among Caucasians) makes it difficult to exclude the presence of the gene in low frequency. Family studies are needed to distinguish whether an Indian of the phenotype $Gm(agfb)$ is genotypically Gm^{afb}/Gm^{ag} or Gm^{fb}/Gm^{ag} .

In any event, the absence or at least, very low frequency of Gm^{afb} in American Indians correlates well with their presumed origin from North Asia in that their ancestral population would have been homogeneous for the absence of that gene. It is not then necessary to invoke drift as the origin of a "difference" between Orientals and American Indians, for the difference may not exist at all between the American Indians and their presumed North Asiatic progenitors.

The Gm(b⁰) difference between North American and South American Indians on the other hand, may have arisen through drift or selection. But a search of "pureblood" individuals of various North American tribes may reveal a population abnormally low in the frequency of Gm^{ab^0} genes, a population from whom South American Indians may have been derived, and thus obviate the need to invoke these two explanations, one untestable and the other unproven. We point out again that genetic drift would have been the explanation of the presumed

difference between North American Indians and Orientals, since Asiatic Oriental populations possess Gm^{afb} and North American Indians do not.

V.4. MATHEMATICAL AND STATISTICAL ASPECTS OF POPULATION GENETICS

Our experience in the analysis of human population data has made it clear that the choice of the technique applied to the study of a number of genetically relevant questions has not always been obvious. Research in theoretical population genetics has developed a body of information that has been the basis for a number of deterministic models for the distribution of alleles among individuals. Such models may be simple functions of allele frequencies or they may reflect a complex relationship between such frequencies and parameters defined to measure allele correlations, mutations, linkage, migration, mating preference, age, etc. The broad range of population genetic models generated by the work of Fisher, Wright, Haldane, Kimura, Crow et al., all have in common simplifying biological assumptions. We have a plethora of models. Are these models useful? Much work is yet to be done to understand the consequences of the simplifying assumptions which are made when such models are applied to human populations. It has been our intent to study the available models to determine which, if any, are most useful to describe the human population. When modifications seem appropriate, they have been incorporated. An equally important consideration in our research has been the evaluation of methods for determining the estimates of parameters of the various models. The research reported below reflects both considerations.

Analyses of statistical data collected by the AEC program from actual populations provide the information which we have used to discriminate among models and to evaluate alternate methods of estimating parameters which make the selected model useful. A good model for human populations is judged in many ways. Is it a biologically realistic definition of the population? Does it predict the characteristics of unstudied individuals? Given a specific model of interest, what are the best estimators of the parameters of the model? The answers to such questions will come from experience with real human populations.

In the past year we have 1) developed a likelihood model for estimation of parameters which define population structure, 2) evaluated the application to human populations of the least squares analysis of variance procedure suggested by

Cockerham, 3) developed a genetic analysis of continuously varying variables defined as indicators of coronary heart disease, 4) completed a linkage analysis in Tecumseh, and 5) developed new insights into the application of demographic models to human populations.

V.4.1. Estimation of Parameters which Measure Human Population Structure (Rothman, Templeton, Sing). - The relationship between allele and genotype frequencies for a gene locus in an interbreeding Mendelian population constitutes the measurable statistical quality which is most often available to estimate a variety of genetic parameters. Three useful genetic parameters for a locus are the set of allele frequencies, the correlation between alleles within individuals and the correlation between alleles in different individuals. Knowledge of the values that these parameters take in a human population provide 1) a description of the consequences of the operation of various forces which occurred in the past or 2) a basis for prediction of the evolutionary fate of alleles in replicates of the population in the future. The values of these parameters at any time, and the rate of change over time, may be considered a function of mutation, migration, population size, and selection.

Data have been collected from human populations by our group and others to estimate these parameters. The preferred procedure for statistical evaluation of such data is dependent on the methods used in sampling and the criteria one sets down for properties of the estimates one wishes to obtain. A comparison of the alternate statistical methods available for estimation of genetic parameters in contemporary human populations is not dependent on the causative forces involved. On the contrary, it is obvious that a meaningful evaluation of the relative roles of such forces depends on the nature of the information obtained from the statistical analysis of the data.

We are concerned here with contrasting the properties of two alternate statistical strategies for estimation of the measures of allele association. Those two procedures are the analysis of variance (ANOVA) and the maximum likelihood method for estimation (M.L.E.). The ANOVA analysis follows from the recent work of Cockerham (1969, Ms.). To date, there has not been a likelihood procedure for estimation of genotype and allele association in small subdivided populations. We have developed the likelihood equations as a basis for an alternative to the least squares approach suggested by Cockerham. The reasons for wanting an alternate procedure are several. Primarily, the M.L.E.

procedure incorporates more of the information from the data than the ANOVA based on the least squares approach to estimation. Secondly, other than this appealing property of sufficiency, the likelihood procedure allows for the convenient incorporation of auxiliary information into the estimation and hypothesis testing scheme. Additional considerations are the efficiency (smallest possible variance) of estimates and power of the associated hypothesis testing schemes to reject the hypothesis of no association.

The data being used to contrast the statistical properties (bias, efficiency, etc.) of alternative procedures come from our studies in the Tecumseh, Michigan community and from a computer simulation of mouse populations. Work on the Tecumseh data has progressed the farthest. Point estimates of F , the correlation between alleles within individuals, and θ , the correlation between alleles in different individuals from the same kindred, are given in Table 3 for the six largest kindreds in Tecumseh. The kindreds range in size from 121 to 59 individuals. A kindred is defined as a group of individuals related by descent. The expected F for this community is zero. The M.L.E. procedure underestimates this parameter but tends to vary less among loci. We expect individuals in kindreds to have alleles in common, hence the expected value of θ is greater than zero. The least square estimates are on the average double those obtained from the application of the M.L.E. procedure. The distribution of kinds of relatives in these kindreds suggests the true θ to be near the M.L.E.

A manuscript giving the derivation of the likelihood equations, the derivation of variance covariance matrix for allele frequencies, F , and θ , and contrasts with the ANOVA procedure on selected sets of data is near completion.

V.4.2. Development of Statistical Procedures to Estimate Parameters Which Define Selection Effects in Parthenogenetic Populations (Rothman, Templeton). - The objectives of this work are (1) to formulate deterministic and stochastic models describing evolution in parthenogenetic populations reproducing by any combination of central fusion, terminal fusion and gamete duplication, and (2) to develop a statistical framework that can be used to make inferences about selection in such populations. (See Section V.5.1. of the Progress Section for a definition of the biological properties of the parthenogenetic population.)

These objectives have been met for the one and two locus models with constant viabilities. Deterministic models have been constructed that allow

Table 3 . Estimates of F and θ from the six largest bands in Tecumseh.

Parameter	Procedure ⁺	Locus used for Estimation					Mean
		IN	Ss	Haptoglobin	Gc	Gm	
F	L.S.	-.0318	.0931	-.0029	.0224	.0458	.0233
	M.L.E.	-.0653	.0223	-.0220	-.0024	-.0021	-.0139
θ	L.S.	.0204	.1100	.0254	.0308	.0487	.0471
	M.L.E.	.0146	.0437	.0076	.0141	.0254	.0211

+ L.S. = Least Squares

M.L.E. = Maximum Likelihood Estimation

the possibility of gamete duplication, central fusion and terminal fusion. Stochastic analogues of these models were then formulated which allow the development of n-generation likelihood functions that explicitly treat the dependence that exists between observations at different generations but from the same population. This likelihood function was used to develop two tests for heterotic selection for the one locus model: the locally best test in the region of no selection and the Bayesian likelihood odds of the hypothesis of no selection versus the hypothesis of selection. For the two locus model, estimators for selection coefficients under a wide variety of fitness models were obtained. Furthermore, a hypothesis testing framework was developed that tests for the presence of selection and distinguishes between various fitness models. In particular, measures of epistasis are proposed which are estimable by maximum likelihood and which can be used to test a variety of hypotheses. These statistical models provide a basis for studying non-random associations between non-alleles caused by natural selection in parthenogenetic populations.

In addition to the above, n-locus, one generation models have been developed for a parthenogenetic population reproducing by gamete duplication. The model allows the direct estimation of viability of an n-locus genotype. Furthermore, this viability can be broken down into fitness components which describe how much of the fitness observed at the n-locus level is ascribable to additive selection (or multiplicative) at some lower level of genetic organization and how much is due to interactions between such levels and higher order levels. The level of genetic organization assumed to be additive or multiplicative is arbitrary and can be chosen to be anything from the individual locus to the entire genome. More importantly, a thorough hypothesis testing framework accompanies the estimation procedures which allows all models to be tested not only against the null hypothesis of no selection but against one another.

V.4.3. An Application of the Analysis of Variance of Allele Frequencies to a Human Population (Sing). - Multilocus analyses which have been completed (Sinnock and Sing, 1972a, b) on 12 polymorphic blood and serum loci indicate that many of the observed allele variations may not simply be random noise in an imperfect biological system. The role which linkage (Glenn, Schull, and Sing, 1973) or small effective population size (Sinnock and Sing, 1972b) may play in determining deviations from multi-locus equilibrium has been considered. Neither factor could account for the observed correlation between nonalleles in

gametes. A third possible influence on observations of multilocus phenotype frequencies may be the subdivision or stratification of the population with respect to allele frequencies of the genetic markers being studied. Such a stratification within a population is measured as a dispersion, or variance, of allele frequencies among the corresponding subdivisions and correlation between alleles of the locus within subdivision. We showed (Sinnock and Sing, 1972b) that the nonrandom distribution of alleles among strata could result in a biased estimate of the correlation between nonalleles when such estimates were computed from a sample from the entire population ignoring the subdivision. Strata could be defined in a multitude of ways. In a paper to be published in a volume on population structure (Sing et al. 1973) we chose to take kindreds in Tecumseh as the strata. A kindred was defined by the collection of all persons who could be related by a pedigree. Kindreds were chosen for this study because any significant variation among strata, whatever the criterion, should contribute to the variation among subdivisions of related individuals. Our null hypothesis was that randomly selected kindreds from Tecumseh will represent the allele frequencies in the Tecumseh community. Heterogeneity beyond that due to sampling variation was considered evidence to favor the hypothesis of stratification. A measure of covariation of alleles within the Tecumseh population due to subdivision into kindreds provides a useful measure of the degree to which pairs of relatives drawn from the population are not genetically independent because they come from the same kindred. This consideration comes into play when we use correlations between relatives as an analysis procedure to evaluate the role of polygenes in determining variation in continuously varying risk variables to be discussed in Section 4.4 of the proposal section.

One other objective of this research has been to evaluate the statistical methodology developed by Cockerham (1969, 1973). This study is the first application of the procedure of analysis of variance to the problem of covariation of ^{single locus} allele frequencies in a human population. It provides a reference for our investigations of statistical methodology reported in Section 4.1 of the progress and 4.1 of the proposal section.

A summary of the results of such an analysis is given in Table 4. Two correlation parameters are estimated. The model for this particular analysis represents each allele by the measure X_{ij} , where i denoted the individual and j the chromosome of the i^{th} individual; $j = 1, 2$. The correlation

Table 4

Least squares analysis of pairs. Point estimates of F and θ .

The number of pairs is given in parentheses

Pair	θ	Parameter Estimated	Genetic System					Average
			MN	Ss	Haptoglobin	Gc	Gm	
Youngest Couple	.0	F θ	.0178 -.0080 (858)	.0198 .0010 (858)	.0055 -.0092 (848)	-.0296 .0018 (853)	.0014 .0272 (803)	.0030 .0026
First Cousins	.0625	F θ	-.1001 .0026 (69)	.1083 -.0608 (69)	.0214 .0563 (59)	-.1026 .0549 (59)	.0126 .0521 (58)	-.0121 .0210
Uncle Nephew	.1250	F θ	-.0241 .1854* (43)	.1317 .2558* (43)	.1316 .1626* (32)	-.0674 .2575* (32)	.0087 .0500 (32)	.0361 .1823
Aunt Niece	.1250	F θ	-.0339 .1053 (50)	-.0918 .1921* (50)	-.0044 .2070* (40)	-.2316 .1043 (41)	.0509 .2136* (38)	-.0622 .1645
Parent Offspring	.2500	F θ	-.0102 .2503* (1071)	-.0175 .2425* (1071)	-.0310 .2372* (977)	-.0303 .2329* (979)	-.0156 .2509* (918)	-.0209 .2440
Full Sib	.2500	F θ	-.0179 .2484* (786)	.0180 .2745* (786)	-.0436 .2446* (660)	-.0034 .2084* (668)	.0207 .2756* (647)	-.0042 .2503
Complete Kindred	?	F θ	-.0024 .1498*	.0097 .1569*	-.0020 .1340*	-.0015 .1508*	.0029 .1470*	.0027 .1477

*notes estimate significantly different from zero at .01 level of probability.

between allele outcomes within the i^{th} individual, ignoring subdivision, is defined as

$$F = \frac{\sigma_{X_{ij}X_{kl}}}{P(1 - P)} \quad i = k, j \neq l$$

where $\sigma_{X_{ij}X_{kl}}$ is the covariance between alleles on the maternal and paternal chromosomes of the i^{th} individual and P is the allele frequency in the population. A second correlation represents the association between alleles in different individuals but in the same subdivision. It is defined by Cockerham as

$$\theta = \frac{\sigma_{X_{ij}X_{kl}}}{P(1 - P)} \quad i \neq k, j, l = 1, 2$$

where now $\sigma_{X_{ij}X_{kl}}$ is the average of the four possible allele covariances between two different individuals of the same subdivision. Complete kindreds and pairs of relatives (one pair per kindred) are considered in Table 4. In general, the estimates of correlations between alleles from pairs of relatives did not deviate significantly from the expected values $F = \theta = 0$, based on the assumption of random dispersion of alleles to the members of the pairs. As expected, because the analysis based on the entire kindred involves a mix of different degrees of relatives, the average $\hat{\theta}$ within a kindred is large, $\hat{\theta} = .1477$. This estimate provides a measure of the degree to which individuals in the entire sample are not biologically independent sampling units because they are "clustered" into genetically related subdivisions. In other words, on the basis of pairs of relatives, one pair from each kindred, we can state that each kindred is a representative sample of allele frequencies. And, although geneticists expect the value of $\hat{\theta}$ within kindreds to be appreciable, the estimate provides a second kind of information which is a useful quantification of the degree of nonindependence in the entire sample due to biological relationships. One would consider such information when planning the sampling procedure for other scientific studies of this population, especially those which would require biologically independent individuals for a valid analysis of the problem.

V.4.4. An Analysis of the Role Genes Play in Determining Cholesterol Variation in the Tecumseh, Michigan Population (Sing). - Data collected in the community

of Tecumseh have provided a unique opportunity to study genetic variability in an unselected human population. A large number of polymorphic genetic markers, physiological variables, socio-economic variables, and numerous measures of biological and environmental status have been recorded for each individual. In the past two years we have completed analyses of the distribution of the alleles of polymorphic marker loci among individuals in families in an effort to detect evidence for genetic selection. The markers were also used to obtain estimates of nonrandom organization of genotypes by measuring the degree to which allelic variation at one locus is correlated with allelic variation at a second. Recently we have turned our attention to the investigation of the relative role genetic variation at these loci (or loci that may be closely linked) and unmarked polygenes in general may play in determining the total phenotypic variation in certain physiological traits which have been identified as predictors of risk for coronary heart disease. The traits of greatest interest are serum cholesterol, blood glucose, blood pressure, and serum uric acid. We have begun to analyze the cholesterol data. Our intent is to subdivide the total variation in cholesterol among the approximately 14,000 measurements that have been made over the past twelve years on the approximately 11,000 individuals into portions attributable to biological and environmental factors. Our first step has been to estimate the contribution to the total phenotype variation of variation in age in each of the examination periods (rounds); 1959-1962 and 1963-1966. A summary is presented in Table 5. For either sex, thirty percent of the total variation is attributable to variation in age and difference between rounds of physical examinations. The distribution properties of the $\log_e$ of the cholesterol values adjusted for age and round are similar to those suggested by Goldstein et al., (1973) of hypercholesterolemic patients hospitalized for coronary heart disease and their normal relatives. They suggested that the log normal properties (skewness) of values in the population at large are likely the result of a mixture of two groups, one homozygous for a normal allele at a single locus and a second heterozygous or homozygous for a dominant allele at a locus which shifts the cholesterol values to the hypercholesterolemic range. However, our attempts to separate a class of hypercholesterolemic individuals in the Tecumseh data by the procedure of method-of-moments have not been successful in identifying two Gaussian distributions, one a large normal group and the other the sus-

Table 5. Contributions of age and round of physical examination to variability of $\log_e$ cholesterol in the Tecumseh data set.

Source of Variability	Degrees of Freedom	<u>Males</u>		Degrees of Freedom	<u>Females</u>	
		Sum of Squares	% Total SS		Sum of Squares	% Total SS
Total	7075	338.93	100.00	7233	353.56	100.00
Age Regression						
Round I	3	47.95	14.08	3	51.85	14.67
Round II	3	49.00	14.40	3	51.91	14.68
Pooled	6	96.95	28.48	6	103.76	29.35
Round after						
Age Adjustment	1	2.24	.66	1	2.17	.61
Residual	7068	241.19	70.86	7226	247.62	70.04

pected smaller hypercholesterolemic group. We suspect that the discrepancy with the Goldstein study is because we are analyzing normal variation in a population unselected for cholesterol values or a disease entity. The plans for continuation of the analyses of the four variables which are thought to be indicators of coronary heart disease are discussed in the Proposal Section 4.2.

V.4.5. Linkage in the Tecumseh, Michigan, Population (Sing). - In the process of genetic analysis of the community of Tecumseh, Michigan, we have obtained linkage data on twelve marker loci. When considering three values of the linkage parameter, θ , no sum of the logs of the likelihood of the data exceeded unity, hence the informative families available do not change the odds in favor of linkage for any pair of loci considered. However, for a number of combinations there are indications that the peak of the likelihood may be between values of θ equal to 0.0 and 0.3. Summed LODS for those combinations are presented in Table 6 for nine values of θ .

The strongest evidence for linkage from these data are for Rh-Kell, Le-Se, and Rh-Le marker combinations. However, all surfaces are relatively flat and if linkage for any combination given in Table 6 is true, more data will be required to establish an acceptable level of confidence. A linkage of Rh with Kell has not been suggested by previous studies, nor has the Rh locus been associated with the Lewis system or the secretor phenomenon. The seeming association of Le and Se may be artifactual and ascribable solely to the complex interaction that is known to occur between these two loci. A manuscript reporting these results has been submitted to the American Journal of Human Genetics.

V.4.6. The Effect of Density-Dependent Population Regulation on the Age-Distribution of Single-Sex Populations (Smouse, Weiss). - Current theoretical models of population growth are concerned with describing the dynamics and predicting the statics of natural populations from two rather different points of view. The ecological approach describes population growth by logistic-type equations, where population size is regulated by density-dependent feedback. Total population size eventually equilibrates at some upper carrying capacity. The demographic approach (also used in population genetics) describes the dynamics of age-structured populations by means of density-independent age

			Interactive Families								
			.05	.10	.15	.20	.25	.30	.35	.40	.45
	K 11-2	12	-1.37533	0.10473	0.58224	0.72523	0.66979	0.53050	0.34985	0.17245	0.04173
	K 11-1	43	-6.01521	-1.63752	0.04710	0.79414	0.93604	0.81637	0.56374	0.28789	0.08251
	A 11-sec	297	-12.31583	-4.66579	-1.43333	0.17083	0.79727	0.93252	0.81448	0.54052	0.26656
	G 11-1	274	-1.55072	-0.12896	0.34495	0.46661	0.44674	0.33337	0.22741	0.10239	0.04433
	C 11-1	16	-3.11009	-0.72150	0.14146	0.44212	0.49898	0.41599	0.29339	0.15060	0.04762
	K 11-2	2	0.19400	0.15833	0.12495	0.09438	0.06718	0.04391	0.02512	0.01130	0.00784
	P 11-2	15	0.47087	0.38507	0.30271	0.22500	0.15842	0.10168	0.05773	0.02563	0.00646
	L 11-2	14	0.01896	0.01737	0.01479	0.01182	0.00873	0.00587	0.00342	0.00156	0.00039
	G 11-2	7	-0.22863	0.13924	0.18494	0.09127	0.02938	-0.03612	-0.03525	-0.02783	-0.00681
	K 11-2	70	-0.40744	0.29939	0.51260	0.53310	0.46422	0.34963	0.22786	0.11269	0.03062
	L 11-2	281	-0.94232	-0.06457	0.22392	0.32901	0.30430	0.24064	0.15413	0.07725	0.02149
	G 11-2	111	-6.08836	-1.98965	-0.41531	0.30411	0.56432	0.58463	0.47641	0.30887	0.13820
	A 11-sec	65	-1.40518	-0.34200	0.00897	0.12573	0.12022	0.07688	0.04339	0.01860	0.00439
	L 11-2	385	0.66146	0.92099	0.87332	0.71790	0.53823	0.35922	0.20919	0.09154	0.02050
	H 11-2	199	-0.26000	0.46626	0.62104	0.58280	0.47074	0.33576	0.20577	0.09803	0.02335

specific survival and fertility rates. The population ultimately converges to a stable age distribution array. The first approach suffers from ignoring age-specific vital rates, treating the total population as a homogeneous assemblage. The second approach suffers from the absence of density-dependent regulation, allowing exponential population growth.

With the current interest in the unification of population biology (population genetics, population ecology, demography), some effort to combine density-dependent feedback control with age-specific vital rates is needed. Leslie (1948) has applied some very simple density-dependent adjustments to standard demographic models. Charlesworth and Giesel (1972) have applied a simple type of density-dependent feedback to cases of genotypic selection in diploid populations. Neither of these models is detailed enough to be useful in the description of ecologically realistic populations. We have developed a more general treatment of density-dependent feedback in age-structured populations. Each survival and fertility rate of the standard Leslie (1945) matrix is differentially damped by the numbers in each age class. The population converges to a unique age-distribution and stable population size. Different density-dependent damping patterns result in drastically different age distributions. Three examples are shown in Figure 1. These examples indicate that very different age distributions may be expected under different types of density-dependent population regulation.

We are currently working on extensions of the basic formulations. We expect to be able to deal with two-sex populations and multiple genotypes (see Sections VI.4.4., VI.4.5.) and ultimately with the social structure of human populations. We expect that this model will be useful in affecting an interface between population genetics, population ecology, and demography.

V.4.7. Computer Use (Griffith, Sing). - During the past year the IBM 1130 system maintained by this grant has been utilized to capacity. Table 7 breaks down the total hours available on the AEC 1130 system during the period January 1, 1972 to December 31, 1972 by use category. Mathematical and statistical applications accounted for 89.9 percent of the total 8784 hours available on this system during the year.

Because of the demands of researchers associated with this project it was necessary to utilize another IBM 1130 system where free time was available to accomplish our computing goals during this period. A total of 1100

Figure 1 : Stable Age Distributions Under Three Different Density-Dependent Growth Regulation Strategies (I- D.D. Survival Regulation; II- D.D. Fertility Regulation; III- Post Infant D.D. Survival Regulation).

Table 7. A summary of IBM 1130 hours utilized by AEC related research during the period January 1, 1972 - December 31, 1972.
(A total of 8784 hours over 366 days.)

<u>Use Category</u>	<u>AEC 1130 System</u>		<u>Other 1130 Systems</u>	
	<u>Hours</u>	<u>Percent</u>	<u>Hours</u>	<u>Percent</u>
System development (Griffith)	260	.0296		
AEC-related mathematical and statistical use	7896	.8989	1100	1.0000
By Investigator:				
Levin	208.5	.0240		
Li	2934.0	.3340	900	.8182
Rossman	220.0	.0250		
Schull	783.0	.0890		
Sing	1737.0	.1980		
Smouse	49.5	.0060		
Spielman	1564.0	.1780	200	.1818
Ward	400.0	.0450		
Total down time due to storms, IBM, and Calcomp	<u>628.0</u>	<u>.0715</u>	<u>--</u>	<u>--</u>
Total	8784.0	1.0000	1100	1.0000

Total IBM 1130 usage equivalent to 1.125 1130's.

hours for an additional 12.5 percent of available time was met in this way. That option will likely be closed to us in the future. We have also used considerable time on the University's 360 System. Programs have been developed over the past six years to utilize efficiently the 1130 computing system. Hence, an expansion of the present 1130 System would not necessitate time-consuming modifications of the complex programs already developed. To meet the anticipated increase in demand for computing on AEC related projects we have invested considerable effort in evaluating the necessary update to our 1130 system. This material is presented in Section VI.4.7. of the Proposal and Appendix B.

V.5. SUPPORTING LABORATORY AND OTHER INVESTIGATIONS

V.5.1. The Genetic Unit of Selection (Sing, Templeton). - During the past year we have continued to conduct experiments which we believe have a high probability of identifying the genetic unit of selection in Drosophila populations. Our objective has been to estimate the degree to which interaction between gene loci determines the fitness of an individual and contributes to the behavior of allele frequencies for single genetic loci. We have been using the strengths of the natural biological design provided by parthenogenetic Drosophila to 1) identify the levels of genetic organization which act as a unit independent of the influences of other such units and 2) determine the role and types of allele interactions between loci which may enter into the organization of such units.

It is patent that an individual's phenotype determines his adaptation to the ecosystem and his reproductive fitness. All biologists realize that the entire genome contributes to the determination of the phenotype. But in the past it has been the belief of many population geneticists (see Crow and Kimura, 1970) that we may be able to use small parts of the genome, even allelic variations at single loci, to predict the adaptation and fitness of an individual. Most multi-locus formulations have been constructed by combining the single locus effects in either an additive or a multiplicative fashion.

The reasons for an oversimplified genetic treatment of phenotypes which are obviously determined by complex integrations of genes and environments have been both theoretical and practical. Firstly, a mathematical theory

has not yet evolved which gives an adequate treatment of the integration of many locus effects in sexually reproducing populations. Secondly, from the standpoint of a practical application of available population models to analysis of data, one is faced with problems of estimation. If we accept the multi-locus formulations which are presently available, it is extremely difficult to compute meaningful estimates of measures of gene interaction.

Analyses in population genetics have been influenced by the theoretical models which have been developed. It has seemed obvious that a consistent theory should build from the gene locus. However, a consistent approach to estimating the parameters of the relationship among loci to form higher order multi-locus genotypes has not been achieved (see Turner, 1971). As a rule the multiplicative interaction model has not been adequate in describing the deviations from additivity of locus effects. More recently, Wills (1970), and Franklin and Lewontin (1970) have suggested that the formulations for genetic load are inconsistent with the observation of extensive polymorphism for the genome because such formulations derive from unrealistic considerations of the single locus as the unit of selection. They suggest that the "cost" to the population of maintaining allelic variation is dramatically reduced if there is an interaction between linkage and epistasis to reduce the number of loci which are independently segregating in the population. The consequence would be a segment, "unit", of the genome representing a combination of nonalleles which survive more frequently than other combinations of alleles at the loci involved. The inadequacy of the current theory 1) to describe properly the multi-locus genotype and 2) to give predictions consistent with the observed polymorphic allelic variation may be corrected if models could be built which properly reflect the unit of genetic integration.

But what is the unit of selection? It is one matter to be aware of the problem, it is another matter to be able to conduct experiments to critically evaluate and to test hypotheses regarding the nature of the role of gene loci in forming such integrated units of selection. Is the gene locus the unit of selection and should we continue to develop models which reflect this? Probably not! To build realistic models which predict the behavior of genic variation in populations over time, it is extremely relevant to have knowledge of the "chunk" of genetic material which behaves as a unit.

Many studies of allelic variations in higher plants and animals have

attempted to define the kinds of forces operating in populations to maintain polymorphic frequencies. Our previous experimental efforts have indicated that the number of isozyme forms in wheat is not correlated with ploidy of the species (Sing and Brewer, 1969). It has been shown that the homeologous chromosomes of the hexaploid species, T. Aestivum, each code for electrophoretically similar isozymes for eleven of twelve enzymes studied (Brewer, Sing and Sears, 1969). Both outcomes are inconsistent with random substitution of alleles implied by the selectively neutral gene hypothesis popularized by King and Jukes (1969). A study of ten enzymes in 22 plant species (Sing and Brewer, 1971), indicated that the number of electrophoretically different forms of some enzymes may not be independent of the number for a second enzyme. The correlation was greatest for the enzymes of the set studied which are known to be related metabolically.

These and experiments by others, coupled with theoretical developments in the field of population genetics, have indicated that proper experiments are difficult to conduct because

- 1) there is a multiplicity of parameters, both demographic and biological, operating simultaneously which tend to be confounded, and
- 2) the unit of selection is the phenotype (genotype), hence a proper study designed to weigh the consequences of the organization of the genotype requires special biological designs. It is patent that field studies are extremely important, yet if one cannot define which variables are relevant measures of the genetic system the data so collected provides an uncritical test of the neutral hypothesis. We will discuss below why the study of parthenogenetic Drosophila may provide insight into identifying the relevant measures.

Research by members of our group has already begun to establish the theoretical framework for studies using parthenogenetic organisms. While a student, Asher (1970a, 1970b) developed deterministic equations to describe the expected decay of variability in parthenogenetic organisms as a function of mode of restoration of diploidy, selection, and map location of the locus. He showed that heterozygosity could be maintained for certain combinations of values of these parameters. A continuation of the theoretical work by Templeton and Rothman (in press) and Templeton (1972) to include a stochastic treatment of the variables involved has suggested experiments which will give unbiased estimates of selection coefficients for single loci and sensitive tests of the hypothesis of selective neutrality. Theoretical work completed by

Templeton (1972) provides the outline for an experimental strategy for obtaining information on the effect of epistasis and map location on the maintenance of inherited multi-locus variation in parthenogenetic species.

This preliminary theoretical and statistical framework developed by students associated with this project allows the direct estimation of fitnesses and epistatic terms for any multilocus genotype in a one-generation experiment, and the viability component of fitness in a multi-generation experiment. Also an extensive hypothesis testing framework has been developed for these problems (see Section VI.4.2. for a brief summary).

Experimental work has been done with sexual and parthenogenetic strains of Drosophila mercatorum. D. mercatorum is a widespread tropical sexual species. Parthenogenetic strains have been selected from wild caught flies by Dr. H. L. Carson and have lived only in a laboratory environment. The response to selection for parthenogenesis indicated to Carson (1967) that parthenogenesis was a multigenic trait and its consequences resulted in the reorganization of much of the genome. Later work by Carson et al. (1969) demonstrated that these parthenogenetic strains are diploid and automictic, i.e., they retain normal meiosis. Diploidy is primarily restored by post meiotic fusion of two cleavage nuclei derived from a single haploid egg nucleus (see Figure 2). This has been termed "gamete duplication" by Carson. The result is total homozygosity in one generation. The actual percentage of eggs undergoing gamete duplication varies from strain to strain and range from 99% to 78% (Carson, personal communication). In the remaining eggs either central fusion or terminal fusion of the meiotic pronuclei restores diploidy (see Figure 3). Recent evidence obtained in our lab indicates that restoration is mostly due to central fusion in those eggs not undergoing gamete duplication.

Two parthenogenetic stocks were obtained from Dr. Carson; S-1-Im and O-3-Im. S-1-Im was derived from flies caught in El Salvador while O-3-Im comes from Oahu, Hawaii. The lower bound for the percentage of gamete duplication was estimated as .94 for S-1-Im (Templeton and Rothman, in press) and as .92 for O-3-Im (Carson, personal communication). Consequently, the pressure to develop total homozygosity is great in these stocks. To insure that each stock consists of a single, totally homozygous genome, both stocks were separately put through 8 single female generations to obtain two isogenic stocks, Iso-8-S-1-Im and Iso-8-O-3-Im.

The ancestors of the Iso-8-S-1-Im and Iso-8-O-3-Im (hereafter

WITH RECOMBINATION

WITHOUT RECOMBINATION

GAMETE DUPLICATION

FIGURE 2

Diagrammatic representation of the meiotic events of gamete duplication. The proportion of eggs reproducing by this mechanism is represented by E_3 in the preceding equations. A single locus with two alleles (A and a) is considered where Y is the probability of a recombination between the locus and kinetochore.

FIGURE 3

Diagrammatic representation of the meiotic events of central and terminal fusion.

referred to as S and O respectively) reproduced sexually but for the past eleven years the S, and for seven years the O, flies have been reproducing primarily by gamete duplication. In other words these stocks have existed in a genetic environment of total homozygosity for the entire genome. A natural hypothesis to make is that the alleles which have been fixed in the genomes that exist in these two parthenogenetic stocks have each adapted to a totally homozygous genetic environment. An isozyme survey of 17 gene loci revealed the two strains to be different at 6 of these loci. Such data imply that the O and S stocks differ at approximately 35% of their loci. We begin with the hypothesis that the O and S parthenogenetic stocks have each adapted in genetically distinct fashions to the genetic environment of total homozygosity.

The experiments being conducted are designed to 1) test this hypothesis and, if such an adaptation does exist, to 2) describe the adaptive mechanism and unit of selection for this system. The essential design is to perturb the O and S homozygote genomes away from their original state and then observe, with respect to several marker loci, the effect of such a perturbation in a totally homozygous genetic environment. An outline for such an experiment is given in Figure 4.

The parthenogenetic F_2 flies are scored for the marker loci at adulthood and the frequencies of the various homozygous genotypes are determined. Since these parthenogenetic F_2 flies shared with the sexual F_2 flies genetically identical mothers who had normal meiosis, and differ only in the post-meiotic diploid restoring mechanism, any difference in the genotype frequencies occurring in the parthenogenetic F_2 from the maternally derived component of the genotypes in the sexual F_2 must be due to viability differences that occurred subsequent to restoration of diploidy. Under our assumption that selection is absent in the sexual F_2 , such viability differences are attributable to the nature of the genetic environment associated with the parthenogenetic F_2 .

A number of predictions can be made at this point. First, if there was no adaptation to a totally homozygous environment in O and S and the alleles represented in their genomes are selectively neutral, no differences in viability should be detected among segregants in the parthenogenetic F_2 . Second, if such an adaptation to total homozygosity was made but was accomplished purely through additive genetic effects, there should be no viability

Score both F_2 's for 2 visible and 4 isozyme loci and identify the types of maternal chromosomes the F_2 's received.

FIGURE 4
Basic experimental design

differences among the parthenogenetic F_2 . However, if the adaptation to total homozygosity exists and if it involves interactions between non-alleles, the predictions are quite different. The O and S parental genomes under such a hypothesis are taken to represent coadapted gene complexes. These complexes are broken down in the F_1 female by meiosis. By chance alone, some of the genotypes of the parthenogenetic F_2 flies will be very close to a parental type while others will be far removed. Under a coadaptation hypothesis, the closer a parthenogenetic F_2 fly's genotype is to a parental, the higher its fitness should be while the more recombinant its genome, the lower its fitness. One prediction under this hypothesis of interacting gene loci is that the fitnesses of "parental" genotypes relative to non-parental genotypes should increase as the number of loci used to determine "parental" type increases. This occurs simply because the more loci one simultaneously used, the more of the genome one is marking. Consequently, a fly that is parental type with respect to two loci will on the average be farther removed from the original O and S types than an F_2 fly that is parental type with respect to five loci.

Under the hypothesis that the adaptation to total homozygosity is due to interactions among loci, several other predictions can be made. If the interactions are between linked loci only, recombination but not segregation should create viability differences among the parthenogenetic F_2 . On the other hand, if the interactions are between unlinked loci only, segregation but not recombination would affect the fitnesses of the F_2 . Finally, if the adaptation involves interactions between both linked and unlinked loci, recombination and segregation should both affect the fitnesses of the parthenogenetic F_2 flies. A summary of the linkage among the loci studied is given in Figure 5.

During the past year data was collected from an experiment using the above design. The results of this experiment are of great interest. With respect to the single locus data, there was no evidence for selection. That is, in the sexual F_2 we expected both F_1 , O, and S type alleles to be found in equal frequency, and they were. In the parthenogenetic F_2 we expected both parental homozygotes to be present in equal frequency and they were. For two or more loci, the situation changes drastically. Table 8 gives summary of these data. The data were first analyzed with respect to two locus systems. Using six marker loci, there are 15 ways of pairing loci. For all

Figure 5

The genetic map for two of the four major chromosomes of D. mercatorum.

TABLE 8

DEVIATIONS OF THE PROPORTION OF PARENTAL (S-Im and O-Im) GENOTYPES FROM
THE EXPECTED PROPORTION IN THE PARTHENOGENETIC F₂ AND THE SEXUAL F₂

Parthenogenetic F ₂				Sexual F ₂		
	Number of j locus systems using 6 markers	Average Deviation of the proportion of parental geno- types from the ex- pected proportion	Average Proportional deviation	Number of j locus systems with signifi- cant (.05) de- viations	Number of j locus systems	Average deviation Average proportional deviation
2 Locus:						
unlinked	8	.05516	.1103	5	8	.0098 .0167
linked	7	.05290	.0964	4	-	- -
total	15	.05411	.1039	9	8	.0098 .0167
3 Locus:						
2 linked, 1 unlinked	16	.0857	.3003	14	16	.0111 .0379
3 linked	4	.0732	.1814	3	-	- -
total	20	.0832	.2765	17	16	.0111 .0379
4 Locus:						
systems with un- linked markers	14	.0956	.5218	13	14	.0118 .0627
4 linked	1	.0693	.2227	1	-	- -
total	15	.0939	.5018	14	14	.0118 .0627
5 Locus:	6	.0949	.7638	6	6	.0128 .1015

fifteen pairs, the proportion of O and S parental genotypes was calculated for both the parthenogenetic F_2 and the sexual F_2 . The knowledge that the segregation ratios between unlinked loci will be 1/2 was incorporated whenever possible. As Table 8 shows, there was always an excess of parental types in the parthenogenetic F_2 . No such excess was detected in the sexual F_2 for those systems that could be tested (i.e., those that involved segregation). For example, for unlinked two locus systems, the average proportion of parental types in the parthenogenetic F_2 was .55516 as opposed to an expected value of .5 while for the same systems in the sexual F_2 it was .5098. This indicates that there is selection favoring the parental types in the parthenogenetic F_2 , but that no such selection is operating in the sexual F_2 . As one goes to other systems involving more loci the same pattern holds but the magnitude of the parental excess greatly increases as the number of loci simultaneously analyzed increases. Thus, by the time we are dealing with 5 locus genotypes, on the average there is a 76% excess of parental types in the parthenogenetic F_2 as compared to the sexual F_2 . The level of significance for some of the 5 locus combinations is on the order of 10^{-9} . This indicates that extremely large selective forces are operating in the parthenogenetic F_2 --as evidenced even by the 2 locus unlinked systems. Moreover, the gene is not the unit of selection because when each locus was treated individually, no selection was detected. The evidence for higher order interactions leading to a co-adapted genome is strengthened by the fact that the more loci simultaneously analyzed, the greater were the deviations from the null hypothesis of neutrality.

A more detailed analysis of the fitnesses of the unlinked two locus systems in the parthenogenetic F_2 was done using the techniques given in Templeton (1972). The results from three of the systems are given in Table 9 which were representative of all eight. The fitnesses were first estimated under the "unrestricted" hypothesis space. Under this hypothesis space, the fitnesses of all multi-locus genotypes are measured directly. A measure of epistasis for two unlinked loci is simply the fitnesses of the parentals added together minus the sum of the fitnesses of the recombinants. For all combinations this was positive and for all but one this measure was quite large. The fitnesses were then estimated under an additive fitness model which allows no epistasis. Finally, although not shown, the fitnesses were estimated under a multiplicative model which does generate some epistasis.

TABLE 9

FITNESS ANALYSIS FOR SOME TWO LOCUS SYSTEMS INVOLVING UNLINKED MARKERS

System	Fitnesses of:				$c = 1 + W_2 - W_3 - W_4$	$\bar{W}$	Log Likelihood Ratio Tests	d.f.	P-value
	S-Im	O-Im	Rec. I	Rec. II					
G-G-PD & v*	Sv	F+	Fv	S+			$2(t_{\Omega} - t_0) = 6.62$	3	$.10 > P > .05$
	Ω 1	.96	.65	.78	+.45	.8225	$2(t_{\Psi} - t_0) = 2.09$	2	$.50 > P > .10$
	Ψ 1	.96	.85	1.11	0	.93	$2(t_{\Omega} - t_{\Psi}) = 4.53$	1	$.05 > P > .01$
sl & v*	sl v	++	+v	sl +			$2(t_{\Omega} - t_0) = 16.70$	3	$P < .001$
	Ω 1	1.34	.85	.73	+.73	.98	$2(t_{\Psi} - t_0) = 4.71$	2	$.10 > P > .05$
	Ψ 1	1.34	1.26	1.08	0	1.17	$2(t_{\Omega} - t_{\Psi}) = 11.96$	1	$P < .001$
sl & Est A	sl S	+F	+S	sl F			$2(t_{\Omega} - t_0) = 3.85$	3	$.05 > P > .025$
	Ω 1	1.28	1.15	.82	+.31	1.06	$2(t_{\Psi} - t_0) = 5.96$	2	$.10 > P > .05$
	Ψ 1	1.28	1.33	.95	0	1.14	$2(t_{\Omega} - t_{\Psi}) = 2.89$	1	$.10 > P > .05$

* These systems had a significant deviation of the proportion of parental (S-Im and O-Im) genotypes from the expected proportion of .5.

Ω denotes the "unrestricted selection" parameter space. Under this space the two locus fitnesses are estimated directly and no restrictions are placed on the additive and epistatic components of fitness.

Ψ denotes the "additive selection" parameter space. Under this space all fitness differences are forced into an additive model with no epistasis.

These various models were then tested against the null hypotheses of no selection and against one another. For example, *sl* and *v* represent a two locus system with a large amount of epistasis. The "unrestricted" fitnesses indicate some selection against the double recessive parental type, but even more intense selection occurs against the two recombinant genotypes. Consequently, segregation of unlinked genes alone breaks up the co-adapted complexes and lowers fitness. When unrestricted selection was tested against the null hypothesis the test greatly favored selection. However, additive selection was not favored over the null hypothesis, and multiplicative selection (not shown) gave an even worse fit than additive selection since negative epistasis was generated.

The selection models were then compared to one another with the result that the additive and multiplicative models were rejected in favor of the unrestricted model. The interpretation of the results thus far is that almost all the selection present in this study is due to an epistatic term different from that defined by additive or multiplicative effects.

The conclusion is that the O and S genomes are indeed adapted to total homozygosity, and that this adaption is a true coadaptation involving interactions between both linked and unlinked gene complexes. Another interesting observation is that no selection was detected with one locus systems, but very large selective forces were detected when multi-locus systems were treated as the unit of selection. This should serve as a warning that selective "neutrality" may be an artifact of using an incorrect unit of selection for a given biological system.

V.5.2. Genetic, Chemical and Serological Studies of Human Immunoglobulins (Gershowitz).

V.5.2.1. Further Subdivision of the Inv System. - The Inv system includes three allotypic marker antigens of the kappa light chains of humans. Three alleles are known; *Inv*¹, *Inv*^{1,2} and *Inv*³. It is known that there is an amino acid substitution of leucine at position 189 in the *Inv*(1,2) kappa chain for valine in the *Inv*(-1,-2) chain. But since most light chains which are *Inv*(1) are also *Inv*(2), the relationship between these two antigens and the single amino acid difference is not known.

We have recently discovered an anti-Rh which, while first believed to be an ordinary Inv(1) coat and reactive therefore with anti-Inv(1), has been shown to differentiate anti-Inv(1) sera. The sera Iesp and Sim both contain anti-Inv(1) antibodies and with the anti-Rh coat Abr gave identical results in hundreds of tests of sera from several populations. A second anti-Rh coat, Kuh, gives the usual Inv(1) pattern with Iesp but red cells coated with Kuh cannot be agglutinated by Sim. We have shown that Inv(2) specificity is not involved so that this combination defines a new antigenic specificity, one possibly associated with Inv(1).

V.5.2.2. Idiotypic Specificity of an Anti-Rh Antibody. - Every myeloma protein or homogeneous antibody preparation can be characterized by having a unique antigenic specificity--its idio^{type}. Idiotypic specificities are so unique that they are used to define the product of a class of immunocompetent cells derived from a single ancestral cell. Attempts are currently being made to study the genetic control of variable regions of certain antibodies by studying whether parent and offspring, both injected with the same immunogen, can produce antibody of the same specificity and with the same idio^{type}.

Normally the injection of a globulin preparation from a normal individual into a rabbit would not be expected to elicit an anti-idio^{type}, for the concentration of any one molecular form in the preparation would be so low so as to preclude the possibility that it could be immunogenic. Thus, red cells coated with anti-Rh antibodies (which are heterogeneous even from a single donor) elicit, normally enough, anti-IgG in the rabbit but not an anti-idio^{type}. A myeloma preparation, homogeneous in molecular composition, elicits anti-Ig and an anti-idio^{type} antibody.

It is generally accepted that myeloma proteins arise as a result of an overproduction of immunoglobulin by a hyperplastic clone of lymphocytes, whose ancestor cell was producing the self-same protein merely as one of many such cells all producing their individual proteins. But if all immunoglobulins are antibodies, it must be presumed that myeloma proteins are antibodies even if the immunizing antigens are not known. One might expect to find a myeloma-like antibody, that is, an antibody produced by the clone derived from a single responding immunocompetent cell, if the clone

become sufficiently large without becoming cancerous. Although homogeneous antibodies (antibodies of restricted, but not unique, primary sequence) have been produced in rabbits, to our knowledge there has been no report of a monoclonal, non-myeloma, antibody. We believe the anti-Rh antibody described below may represent the first such report.

As part of our efforts to produce anti-Gm antibodies in the rabbit, anti-Rh was coated onto red cells and injected into rabbits. One anti-Rh (Spr) coat was discovered to have elicited an anti-idiotypic in all three rabbits injected with this anti-Rh antibody. The reaction was identified as being idiotypic for this anti-Rh antibody by the facts that none of 250 anti-Rh sera could inhibit the reaction between the anti-idiotypic and red cells coated with the anti-Rh, while homologous serum (Spr) did inhibit and removal of the anti-Rh antibody by simple absorption with Rh-positive red cells removed the inhibiting capacity of the serum. Thus, the idiotypic reaction was associated with the anti-Rh molecule. Neutralization of the serum with increasing amounts of anti-idiotypic antibody had no apparent effect on anti-Rh titer, suggesting that the two sites, anti-Rh and idiotype, were distinct.

Since it is normally expected that anti-Rh is produced even in the same individual in several different molecular forms, we set out to determine whether the idiotypic specificity could be found on all the forms of anti-Rh in the Spr serum. The serum had originally, by a routine antibody screen, been identified as an anti-CD, that is containing anti-C and D antibodies. During an attempt to fractionate these antibodies we realized that this serum actually contained only one antibody, known as anti-G. When anti-G is found, it is usually found in the presence of anti-D or anti-C or both, but absorption of this serum by the rare $r^G r$ (Rh negative, G positive) blood removed all anti-Rh activity, demonstrating that only one anti-Rh specificity was present.

This antibody had been used in our usual typings for Gm^a as the standard Rh coat for this IgG_1 heavy chain marker although we had believed that the serum also possessed the weak ability to confer Gm^b , the IgG_3 heavy chain marker. Careful reinvestigation now led to the realization that only one marker, Gm^a , could be conferred by this antibody, and so, since the phenotype of the donor serum was $Gm(ag)$ the anti-Rh was of only one of the two potential types.

It is also generally accepted that in monoglobulin heavy chains of all classes occur in combination with kappa or lambda light chains and that both kinds of light chains will be found in any "normal" antibody preparation. Tests of red cells coated with anti-Rh sera indicated in fact that the vast majority of anti-Rh do react with both anti-kappa and anti-lambda. The anti-Rh Spr, however, completely failed to react with anti-lambda.

Thus, the anti-Rh Spr has a single (albeit highly unusual) anti-Rh specificity, confers a single heavy chain marker and is of a single light chain type. All of these facts, together with the highly unusual fact of the production of an anti-idiotypic, led to the suspicion that the idiotype specificity is associated with a monoclonal antibody.

It is still possible that two populations of anti-Rh antibody exist in the serum, one monoclonal and one not. The demonstration that only one antibody type exists could be demonstrated by the complete removal of anti-Rh activity by the rabbit anti-idiotypic. Experiments to effect this proof have not been successful; solidification (gellation) of the anti-idiotypic serum by glutaraldehyde causes the antibody to react with other anti-Rh which are clearly non-reactive when the antibody is in the soluble form. Since it is essential to absorb with a particulate antibody, we intend to try coupling the anti-idiotypic to Sepharose with subsequent passage of the anti-Rh serum through an immunoabsorbent column.

The unambiguous identification of this antibody as the product of a single clone of cells will have important implications in understanding the relationships between myeloma protein production and normal antibody production, and in estimating the protein output of a single clone under normal in vivo conditions.

V.5.2.3. Gm Antigens in Lymphocyte Culture Media (Kohler and Gershowitz). - Differences in Gm allotypic antigens of human IgG heavy chains have been shown to be associated with amino acid substitutions and the causative genes have therefore been considered to be allelic forms of a structural locus. In 1967, Lobb, Curtain, and Kidson, and again in 1968, Lobb, presented evidence that apparent phenotypic differences in Gm types may not be due to the presence or absence of structural genes but that each individual carries the genes for many Gm markers, whose phenotypic expression depends upon the segregation of regulatory genes controlling synthesis of gamma-G heavy chains.

Using, like stimulated short term cultures of lymphocytes and fluorescein conjugated anti-Gm^a and rhodamine conjugated anti-Gm^b they found that lymphocytes from all donors regardless of serum Gm phenotype showed two populations of cells, one Gm(a+) the other Gm(b+)

Rivat, Gilbert and Ropartz (1970) demonstrated, in the spent culture medium of PHA stimulated short term lymphocyte cultures, that an occasional culture medium possessed a Gm antigen not found in the cell donor's original serum.

In contradistinction to the results of these studies of cells in culture, Fröland and Natvig (1972) demonstrated by specific uptake of labelled anti-Gm antibody by fresh peripheral blood lymphocytes, that only antigens corresponding to the donor's serum phenotype were found on the lymphocytes. It would seem that incubation in culture in the presence of PHA may reverse the regulatory inhibition.

The present study seeks to confirm the work of Lobb, Curtain and Kidson and of Rivat, Gilbert, and Ropartz and to extend such observations by inquiring into the nature of the regulation of immunoglobulin synthesis, using the Gm-markers a, x, g, f, and b as labels for the G1 and G3 heavy chains of IgG. These studies make use of long-term lymphocyte lines (from Dr. A. D. Bloom's laboratory), which have the advantage over short-term cultures in that 1) they can be maintained indefinitely, 2) they can be cloned, and 3) large amounts of cells and spent culture medium can be obtained.

Preliminary results confirm the observations of Lobb, et al and Rivat, et al. The culture media from three cell lines established from 2 donors with serum Gm-type (f+b+) have been typed as follows: 1) Gm(g+), 2) Gm(a+) and later Gm(b+); and 3) Gm(a+b+). Thus, lymphocytes in long-term culture also appear to be capable of synthesizing IgG with Gm specificities not present in the donor's serum.

The culture media from the first two lines were tested by immunoelectrophoresis against anti-IgG and gave a monoclonal precipitin line similar to those seen in the sera of multiple myeloma patients. The culture medium from the third line gave an apparently bclonal precipitin line on immunoelectrophoresis. This suggests that a single clone of cells within the heterogeneous population was synthesizing the IgG of the particular Gm type. Culture medium collected from these lines at later date

However, being positive for Gm A, X, G, F, and D. Although IgG is present in these later cultures, a diffuse line is obtained on IEP, indicating the presence of a heterogeneous population of IgG molecules. It is not known whether this IgG is IgG₂ or IgG₄ subclass and thus not typable for Gm markers, or whether some alteration of the Gm antigenic site has occurred, or whether production of IgG₁ or IgG₃ corresponding to the Gm-specificities which can be identified, is too low to be detected.

Culture media pools of 3-4 subculture transfers from additional lymphocyte lines have also been Gm-typed, with the following general results:

1. Serum type, (agfb); culture medium, (a+); immunoelectrophoresis, monoclonal
2. Serum type, (axgfb); culture medium (g+b+); immunoelectrophoresis, possibly biconal.
3. Serum type, (agfb); culture medium (b+).
4. Serum type, (ab); culture medium (a+).
5. Serum type unknown at present; culture medium Gm(b+); immunoelectrophoresis-heterogeneous IgG.
6. Serum type unknown at present; culture medium Gm(f+).
7. Serum type unknown at present; culture medium Gm(f+); immunoelectrophoresis-heterogeneous IgG.
8. Serum type unknown at present; culture medium--no Gm type; immunoelectrophoresis heterogeneous IgG.

The possibility that a single cell line is producing the IgG in at least some of these cultures is suggested by the finding of monoclonal precipitin lines on IEP and by the fact that, in cultures in which Gm antigen was detected, 9 media possessed only one antigen at a time, and only two possessed two antigens simultaneously. Since all individuals possess at least two serum Gm antigens (one for IgG₁ and one for IgG₃) equal activity of all IgG-producing cells should have resulted in the finding of two antigens in all cultures.

Confirmation of the finding of anomalous Gm antigen on short term cultured lymphocyte cells and in short term cultured lymphocyte media has been obtained in these studies of long term cultured lymphocyte media. The body of evidence is sufficiently impressive to raise serious

questions about the nature of the Gm genes.

V.5.3. Population Studies of Microbial Ecosystems (Smouse). Our experiments for the past year have been primarily addressed to a consideration of the rate of gene substitution in haploid populations and to the "cost of natural selection". Haldane (1957) claimed that the maximal rate of gene substitution must be small; otherwise the accumulated genetic death resulting from the substitution would be incompatible with population survival. The rate of substitution would increase with the advantage of the new mutant, and the greater the advantage of the new mutant, the more precarious the future of the population. Recent theoretical developments have raised serious questions about the validity of the argument (Feller, 1967), but little experimental evidence exists on this point.

A lactose negative strain of *E. coli* K-12 is grown to maximal density on a liquid medium consisting of minimal salts, glucose, arabinose, and varying concentrations of lactose. Lactose concentration is the limiting factor in population growth and final density, but since the Lac^- strain is not responsive to lactose, the final population density of the Lac^- culture is the same at all lactose concentrations. An isogenic Lac^+ strain is added, at an initial frequency of .001, and begins to replace the Lac^- strain, due to its ability to utilize all three sugar sources. Figure 6 shows the rate of replacement for four different concentrations of lactose. The forty (40) day time span of the experiment represents about 133 generations.

From the trajectory of the frequency of the Lac^+ strain, it is possible to estimate the selective advantage enjoyed by this genotype in each set of experiments. These advantages are 6.2%, 7.1%, 8.5%, and 12.0% for lactose concentrations of .00, .08, .16, and .25 gm/liter. The selective advantage and the rate of substitution are increasing functions of lactose concentration. The initial trajectories of total population size for the evolving mixes are shown in Figure 7. Although total size oscillates, the oscillations were progressively damped, and total size never dropped below that of the pure (Lac^-) control. In each case, the total population size finally equilibrates at the value attained by the pure (Lac^+) strain: 103%, 145%, 165%, and 185%, respectively for .00, .08, .16, and .25 gm. lactose/liter.

These experiments demonstrate that population size increases during genetic substitution, rather than decreasing, as could be predicted

ure 6: Gene frequency changes in mixed cultures of wild type (Lac^+) and lactose negative (Lac^-) strains, as a function of lactose concentration in a basal medium consisting of mineral salts, .05 gm glucose, and .20 gm arabinose/liter.

Fig. 2.7: Relative population sizes of evolving mixed cultures of Lac⁺ and Lac⁻ strains, as a function of lactose concentration in a basal medium consisting of minimal salts, .05 gm glucose, and .20 gm arabinose/liter. Pure cultures of Lac⁻ are defined to have a relative size of 100.

by Haldane's postulation that the amount of the improvement and the rate at which it is achieved increase with the adaptive superiority of the mutant (Lac^+ here). It appears that the rate of evolution is limited only by the adaptive superiority of the available mutants, and that Haldane's argument is not generally valid.

V.5.4. Admixture Studies and the Detection of Natural Selection (Ward, Adams). - Ward and Julian P. Adams (Botany Department) have examined the genetic data available for 5 U.S. Black populations to see if there was any evidence for natural selection in these populations with respect to 8 different genetic systems. The initial prerequisite for such studies is an adequate estimate of the ancestral gene frequencies. This estimate was made for the African population, by taking account of the relative contribution of various geographic areas to the slave trade. Attention was also paid to the variance associated with admixture estimates, and to the possible sources of bias in estimating admixture. Three methods which have been devised by various investigators to detect selection were then applied to these data. The final conclusion was that the available data gave no consistent evidence for the action of natural selection in these populations. Since data on the HbS allele were available for some of these populations it appears that the ability of any of these three methods to detect selection is somewhat limited.

V.5.5. Continuing Participation in the Cytogenetic Studies of the AFCC (Bloom). - Progress in this area is summarized in Appendix C, which is a reproduction of Bloom's Trip Report to AEC following 3 weeks in Japan in early 1973.

VI. PROPOSED PROGRAM

VI.1. THE CONTINUING PROGRAM ON THE AMERICAN INDIAN

The important finding, that the mutation rate gene/generation resulting in biochemical variants may for the relatively unacculturated Indian tribes of South America approach 10^{-4} , rather than the rate of 10^{-5} or even 10^{-6} thought from other types of studies to be an appropriate guideline, has forced us to reconsider our projections for field work. We had thought we had completed our field work among the Indians, but now the analysis of a further series of Indian blood samples for rare protein variants is highly desirable. We therefore now wish to project a period of intensive field work in the early summer of 1974. Two parallel developments have facilitated this decision. The first is the failure of the program in Japan to proceed on schedule; had it gone forward as projected, this field work would have been impossible. The second is the fact that the Alpha Helix, the NSF's floating laboratory for general research purposes, is scheduled to be on the Amazon the summer of 1974. We have been approached concerning our interest in availing ourselves of its facilities, and have submitted an application to that end (Appendix 1).

In this section, which is the portion of the total program funded by NSF, we will divide the projection for next year into three portions, namely, the field work, the continuing analysis, and the population simulation. It is a propos to recognize at this juncture that this represents the last year in which we expect this portion of our total program to be funded as a component of the International Biological Program. Most of our analyses of the demography and population structure of the Indian will be completed this year. On the other hand, if the field work is successful, we will have a large backlog of specimens for analysis. In round figures, we project approaching NSF for "terminal support" during 1974-1976, of about half the amount of the current support.

VI.1.1. Further Field Work. - In the past, we have prided ourselves on the multidisciplinary nature of our field work. For 1974, our objectives

are much more limited than previously: we will collect as many blood samples as possible, from relatively unacculturated Indians, for the determination of the frequency of rare variants and an indirect estimate of mutation rates. The activities of the Alpha Helix will be a joint U.S.-Brazil undertaking. We propose to collaborate closely with Professor Francisco Salzano of the University of Rio Grande do Sul, Brazil's chief authority on the genetics of the South American Indian, and a long-time associate. Details are still being worked out. However, we propose to begin our collections among two rather large groups with which we are already familiar, the Macushi and Wapishana of the Roraima Territory, in Northern Brazil. It is anticipated that for this purpose the Alpha Helix will ascend the Rio Branco as high as the impassable rapids at Caracarasí, and we will fly out of the airstrip there. After 2-3 weeks on this location, we will proceed up either the Rio Negro or Amazon, depending on plans to be worked out with Salzano. Attention is directed to the efforts of the Brazilian Government to open up this general area through its program of road building. Now would appear to be an unusually propitious time for this sort of undertaking, since new tribes are being contacted. Furthermore, it is quite likely that if we develop the proper credentials, we can anticipate important support from the Brazilian Air Force, which is quite active in the work of opening up the interior.

While the major emphasis is on the collection of blood samples for a very specific purpose, the estimation of mutation rates, it is clear these samples will also be useful for a variety of other purposes--clarification of tribal relationships, degree of intratribal heterogeneity, etc. Furthermore, we anticipate several special studies. One is the collection of anthropometric measurements on selected tribes, to further a subject in which we have an established interest, the comparison of dendrograms based on genetic markers with those based on physical measurements. Furthermore, we see this as an opportunity to extend our knowledge concerning glucose metabolism (glucose tolerance curves with accompanying insulin determinations), electrolyte metabolism (see Progress Report), and possibly other matters of biomedical interest.

VI.1.2. Analysis of Data at Hand on the American Indian.

VI.1.2.1. Implications of Frequency of Occurrence of Rare Variants in Indian Villages (Li, Neel, Smouse). - As noted earlier, to date at least 9 rare variants have been discovered in a total of 65 Indian villages. For each village, the number of occurrences of each variant (1, 2, 3, ..., n) and the total population sampled from each village are known. The simulation program which led to the derivation of $\bar{t}_0$ (see Sect. V.1.6.) also resulted in a distribution of mutants by village, the mutants by definition all being neutral from the standpoint of natural selection. Finally, Ewens (1972) has derived the expected distribution of numbers of neutral mutants in a population in the generations intervening between introduction by mutation and random loss. We propose to compare the observed distribution with those predicted by theory and by simulation, with the thought that significant discrepancies might be an indication of the action of selection. It is already clear that the execution of this project will encounter a number of mathematical difficulties, but the issue is of sufficient importance that a considerable investment of time seems justifiable.

VI.1.2.2. Extensions and Applications of Matrix Theory (Ward). - Considerable extension of the work on the effect of migration in a tribal population is anticipated for the coming grant year. Our preliminary studies have indicated that the models commonly utilized to investigate the magnitude of the effect of migration are not wholly applicable to the processes occurring within the Yanomama. Since the processes observed in the Yanomama are likely to be similar to the processes that occurred in the majority of tribal populations, and since the maintenance of the high degree of genetic diversity so commonly observed in these populations can only be fully understood in the context of the effect of migration, the formulation of a more adequate model of migration appears as a necessary prerequisite to further analysis of the genetic parameters of this population. Following the initial progress described in Section V.1.5, three aspects of migration will be considered.

a) Migration matrices: As an extension of the model developed by Bodmer and Cavalli-Sforza (1968), an attempt will be made to define a set of operations that will allow changing migration probabilities from one generation to another as well as allowing village fissioning and fusion. The only previous attempt at extending the basic model was made by Wagener (1973), who following a suggestion by Ward and Neel (1970), took account of the existence of a non-isotropic vector of migration from the outside world. While her formulation does allow for a more appropriate analysis of the data as it exists, the net effect of a non-uniform migration rate from the outside was small compared to the effect of the migration rates between villages. The effect of altering the between village migration rate is likely to be much more marked since the time to stability appears to be markedly affected by changing migration rates.

b) Genetic identity of populations under migration pressure: Nei and Feldman (1972) have recently extended the results of Malécot (1950) to derive measures of genetic identity by descent between a pair of populations under general conditions of migration and mutation, while Latter (1973) has developed a general formulation of the island model in terms of identity by descent. The Yanomama represent an exceptionally favorable population to which these methods of analysis can be applied by reason of the known relationships between a subset of villages, the existence of a set of measures of distance for these villages, and a large amount of genetic variability which may well form a limiting case for the model of Nei and Feldman.

c) Expectation of genetic diversity conditional on the geographical distribution of demes: In any study of the effect of genetic migration, consideration should be given to the underlying geographic distribution of the demes for the migration rates between demes will be some function of this distribution. (In the absence of perturbing factors, such as sociopolitical relationships between human sub-populations, this function will be based solely on the distance relationships between demes.)

i) The effect of settlement expansion on genetic variability. A simple model of population expansion by repeated settlement fissioning will be examined in terms of the resulting genetic variability. Such a model, derived from a deterministic model of Bylund (1960), is particularly appropriate to the Yanomama where village fission has played an important role in recent times. The results will be contrasted with the expectations generated by the 'diffusional' and 'gravitational' models. In addition, a variety of contrasts are possible through the use of the Monte Carlo methods and by comparing the results with those obtained by Morrill (1962).

ii) Assuming that settlement expansion has finished because of some environmental limitation, then the resulting distribution of colonies can be examined in terms of a Lüschan system. The actual distribution of settlements may incline to a simple packed system (Lüschan, 1960) or secondary agglomerations may occur (Isard, 1956). The exact form of the settlement pattern is expected to have a different consequence in terms of the effect of migration and the distribution of alleles among the settlements.

VI.1.2.3. Genetic Relationship Between Individuals Within Villages. - So far the majority of genetic analyses of this project has focussed on the genetic relationships between villages, with little attention paid to the distribution of genotypes within villages. Until recently there was little theoretical development to encourage such an analysis, but Nei (1971) has recently extended some results of Malécot (1967) in a manner that enables the genetic relationships between individuals within a population to be compared to the genetic relationships between populations. In this analysis we shall follow Nei and base our measures of genetic relationship on the concept of genetic identity by descent. The measures of genetic relationship between populations will be thus based on the distance measure introduced by Nei (1972). The measures of relationship between individuals within a village will be derived by considering the phenotype to be a vector of observations on a set of discrete distributions (cf. Friedlaender, 1969). A measure of distance, analogous

to Nei's, can be readily developed and the relationships of individuals within village analyzed with respect to genealogical relationships, when this information is available, and compared to the relationship between villages.

VI.1.2.4. Partition of Genetic Variation Within and Among Yanomama Villages (Spielman, Sing, Neel).

VI.1.2.4.1. "Conventional" Cockerham Analysis. -

For the last half-dozen years, C. C. Cockerham has been developing techniques, in analogy with classical analysis of variance, for the estimation, partitioning, and testing of hypotheses concerning allelic variance components. Direct applications of the method have already been made by Sing et al., using the Tecumseh data (Sect. V.4.3.). More recently, Rothman, Templeton, and Sing have been revising the model on which Cockerham's analysis was based, identifying additional parameters of interest and comparing different estimation procedures (Sect. VI.4.1.). Cockerham's chief conceptual contribution was to focus attention on a measure which characterizes the alleles of each individual. The correlations (or variances), within and among individuals, of this measure of allelic state, are used instead of allele frequencies. Thus the smallest unit of analysis is the individual and his two alleles, replacing the group and its allele frequency. The new model and its analysis continue to use this foundation, but there are preliminary indications that it may lead to substantially different estimates for some parameters. In addition, the new parameters are expected to convey fundamentally more important information about population structure (Sect. VI.4.1). There are available therefore, several partially overlapping strategies for this partitioning analysis of the genetic variation within and between villages, and one outcome of this study will be a framework for comparison of the various models and procedures for estimation.

The "conventional" Cockerham analysis will involve first the determination of the values for f , F , and θ . These are basic population structure parameters which may be given both a (intra-

class) correlation and a variance interpretation. Viewed as the former, f is the correlation between (allelic state measures of) uniting gametes within sub-populations, F is the same kind of correlation ignoring subdivision of the population, and θ is the correlation of the allelic state measures of alleles in different individuals (same sub-population). These parameters and others for the improved model of Rothman et al., will be estimated using methods developed by Cockerham and the latter authors. The extensive inter-village migration described elsewhere (Sect. V.1.5) may make it necessary to choose widely separated villages, thus minimizing the possibility of genetic exchange among the population studied.

Yanomama population structure has been extensively treated by Neel and Ward in a manner with parallel goals, but using functions of genotype and gene frequency. Thus, there is to begin with a useful comparison of new results with theirs. Using the same 37 villages they worked with, or the newly extended data on 48 villages, is f (significantly) negative on the average, as they found for F_{IS} ? Do the largest sub-populations (N_e) have the most negative f 's? (see Cockerham ms. (1973), p. 35). Is there a correlation between f and $1 - H/2\hat{p}\hat{q}$?

A principal aim of this part of the study is to look for heterogeneity of f and θ values for different loci. Under the hypothesis that the loci are not undergoing differential selection, these parameters might be expected to be equal for all loci. Heterogeneity by locus or by village would be evidence rejecting the hypothesis, a finding of considerable importance. Another interesting potential source of differences in structure parameters is sex. Abundant evidence has accumulated that men and women do not contribute equally to the succeeding generations' gene pools. Is this distinction recognizable in the genetic parameters?

VI.1.2.4.2. The Allele Correlations and Lineal Effect. - The availability of the correlations described above makes it possible to determine the relevance of allele correlations for genetic distance and microdifferentiation. Neel has repeatedly stressed the special character of founder effect in the Yanomama. It is likely that

when a group splits to form new villages, the members of the old group are not randomly distributed between the fragments produced. The presumed tendency to preserve lineage integrity or "lineal effect" as Neel has called it, is proposed in part in explanation for what is taken to be striking microdifferentiation of villages in allele frequencies. Following a similar argument, Spielman has shown that men and women from the same village are much more similar in shape (overall morphological proportions) than men and women from different villages. If the shape measure reflects the operation of genetic variables, this demonstration is equivalent to a finding of relative genetic homogeneity within villages, the necessary counterpart to heterogeneity between them. The genetic microdifferentiation and the morphological homogeneity suggest that it should be possible to demonstrate the genetic similarity within villages directly.

Specifically, we wish to determine: when a village splits, are members within each new village more similar genetically to each other than if the villages were formed by random splitting of the parent village? The correlation of allele measures in different individuals seems ready-made for this question. For many villages and for some clusters we know how to group the individuals to re-constitute an ancestral village which no longer exists; the correlations (θ) in such reconstructed villages can then be compared with those in the descendant villages for a direct test of the prediction. In addition, this approach should find a fruitful comparison with the genealogically oriented approach to the question of how much differentiation is introduced at the time of a split (Ward, Neel).

Finally, it should be clear that an extension to a kinship/isolation by distance sort of model is possible. The correlation between allele measures in individuals of different groups can be determined as a function of the distance (even of the genetic distance) between groups. This can be compared with the standard approach of Malécot (or Morton) which has already been pursued with these data (Schull and Neel).

VI.1.3. Further Monte Carlo Simulation Based on Yanomama Population Structure (Li). - The results of our Monte Carlo simulation experiment,

designed to study the fate of a single mutant gene in a primitive population, suggest that the demographic parameters and the mating structure which characterize the Yanomama population lead to consequences that cannot at present be adequately predicted by analytic models based on diffusion equations or other Markov process procedures.

Two simulation experiments have been planned for the coming year. The first is designed to study the distribution in later generations of genes contributed by the founders of a primitive population. We shall approach this problem by assigning 4 unique neutral alleles to each of the two multiallelic loci carried by every member of the initial (founding) population. Thus, it is possible to measure (a) the proportion of genes contributed by each founder, by examining the distribution of these alleles after 400 years of simulation; (b) the rate of decay of genetic variation due to population subdivision and non-random mating; (c) the extent of heterogeneity of the F-values caused by factors such as different fertility or other biological factors; and (d) the degree of hybridization among villages due to the systematic pressure of intervillage migration.

The second experiment is designed to study non-random association of neutral genes. Each member of the simulated population will possess 60 biallelic loci. The program to be developed will provide for independent segregation of 20 of these biallelic loci. The remaining 40 loci will be organized into 20 independent linkage groups, with two loci for each. Recombination frequencies of 5%, 15%, 25% and 35% between loci will be assigned four sets of five pairs of loci each. Thus, this study will not only enable us to measure non-random association of non-alleles in the absence of natural selection but also the effects of intensity of linkage on non-random association of non-alleles in a primitive population such as the Yanomama. In addition, the results from this same experiment may be used to measure the genetic drift to be expected under the mating structure characterizing the Yanomama.

VI.2. ORGANIZATION OF THE PROGRAM IN JAPAN

VI.2.1. A Hopeful Projection. - With respect to the development of the program in Japan, this grant year may be divided into two portions:

1) Between November, 1973 and August, 1974, the Pilot Study will be continued. If and when word of approval of the Expanded Project is received during that period, we would plan to order a GEMSAEC-type Autoanalyzer (see below) and adapt it to the needs of the Project, building up a battery of procedures which in due time will be transplanted to the Japanese Project. We would also recruit Japanese personnel for the Project.

2) Beginning in September, 1974, we would expect to initiate the major project in Japan. Neel expects to be in Japan during the entire first year of this Project.

It is anticipated that one of the two Japanese professionals currently enrolled in the Pilot Project will, on the basis of a commitment at the time of their recruitment, have been brought to Ann Arbor for one year, for advanced training, beginning August, 1973, and the other will come in August of 1974. It is anticipated that in addition to receiving training, while in Ann Arbor they will be working on the detailed characterization of some of the variants already detected in Japan and also participating in the analysis of the many specimens to be collected in connection with the Alpha Helix expedition during the late spring-early summer of 1974.

With respect to adapting the GEMSAEC* to our needs, given the approval of the program in Japan, the plan is to initially apply GEMSAEC techniques to the following four enzymes: isocitrate dehydrogenase, glucose-6-phosphate dehydrogenase, malate dehydrogenase and glucose phosphate isomerase. All are red cell enzymes which can be assayed by following absorbency changes at 340 m μ since all involve NADH-linked reactions. A Spectro-Film filter, with a band width of 10 m μ , is available for this

* GEMSAEC is an acronym derived from the major sources of developmental funding (National Institute of General Medical Sciences and the Atomic Energy Commission) for a fast analyzer.

wave length which is within the operating capabilities of the "NASA version" of the fast analyzer. Since we have already shown the feasibility of G-6-PD measurements using this spectral assay technique in the GEMSAEC it is reasonable to expect that the method can be quickly applied to the other enzymes listed above.

Thus, the potential exists to use such instrumentation to screen a population for protein variation based on molecular properties other than charge alteration. We see in the GEMSAEC the possibility of effectively screening for such additional properties as K_m variation and altered susceptibility to thermal inactivation. The latter seems particularly attractive as an initial approach because of ease of adapting to GEMSAEC instrumentation. Further, based on existing reports in the literature, it is reasonable to expect that primary structure variations will, in a number of instances, be accompanied by alterations in susceptibility to inactivation by heating. For example, Kirkman and co-workers (1964), among others, have shown the occurrence of a number of heat labile G-6-PD types in Western European populations. The other three systems we propose to use in the initial screen have not been characterized in a similar manner.

Although the thermal inactivation approach appears to hold good promise as a screening method we feel, given the capabilities of the fast analyzer, that variation expressed through altered kinetic behavior also provides a potentially fruitful method of screening. The parameter likely to reflect variation is K_m . Again, the G-6-PD system has been looked at for K_m variation and there are a number of documented cases where such variation is seen in human populations. As is the case in thermal inactivation, K_m measurement does not depend on the absolute amount of enzyme present so that concentration variation from sample to sample does not present a difficulty. Further, because of the sensitivity of the current generation of fast analyzers, very little sample material is required for either of these approaches.

VI.2.2. Contingency Planning Should the Program Currently Proposed for Hiroshima and Nagasaki Not Materialize. - As of this writing, there

are a number of administrative considerations on both sides of the Pacific which may either indefinitely delay or even abort the proposed new search for evidences of increased mutation rates in Hiroshima and Nagasaki. In view of the extent to which that program is linked with this program at Michigan, contingency planning seems only prudent. Looking ahead several years, we see three alternatives to proceeding as projected:

1) In the event that it proves impossible to continue ABCC in its present form, or in the event that ABCC for a variety of reasons is unable to accommodate this program, it is proposed that consideration be given to entrusting the program to a consortium of universities, of which the University of Michigan would be the U.S. component and three or four Japanese universities would be the Japanese component. This plan is a poor second to the present projection. However, the scientific significance of these new genetic studies is such that exploration of every legitimate avenue of implementing them seems indicated.

2) In the event that no program can be developed in Japan, we can visualize two "compensatory developments" here at Michigan:

a) As noted in the Introduction to Section VI.1, we believe that the possibility that Indian mutation rates for the cistrons coding for protein structure approach 10^{-4} is a development calling for vigorous follow-up. We are prepared to propose a major program in which blood specimens from the town of Tecumseh, Michigan are subjected to the same type of biochemical studies as those employed for the Indians, while concurrently, for essentially the same proteins being studied in man, experimental studies of mutation rates be undertaken in *Drosophila* and the mouse.

b) Alternatively, given the continuing concern for the potential mutagenic effects of all kinds of environmental contamination (including radiation), we are prepared to propose a major pilot study on the monitoring for the possible effects of mutagenic agents, employing cord blood specimens from newborn infants, along the lines discussed by Neel (1972) and Neel, Tiffany, and Anderson (in press).

Which of these programs should be selected for development should the Japanese program not materialize depends on a number of

scientific and administrative considerations which we are prepared to discuss in detail at the appropriate time. At this juncture we wish only to indicate that considerable thought has been devoted to alternatives should it prove impossible to develop the very important potentialities in Japan for advancing our knowledge of human radiation genetics.

VI.3. SUPPORTING LABORATORY ACTIVITIES

Aside from efforts to automate the search for biochemical variants in man, the activities of the biochemical and serological screening laboratories for this next year are expected to fall under two categories.

VI.3.1. The Development of New Systems. - It is proposed to introduce into the laboratory an additional battery of 6-8 genetic systems in which variants can be detected by electrophoretic techniques. This will broaden the base of our knowledge of the frequency of variants among natural populations and furnish additional data for the types of studies described by others in this report. In selecting additional systems to be added to our present battery, we will be guided by two considerations. The first is the prospect that these systems may be used in the Biochemical Genetics Program in Japan. We will choose systems which give the greatest potential for the detection of electrophoretic variants, and also consider the usefulness of these systems in any attempts to evaluate kinetic or stability parameters using, for example, the GEMSAEC system.

The second consideration will be to attempt to develop maximum overlap between our systems and those being developed by Dr. H. Malling and his associates at the National Institute of Environmental Health. They are currently setting up a battery of nine protein systems in the mouse for a study of the effects of chemical mutagens, and we are exploring the usefulness of four of these systems for our purposes. Dr. Larry Valcovic is attempting to introduce an additional 13 polymorphic or mutant enzyme systems into a background of C57BL/6 mice to provide a large number of genetic markers which can be scored using electrophoresis. We would attempt to introduce as many of these systems as possible into

our screening program. Several of the systems now under evaluation meet these requirements (see Table 10). When a satisfactory level of competence has been achieved, some 5000 stored specimens, available for study, will be investigated.

VI.3.2. The Alpha Helix Expedition. - As noted earlier, there is a high probability that a Brazilian-American team of some 10 persons will be able to avail itself of the facilities of the Alpha Helix for some two months in 1974, on the Amazon and such tributaries as the Rio Branco and Rio Negro. Preparation for the field work will require approximately one month. The general strategy will be to collect as many blood specimens as possible from appropriate groups during that period but to do relatively little with these specimens in the field. Since if all goes well the rate of input into the laboratory will far exceed its capacity to make determinations, for the most part the specimens will simply be placed in storage as they come to the laboratory, with the typings to be performed later. The current projection is for the collection of 200-300 specimens per week, or 1600-2400 for the two-month period. This number of specimens is beyond the capacity of the laboratory for 1973-1974; some continuing work is visualized for 1974-1975.

VI.4. MATHEMATICAL AND STATISTICAL ASPECTS OF POPULATION GENETICS

In the next contract period we plan to expand our work in the theoretical aspects of population genetic models and statistical estimation procedures. In Section V of the Progress section we outlined our general objectives. Firstly, we plan to address ourselves to evaluation of the parametric definitions and assumptions implied by certain models of human populations, both genetic and demographic. Secondly, we plan to explore the properties of estimators available for parameters of these models. A major effort will be made to extend the likelihood method in this problem area. Thirdly, we have bodies of data which are particularly appropriate for the investigation of these theoretical problems, and which will very likely generate new theoretical problems and models. The

Table 10. A list of electrophoretic systems under consideration for addition to the present battery.

Enzyme System	Polymorphic Human Populations	Nonpolymorphic	Reference
Glyceraldehyde-3-Phosphate Dehydrogenase		+	Charlesworth , D., Ann. Hum. Genet. Lond. <u>35</u> , 477 (1972)
Isorbital Dehydrogenase		+	Charlesworth , D., Ann. Hum. Genet. Lond. <u>35</u> , 477 (1972)
Fructoaldolase		+	Charlesworth , D., Ann. Hum. Genet. Lond. <u>35</u> , 477 (1972)
Glyoxalase		+	Charlesworth , D., Ann. Hum. Genet. Lond. <u>35</u> , 477 (1972)
Glucose Phosphate Isomerase		+	Charlesworth , D., Ann. Hum. Genet. Lond. <u>35</u> , 477 (1972)
Glutamic-Pyruvic Transaminase	+		Chen, S. et al., Ann. Hum. Genet. Lond. <u>35</u> , 401 (1972)
MDH-Diaphorase		+	Hopkinson, D. A. et al., Ann. Hum. Genet. Lond. <u>34</u> , 1 (1970)
Glutathione Reductase	?		Kaplan, J., Nature <u>217</u> , 256 (1968)
1,3-Diphosphoglycerate Mutase		+	Chen, S. et al., Biochemical Genet. <u>5</u> , 481 (1971)
Phosphoglycerate Kinase	+		Chen, S. et al., Amer. J. Hum. Genet. <u>23</u> , 87 (1971)

These systems are to be evaluated to determine whether they can be run on the S. American Indian samples stored in liquid nitrogen. Maximum effort will be made to develop conditions that will allow multiple determinations from a single gel.

availability of computing facilities makes it possible to conduct extensive statistical analyses or to generate, by simulation techniques, specific bodies of data drawn from distributions with known parametric properties.

VI.4.1. Application of the Likelihood Principle to Problems of Estimation in Population Genetics (Rothman and Sing). - We recognize a basic problem in deriving useful inferences from genetic data collected from human populations concerning the forces which operated in the past on the gene pool or about the behavior of the gene pool in the future. Two general issues are involved. The first is the definition of a model which has realistic biological meaning and can provide utility in answering either or both aspects of the problem. The second is the choice of statistical procedures used to estimate the parameters of the model. Models may be simply functions of allele frequencies and selection coefficients or they may be complex relationships between selection coefficients and parameters which describe the demographic distribution of alleles among individuals in the observed population. It is our goal to consider the properties of more complex models which purport to separate three major categories of influence on the observed frequencies of genes (genotypes) in samples collected from human populations. Those three influences are the effects of subdivision, the effects of time (drift) since subdivision, and selection (biology of the genotype).

The choice of estimation procedures for parameters which define these three influences is at first of greatest interest to us. We have a number of realistic models. Very little theoretical work has been done to indicate the most reasonable way to estimate parameters in these models. While we cannot expect every problem in estimation to be solved by the same procedure, certain strategies have advantages in specific situations. For certain problems involving a multiplicity of parameters we intend to explore the properties of likelihood estimation procedures. The convenience of accounting for prior information, the fact that all the information in the data resides in the likelihood, the efficiency property of maximum likelihood estimates and the possibility of deriving

locally most powerful tests of hypotheses regarding selection are reasons for considering further development of the likelihood concept.

Our efforts will involve 1) development of likelihood expressions for real data sets, 2) derivation of the efficient variance-covariance matrix for estimates, and 3) applications to data sets with known parametric properties. We will be especially interested in resolving procedures to partition the effects of subdivision, drift, and selection and in derivation of procedures for testing hypotheses about these three factors.

VI.4.2. Genetic Statistics for Mutation Problems (Jensen). - Simulation studies of the average number of generations until a rare mutant becomes lost in a finite diploid population have shown that the estimates of the mean time to loss are much smaller than that predicted by theory (Sect. V.1.6.) This theory has been derived by Kimura and colleagues from the study of haploid populations with non-overlapping generations, through the use of diffusion theory approximations. These results may have little applicability to human populations. Jensen proposes to examine the mean time to fixation or loss of a mutant in a diploid population with overlapping generations by examining the underlying stochastic process, i.e., to extend Kimura's treatment so that it is appropriate to human populations. With the appropriate model, it is then possible to obtain an exact statement of the mean time to loss, an estimator of the mean time, and its variance.

Directly related to estimating the fate of a new mutant are problems in estimation and hypothesis testing dealing with rare events, the mutation rate. The available procedures are based on the assumptions of independent observations for haploid populations. For diploid structured populations these procedures are not valid since individuals are not independent, and in many cases we have the additional problem of detecting the existence of a mutant. The mutant cannot be observed quantitatively; its presence is marked by a co-dominant allele. Jensen also proposes to study and develop valid statistical estimation and hypothesis testing procedures for problems of this nature.

VI.4.3. Application of Yule's Law to Fixation of Alleles in Populations with Complex Demographic Structure (Rothman, Templeton). - Yule's Law is a general probability model that described, quite accurately, word frequencies (Simon, 1954), distributions of responses to psychological tests (Haight, 1966), and originally associated with an evolutionary process by Yule (1923). A computer simulation of a complex human population by Li and Neel (Sec. V.1.6.) was developed to study the loss of neutral alleles. After examining the output it was found that the proportion of alleles lost at a given generation fit a simple form of the Yule Law remarkably well. We therefore investigated the set of assumptions that imply such a Law. In particular, we found that if the underlying process is Markov and if the distribution of the neutral allele is geometric then for infinitely large populations Yule's Law is obtained. The geometric distribution arises as a result of maximum entropy subject to the neutrality assumption. This then suggests that the effect of including more complex demographic conditions tends to increase the entropy. We propose to investigate further this result, first by taking into account the finiteness of the population and secondly by examining how each demographic factor can contribute to entropy. The tools to be applied to this problem include direct product branching processes.

VI.4.4. Density-Dependent Selection in Age Structured Populations (Smouse and Weiss). - We have previously described our effort to develop a density-dependent model of growth regulation for an age-structured population (see Sect. V.1.8.). It is our intent to extend this treatment to the consideration of the joint behavior of populations of genotypes. Initial progress has been made in this area by Charlesworth and Giesel (1972), who have investigated the situation where diploid, single-locus genotypes have growth rates damped by total population size. They demonstrate that heterosis (of a density dependent sort) is necessary and sufficient for coexistence of two alleles at a single locus.

We plan to generalize the treatment of density-dependent selection in age-structured populations so as to include various types of interactions among the genotypes. We expect to be able to demonstrate

that genotypes may coexist whenever they are ecologically complementary, with or without heterosis.

VI.4.5. Consistent Two-Sex Models for Human Population Demography (Weiss, Smouse). - Up to now very little has been accomplished in the way of internally consistent two-sex models of human demography which would predict both the proper age-specific sex ratios and the rates of maternity and paternity by age, depending on the social mating patterns in force. Only the most oversimplified models have been developed. The difficulty has been that observed birth and death rates for the sexes produce different 'intrinsic' growth rates, which means, in the Malthusian extreme, that one sex displaces the other. This is inconsistent and must be avoided by anyone desiring realistic models.

We will take the basic density-dependent matrix method already derived by us and extend it to cover two sexes; by introducing a means for relating mating behavior between members of each sex we will derive a stable, zero-growth population in which the proportion of the sexes at each age is determined by the density feedback system, the basic vital rates, and the social structure. The method will produce consistent models even though in general the 'intrinsic' growth rates of each sex will differ. It will be shown that secondary and tertiary sex ratios do not have to correspond to inevitable parity at particular ages, but that different ratios result from different conditions. The model will improve the basic analysis of two-sex populations from both demographic and genetic viewpoints, and relieve the problems which have plagued many past studies on real populations.

VI.4.6. The Analysis of the Role Genes Play in Determining Normal Variation in Predictors of Risk for Coronary Heart Disease in the Tecumseh, Michigan Population (Sing). - The goal of our work in this area of research is to utilize the unique collection of data available from the Tecumseh Community Health Study to obtain estimates of the role of genetic variability in determining variation among individuals for variables which are defined as predictors of coronary heart disease (CHD). During the course

of this study we anticipate the development of a method of applying quantitative statistical genetic concepts to the study of an unselected human population. Both mathematical and statistical problems unique to human data are anticipated.

Our experience with serum cholesterol (Sect. V.4.4 of the Progress section) will serve as guidance for a similar analysis of three other "risk" factors. They are blood pressure, blood glucose, and serum uric acid. It is our intent, for each of these variables, to subdivide the total phenotypic variation among individuals into:

- 1) a component due to age and temporal variation,
- 2) a component due to allelic variation at a single locus whose genotypes have major effects which lead to the skewness or non-unimodality of the distribution of normal values in the population,
- 3) a component due to allelic variation for each of the thirteen blood and serum genetic markers taken separately or in combination,
- 4) a component due to the cumulative effects of polygenes each of which has a small effect on the variation of the risk trait, and
- 5) a component due to residual nonbiological variation which we will further subdivide into fractions identified by socio-economic indicators such as occupation, household affiliation, income, education, etc.

After completing a description of the variable with respect to age, round of examinations and possible bimodality due to a major locus (steps 1 and 2 above), adjusted values will be regressed on each of the available 13 bloods and serum marker loci (step 3). Standard multiple regression techniques will be employed. Step 4 will employ an analysis of the correlation between relatives to determine whether the degree to which polygenes not identified by the 13 genetic markers may be contributing to variation in the risk variable. Several major analysis features are possible in this 4th step of subdivision of the sources of variation that are uniquely possible in this body of data. They are:

- 1) the use of concomitant socio-economic variables to define the common environment of pairs of relatives,

2) the use of the blood and serum marker loci as "treatment" variables to identify sections of chromosomes which may carry loci affecting the trait, and

3) the use of kindred definitions already established by Sing et al. (1973) (and Sect. V.4.3. of Progress section) to evaluate correlations among pairs of relatives with varying degrees of sampling independence. These three facets will aid in a more realistic evaluation of the role of polygenes. The first facet is made possible by the availability of a large number of nonbiological variables for each of the individuals in our sample. Variation among family or household units contributes (is equated to) covariation between individuals within such a unit. This effect in human populations is a bothersome bias in the evaluation of phenotype correlations. We believe we have adequate data to properly subdivide the effect of variables which contribute to a correlation due to "common environment." Step 5 will involve identifying the contribution of such variables to the total phenotypic variation.

The second facet will involve selection of those polymorphic loci accounting for the greatest variability in the risk variable to establish treatment genotypes. The correlation between relatives for the entire data set will be compared with the same correlation based on individuals having a common single (or multiple) marker genotype. In this way, an estimate of the proportion of variance due to polygenes identified by the marker loci can be partitioned from the effects of variation in the remainder of the genotype.

The third facet involved in our evaluation of correlations we believe will identify a bias which is often overlooked in human data analysis. When a large sample, which is approximately the size of the interbreeding unit, is collected, one is faced with pairs of relatives which are not statistically independent. Our investigations of genetic "structure" of the Tecumseh population (see Sect. V.4.3. of Progress) will enable us to define our sample of relatives in such a way that the bias due to the nonindependence of pairs of relatives drawn from the same kindred will be estimable. We will begin step 4 by partitioning the variation among and within the kindred units defined in Table 11. The

Table 11. Number of Kindreds in Terenure by size for each genetic system.

Kindred Size	GENETIC SYSTEM ¹											
	M.Ss			gab.onloaia			sGo			Ga		
	All	Couples 1	2	All	Couples 1	2	All	Couples 1	2	All	Couples 1	2
2	151	10	10	214	73	73	211	72	72	234	76	76
3	200	136	136	216	154	154	215	155	155	206	137	137
4	247	195	195	236	189	192	241	192	195	232	180	183
5	170	149	154	147	128	139	138	120	130	138	119	128
6	119	106	116	94	85	100	103	94	111	97	84	100
7	59	54	72	43	39	57	42	39	56	37	34	50
8	42	40	65	35	34	56	36	34	57	31	29	47
9	20	20	35	24	23	43	25	24	45	26	25	47
10	28	27	50	24	24	57	23	23	54	21	21	48
11	18	17	36	16	16	38	15	15	35	14	14	33
12	22	22	57	13	13	29	13	13	31	16	16	36
13	10	10	24	7	7	21	9	9	27	8	8	25
14	7	7	20	11	11	34	10	10	30	9	9	31
15	9	9	26	7	7	29	8	8	33	4	4	14
16	4	4	15	6	6	24	4	4	18	7	7	28
17	3	3	10	6	6	21	7	7	25	7	7	23
Total (2-17)	1109	809	1021	1099	815	1067	1100	819	1074	1085	772	1008
18-121	49	49	327	33	33	261	34	34	266	31	31	242
Remarriages			14			4			4			4

¹ The number of kindreds of size s with one or more couples.

² The number of couples in kindreds of size s. Remarriages occurred only in kindreds larger than 17.

variable in one sex in the Tec each DNA set.

Step in analysis	Source of variability
1	Age in Round I Age in Round II Round differences Replicate differences.
2	Bimodality in the initial distribution
3	Polymorphic marker genetic loci: <ul style="list-style-type: none">- among genotypes each polymorphic locus- combinations of genotypes determined by polymorphic genetic loci
4	Polygenes: <ul style="list-style-type: none">- Among kindreds- Among individuals in kindreds- Among nuclear families in kindreds- Among individuals in nuclear families in kindreds<ul style="list-style-type: none">- Among spouse pairs²- Among sib pairs- Among 1st cousin pairs- Among parent offspring pairs

¹The analysis will proceed from 1 through 4. Each successive analysis will be of the variable of interest adjusted for the effects estimated in the previous steps. In this way a stepwise orthogonal subdivision of variation will be obtained.

²Pairs of relatives will be grouped according to kindred and household classification so as to 1) remove bias attributable to non-independent genetic influences and 2) estimate the effect of common household environment on the computed correlations among pairs of individuals.

correlations between relatives which make up these kindreds will be biased by the common environmental effects of certain combinations of relatives constituting nuclear families and the nonindependence of these nuclear families within kindreds. The former bias was considered above by step 1 of the analysis. Proper consideration of the latter will be possible by using the information about the coefficient of kinship within kindreds available from our previous work on population structure.

A summary of the sequence of analyses which are meant to subdivide the normal variation in a risk trait is given in Table 12.

VI.4.7. Applications of Statistical Procedures to Estimate Selection Effects in Parthenogenetic Populations to Human Populations (Rothman, Templeton). - In the Progress Report, Sections V.4.2. and V.5.1., we discussed the theoretical and applied work being carried out by our group using the parthenogenetic biological design. Parthenogenesis is a system of reproduction which we realize does not fully represent all factors which are operating in the human population. However, it was obvious to us several years back that new approaches were needed to gain insight into the nature of complex human populations. Models derived for and experiments conducted with parthenogenetic *Drosophila* have aided in development of new approaches. In the next grant period we intend to extend these approaches to sexual populations. To date, only one-locus models have been applied to sexual populations. These developments represent the first estimators and a test scheme for selection in sexual populations that take into account the Markovian dependence that exists in data drawn from a single population over time. The problem is much more complex. The single locus is not the unit of selection. A useful theory for sexual populations must incorporate multiple loci. Much work is yet to be done, even with one locus models, to extend our work with parthenogenesis to sexual populations in general. We plan to continue to explore how we might make further applications of the developments of our work with the parthenogenesis to problems in sexual populations.

VI.4.8. Proposed Updating of the Computing Facility (Griffith, Sing). -

In 1967 the AEC purchased a basic IBM 1130 system consisting of 8K core, a printer, and a card read-punch device to meet the needs of this project. This initial system has been modified over the years to meet the growth in research supported by this project. A faster printer was added in 1968. Then a faster card read-punch unit was installed. This was followed by an additional 8K of core and an IBM 2310 peripheral disc storage unit. Last year we added a high speed Calcomp DS 13 disc drive. We have gone through a period of data collection and organization which will put new demands on our computing machinery. It is obvious that at the present rate of growth in numbers of analyses on existing data, the computing demands which we anticipate over the next three to five years will increase and hence will far exceed our present hardware capacity. Furthermore, such a projection ignores the new data which will be collected during the same period which will surely tax the system for collating and organization for the more complex analysis to be done.

Even at the present level of involvement of the investigators, we have extended our demand by 12 percent beyond our ability to deliver computing time (see Progress section V.4.7. for a breakdown of computing during 1972).

Consequently, we believe it is proper to evaluate the totality of our computing system at this time in terms of how best to meet the computing load for the next three to five years. We have conducted an evaluation of our requirements and an intensive survey of the computing hardware available to meet those requirements. In Appendix B we present the results of our research on this matter. The cost of computing hardware has decreased dramatically in the past six years. For about the same dollars we have been spending each year to modify the IBM 1130 it is now possible to make major alterations in this computing system with a resultant capability which will keep well ahead of the growth in our computing requirements in the next five-year period.

VI.5. SPECIAL STUDIES

VI.5.1. The Genetic Unit of Selection (Sing, Templeton). - The objectives, experimental design, description of biological material, and rationale for our work is presented as a preamble to our Progress Report, Section V.5.1. Preliminary experiments have shown that the proposed design for studying gene interaction is one which can detect and accurately estimate selection at a multi-locus level. We have shown in the past year (reviewed in Progress Report, Section V.5.1.) that very large selective forces are indeed operating in the biological system represented by parthenogenetic *Drosophila*. Detailed analyses of the selective forces operating and the nature of the genetic integration reflected by such forces are yet to be made. We propose the following experiments to make such a detailed analysis.

1. Extend the preliminary experiment to get larger numbers:

In particular, we need to have considerably more sexual F_2 flies to accurately estimate recombination frequencies between linked loci. This will allow us to make a quantitative study of the multi-locus systems. Many questions can then be answered. For example: what happens when the chromosomes are internally balanced but an imbalance exists between them due to segregation; does a double crossover counteract the fitness decrease due to a single crossover in a three locus linked system; at what genetic unit (if any) can we adequately combine the fitnesses in an additive or multiplicative fashion to predict the fitness of some higher order genetic unit?, etc. This last question is essentially the question, "What is the unit of selection?" The preliminary experiment has already shown the gene is not the unit of selection since two locus fitnesses cannot be predicted by a model based on additive or multiplicative combination of single locus fitnesses. However, perhaps at some higher genetic unit an additive or multiplicative model can be used.

It is also desirable to obtain more parthenogenetic F_2 . This would increase the sensitivity of the design for all systems and allow us to analyze the six locus genotypes for which we simply do not have sufficient numbers at present.

2. Experiment to study maternal effects: In the preliminary experiment, the O genome was derived from females, and the S genome was derived from males. It is possible that a maternal effect might influence the genotype frequencies of progeny of these S-O F₁ hybrid females. To control for this possibility, the preliminary experiment should be extended by using the following S-O F₁ hybrid females:

$$\begin{array}{ccc} \text{Br}_{14}\text{-Iso-8-Oct} & \times & \text{S-sl-v-pm-vl-Im}\varnothing\varnothing \\ & \downarrow & \\ & \text{S-O F}_1 \text{ hybrid}\varnothing\varnothing & \end{array}$$

The data would then be analyzed in a fashion identical to that of experiment 1. Such an experiment would be expected to detect and define any maternal effects on fitness if, in fact, any were present.

3. The effect of whole chromosomal interactions: In both of the above experiments the F₁ hybrid females are heterozygous at all five chromosomes, and consequently the entire genome is being perturbed by meiosis. One way of dissecting the fitness interactions among different chromosomes or groups of chromosomes is to create females heterozygous for only certain of the chromosomes while the other chromosomes are fixed to one of the parental types. Consequently, only a portion of the genome would be perturbed during meiosis. This gives us another method of obtaining a more precise answer to the question: What is the unit of selection? Such females are easily made by taking advantage of the fact that there is no crossing over in the males. For example, one potential breeding scheme is (ignoring the dot chromosome):

$$\begin{array}{ccc} \text{Br}_{12}\text{-S-sl-v-pm-vl}\varnothing\varnothing & \times & \text{Iso-8-O-3-Im}\varnothing\varnothing \\ \downarrow & & \downarrow \\ \text{F}_1\text{S-O} \times \varnothing\varnothing & \times & \text{S-sl-v-pm-vl-Im}\varnothing\varnothing \\ \downarrow & & \downarrow \\ \varnothing\varnothing \quad \varnothing\varnothing \quad \varnothing\varnothing \quad \varnothing\varnothing & & \varnothing\varnothing \quad \varnothing\varnothing \quad \varnothing\varnothing \quad \varnothing\varnothing \end{array}$$

(next page)

<u>F₂ Phenotype</u>	<u>Chromosomal constitution</u>	<u>Result of meiosis</u>
wild ♀♀	00 00 00 00	All chromosomes perturbed
vl ♀♀	00 00 00 00	"vl" chromosomes unperturbed
pm ♀♀	00 00 00 00	"pm" chromosomes unperturbed
v ♀♀	00 00 00 00	"v" chromosomes unperturbed
pm vl ♀♀	00 00 00 00	"pm" & "vl" chromosomes unperturbed
v vl ♀♀	00 00 00 00	"v" & "vl" chromosomes unperturbed
v pm ♀♀	00 00 00 00	"v" & "pm" chromosomes unperturbed
v pm vl ♀♀	00 00 00 00	Only X chromosome perturbed
sl ♂♂	00 00 00 00	-
etc.		-

These males are then used to create additional chromosomal backgrounds, for example:

sl pm vl♂ x S-sl-v-pm-vl-lm♀♀
 00 00 00 00 / 00 00 00 00
 sl pm vl♀♀ = 00 00 00 00 Only "v" chromosome perturbed in meiosis

Other similar breeding schemes can be performed such that a female can be bred which has any particular chromosome or group of chromosomes homozygous for either the S-type or the O-type parental chromosomes. These females with their various genetic backgrounds can then be studied with the same design used on the F₁ S-O hybrid females, but here only experimentally determined portions of the genome are being perturbed. Because all conceivable chromosome backgrounds fixed for either O or S chromosomes can be constructed, we have an extremely powerful tool to quantitatively analyze interactions among groups of chromosomes and their fitness effects on the chromosome segments marked by the loci studied.

4. The internal balance of a chromosome: The above experiments will characterize the fitness interactions that occur among intact

chromosomes. The next step will be to examine the fitness interactions among loci in the same chromosome. One chromosome in particular will be looked at in detail; the autosome bearing the marker vermillion (*v*). This is the largest autosome of *D. mercatorum* and presently has the most genetic markers available. Our preliminary experiment has already shown strong selective forces are operating on this chromosome as a unit. How much of this selective force is due to interactions with other groups of chromosomes will be answered in the previously discussed set of experiments. The experiment that will be described now will look at the internal (within chromosome among loci) selective balance of this chromosome.

First, as previously discussed, the following flies can be bred:

sl pm vl♀♀ *00 00 00 00* and + ♀♀ *11 11 11 11* (1)

Any selection detected in the parthenogenetic offspring of these females is due to perturbation of the *v*-chromosome only. In addition to these females, the corresponding males may also be bred, for example:

sl pm vl♂♂ *0° 00 00 00*

This male can then be used in the following breeding scheme:

As can be seen, recombination in the *F*₁ females breaks down the original coadapted *v*-chromosome so that the *F*₂ females have even less of a perturbation than the females described in (1). One can then select those *F*₂ females heterozygous for the marker loci and then allow them to reproduce parthenogenetically as before. Selective differences in these partheno-

genetically produced flies would no longer be due to perturbation in the entire v-chromosome, but only in those chromosomal segments surrounding the markers. This backcrossing scheme can be repeated for several generations, and the selective forces detected regressed against the number of generations backcrossed. As the number of backcross generations increases, the average length of chromosomal segments being perturbed in the experiment decreases. Therefore, the regression of genotype frequencies (or measures of epistasis) against generations will give direct information about the relative size of chromosome segments with interacting gene loci.

5. A multi-generation experiment. All the previous experiments are designed to study selection within a single generation. The only type of selection detectable in such a design is viability from the time of diploid restoration until the time of sampling which, in our experimental design, is within 24 hours of emergence from the pupal case. Consequently, fecundity components, and possibly some viability components of fitness in the F_2 females of the basic experimental design are not detectable. To get at these additional components, a multi-generation experiment will be done so that the parthenogenetic descendants of the S-O F_1 hybrid females will be exposed to selection at all life stages including parthenogenetic reproduction. A group of S-O F_1 ♀♀ will be used to start a parthenogenetic population which will be carried for several discrete generations. This population will be sampled for the genetic markers at each generation. Under such conditions, the multi-locus, multi-generation fitness estimators and hypothesis-testing framework developed by Templeton (1972) can be used to study selection.

In summary, the above experiments will allow us to make a thorough quantitative study of the selective forces operating in parthenogenetic Drosophila at the genomic level, inter-chromosomal level, and intra-chromosomal level. Furthermore, by studying individual gene loci we have a unique opportunity to study the behavior of allelic variation for single locus and multi-locus systems in all of these genetic backgrounds. In addition to measures of selection, these experiments allow a test of how the selective forces may be determined by interactions among loci and between these loci and various genetic backgrounds. It is a series of

experiments utilizes one of the few biological systems in which quantitative studies can be done to describe the genetic unit of selection.

VI.5.2. Genetic, Chemical and Serological Studies of Human Immunoglobulins (Gershowitz).

VI.5.2.1. The Search for Subclass and Allotypic Variation Among Non-IgG Immunoglobulins. - The identification of allotypic variation in the IgG molecule was greatly facilitated by the fact that anti-Rh antibody could be used as the source of allotypic antigens. Using anti-Rh coated red cells, anti-Gm antibodies (detecting allotypic variation of the IgG₁ and IgG₂ subclasses) have been found in normal human sera. IgA and IgM myeloma proteins have been coated by the chromic chloride method on red cells and these coated cells used in widespread searches for anti-allotype antibodies among normal human sera, with very little success.

Because of our interest in developing new genetic systems for the purpose of studying human populations we have proposed to resort to heteroimmunization of rabbits, in an attempt to find allotypic variation of the IgA and IgM molecules. As an antigen, we have proposed using IgA and IgM molecules selected on the basis of a single combining reactivity, on the supposition that there is a greater probability that such an antibody population may be derived from a single clone of cells and thus contain a single allotypic site. In addition, preparations of antibody which might be composed of a single allotype specificity would certainly be useful in the proper analysis of immune sera.

From among a number of blood bank donor sera screened, we have selected ten with high titers of anti-rabbit red cell agglutinating activity. Simple salt fractionation and column chromatography on Sephadex G-200 has not yielded sufficiently pure IgM preparations. It is obvious that these preparations are contaminated with IgG, which also has anti-rabbit-red-cell activity. We propose to clear these preparations of their IgG activity by absorption on glutaraldehyde-gelled rabbit anti-human-IgG. Since all human sera possess several heterospecific agglutinins, the IgM preparations will then be used 1) to coat rabbit red

cells which will be used as immunogen in some rabbits, 2) to coat chicken red cells which will be used as immunogen in other rabbits, and 3) for those sera which have anti-human A, to coat type A red cells for use as an immunogen in still other rabbits. As we have stated above, the supposition is that each combining activity may be derived from a different clone or clones of cells and that this should increase the probability that each antibody preparation differs from the other by at least a single allotypic specificity.

VI.5.2.2. A New Allotypic Specificity of the Human Kappa Light Chain. - The Inv allotypic system of the human kappa light chain is composed of three antigens (Inv(1), Inv(2) and Inv(3) and three alleles, Inv¹, Inv^{1,2} and Inv³. (Note that Inv(2) is found only in the presence of Inv(1) and that most Inv(1) positive light chains are also Inv(2) positive.)

We have recently discovered an agglutinating serum (agglutinator) anti-Rh coat combination which has served to identify a new Inv antigen. Table 13 summarizes the reactions observed between the anti-Inv agglutinators and the anti-Rh coats we have studied. It should be borne in mind, however, that an unknown serum is typed for the specificity or specificities indicated by its ability to inhibit the reaction between a specified combination of anti-Rh coat and agglutinator.

Before the finding of the anti-Rh coat Kuh, both Lesp and Sim had been thought to be identical anti-Inv(1) agglutinators. Kuh behaves, with Lesp, as an ordinary Inv(1) coat in a very large number of tests run in parallel with Abr and Lesp but obviously differs from Abr in reactivity toward Sim. The simplest explanation is that Lesp is anti-Inv(1) and that Sim is anti-Inv(x). Inv(x) may be associated with Inv(1) or may be another form of Inv(1).

Several hundreds of Oriental, Negro and Caucasian sera have been tested with the Lesp-Kuh (anti-Inv(1), Inv(1)) and Sim-Abr (anti-Inv(x), Inv(1,x)) combinations and no differential inhibition has been found. It may be that the interpretive scheme is incorrect for these tests ought to distinguish between the Inv(1) sera which are (x+) or (x-), depending, of course, on their relative frequencies. The Inv (-1,x) serum could have been detected by the test combinations if it existed with any appreciable frequency.

Table 13. The Pattern of Reactions Observed Between Inv Coats and Agglutinators

<u>Agglutinator</u>	<u>Antibody Specificity</u>	<u>Anti-Rh Coats</u>				<u>Kuh</u> (1,2)	<u>Lewis</u> (3)
		<u>Abr</u>	<u>Dul</u>	<u>Peters</u>	<u>2290</u>		
		<u>(1,x,2,3)</u>	<u>(1,x,2,3)</u>	<u>Proposed Specificities</u> <u>(1,x)</u>	<u>(1,x,2,3)</u>		
Lesp	1	+	+	+	+	+	-
Sim	x	+	+	+	+	-	-
Vir	2	+	+	-	+	+	-
Metrix	3	+	+	N.D.	+	-	+

The frequency of Inv(x) and its true relation to Inv(1) can only be determined by a specific Inv(x) test. We intend then, to search for an anti-Rh coat which will react with Sim but fail to react with Lesp. The Yanomama will be very useful in understanding the relation of this new specificity to Inv(1), for they are characterized by an unusually high frequency of Inv(1). Conversely, the new specificity may reveal an additional allotypic polymorphism among the Yanomama. In addition to its possible utility as a new population marker, the system may be useful in studying the relation between antigenic specificity and primary sequence.

VI.5.2.3. Further Investigation Into the Myeloma-Like Nature of an Anti-Rh Antibody. - Individual antigenic specificity (idiotypy) has been found in globulin preparations consisting either of entirely homogeneous myeloma proteins or in preparations of restricted antibody. Although it is possible to categorize certain immunoglobulin preparations as derived from either a single or very few initial cells based on the combined findings of a single heavy chain marker, a single light chain, and restricted migration pattern on electrophoresis, only amino acid sequencing can distinguish the two. (Primary sequencing reveals the existence of only one amino acid at each residue position in the case of the monoclonally derived protein but two or several amino acids may be found at certain residue positions in the case of restricted homogeneous antibody of polyclonal origin.)

Homogeneous antibodies have been produced in the rabbit with carbohydrate antigens of bacterial origin, with rigorous immunization schedules. We do not know of any occurrence of a homogeneous, if not actually monoclonal antibody in humans but suggest an anti-Rh antibody herein described as a likely candidate.

We have described, in the Progress Report, the following:

- 1) Anti-Rh Spr when coated on Rh positive red cells, caused the production of an anti-idiotypic antibody in all of three rabbits

immunized. This fact alone suggests a high concentration of similar, if not identical, antibody molecules in the anti-Rh serum.

- 2) The idiotypic activity is not found in the donor serum after absorption of the anti-Rh.
- 3) The idiotype site and the anti-Rh site on the molecule are not identical.
- 4) The anti-Rh antibody has only one serological activity.
- 5) The anti-Rh antibody has one heavy chain allotypic marker.
- 6) The anti-Rh antibody has only one light chain (kappa).

Additional verification of the restricted nature of this antibody, short of amino acid sequence, can be obtained in two ways. One, the demonstration that the antibody preparation has restricted electrophoretic migration will be strong evidence for a restricted, but not necessarily monoclonal, origin. Two, if the antibody has been produced by a single clone, all the molecules should possess the idiotypic specificity; the finding of non-idiotype-possessing molecules is proof of at least biclonal origin.

We have already begun preliminary attempts to investigate the optimal method of elution of anti-Rh from red cells. We intend, when we are satisfied with the method, to obtain an eluate from Rh positive cells coated with anti-Rh Spr which will be subjected to polyacrylamide electrophoresis. If sufficient eluate can be obtained, the antibody will be further examined on polyacrylamide, after reduction and alkylation, to determine whether one, or more than one heavy chain is involved.

In addition, it is crucial to determine whether all the anti-Rh molecules in the serum possess the idiotypic specificity. The proper performance of this experiment depends on the use of a solid anti-idiotype preparation to effect complete removal of all anti-Rh antibodies which have the idiotypic antigen. The relatively easy approach

to a solution of this problem failed to yield an answer because gellation of the anti-idiotypic antibody caused it to lose specificity. We intend to proceed to couple the anti-idiotypic to Sepharose for use in a column. Passage of a limited amount of the anti-Rh serum through the column should reveal whether non-idiotypic anti-Rh can be eluted.

The belief that this anti-Rh antibody may be restricted or perhaps even monoclonal in origin has led us to re-examine the generally accepted notion that all antibodies are heterogeneous. If this antibody is unique, then it may be only a curiosity, but a curiosity, nonetheless, which will permit an estimate to be made of the amount of protein produced (perhaps) by a single clone. It will also be part of the proof that myeloma proteins are normal antibodies, for the belief that a monoclonal protein is an antibody would have support if it could be shown that monoclonal antibody can be produced. If this antibody is not unique, then other examples of monoclonal antibody might be found (perhaps among other anti-Rh antibodies) and a test system devised to relate normal antibody production to myeloma production. To that end, we propose to search for other anti-Rh antibodies which might be looked upon as restricted (single antibody specificity, single heavy chain allotype and single light chain). If found, they will be investigated for their abilities to elicit anti-idiotypic in the rabbit.

(Parenthetically, we might point out again that the ability to produce restricted or homogeneous antibody in the rabbit has been associated with carbohydrate antigen. If it is proven that the anti-Rh Spr is homogeneous, then one might conclude by analogy that the immunizing antigen might well have been carbohydrate, a conclusion of some significance for it has not yet been possible to identify the nature of Rh antigen.)

VI.5.2.4. The Nature of Gm Genes (Kohler, Gershowitz). - It has been generally assumed that the Gm structural genes of IgG heavy chains represent allotypic differences between individuals. Immunofluorescent staining and Gm-typing of culture medium of PHA-stimulated lymphocytes, has, however, shown that such "dedifferentiated" cultured cells are capable

of synthesizing Gm-antigens not found in the serum of the cell donor. Thus, it appears as if the Gm structural genes may not be allotypic after all, but may be isotypic, i.e., present in all individuals. The expression of these genes in an individual depend on the inheritance of regulatory genes controlling the synthesis of the various γ G heavy chains.

Preliminary studies using spent culture medium from long-term lymphocyte lines for Gm-typing provide further evidence for this new appraisal of the Gm system. Gm-specificities not present in the donor's serum have been synthesized by their cultured cells. It appears that dedifferentiation of peripheral blood lymphocytes has occurred, possibly through Epstein-Barr virus transformation, and regulatory inhibition of Gm-specificities no longer exists.

The culture lines are now being grown in large Blake bottles to eliminate the necessity of pooling culture media from several subculture dates. This will allow more precise definition of the culture conditions and determination of IgG production during specific time intervals.

Efforts will be made to locate lymphocyte donors whose serum has not yet been Gm-typed, to permit the comparison of serum to cell-line Gm-types in a larger sample. If insufficient numbers of Gm homozygotes are included in lines already established, long-term lymphocyte lines will have to be established from homozygous donors. (Anomalous antigen can only be detected in cultures derived from homozygotes.)

It is obvious that a proper understanding of the nature of this anomaly necessitates the production of cloned lines. It will then be possible to determine if a particular clone is capable of switching from synthesis of one Gm-specificity to another. Assuming that the production of IgG molecules of different Gm-types is under the control of regulatory genes, this system will make it possible to determine whether these genes can be influenced by external factors, such as cell density, pH, temperature, or by the addition of agents which either stimulate or depress synthesis of IgG molecules. Attempts to produce cloned lines are already underway.

It should also be possible to Gm-type single cells through the use of immunofluorescent anti-Gm antisera, in order to determine whether allelic and genic exclusion hold true for lymphocytes in long-term culture.

During the establishment of a long-term lymphocyte culture, the cells become transformed by virus and appear to dedifferentiate to a more primitive, blast-like state. These cells may regain former functions and thus mimic various stages of differentiation of embryonic or immature cells. Rather than being committed to production of only one specific immunoglobulin molecule, they may be capable of synthesizing several forms and could ultimately provide evidence bearing both on the two-gene-one immunoglobulin polypeptide chain hypothesis, and on the nature of genic and allelic exclusion of immunoglobulins, as well.

VI.5.3. An Additional Approach to the Estimation of Mutation Rates in a Small Genetic Isolate (Ward, Neel). - One drawback in deriving accurate estimates of mutation rate in natural populations lies in our inability to actually measure mean mutant survival, $\bar{t}_0$. Lacking this parameter any estimates of locus specific mutation rates must be necessarily approximate, even if the ascertainment of genetic variants was complete. A method of analysis that promises to yield estimates of $\bar{t}_0$ as well as a determination of how $\bar{t}_0$ is affected by the population structure of the deme (mating patterns, migration rates, differential fertility, etc.) is to simulate the fate of a neutral gene in a pedigree for a given deme. Given an actual pedigree, introduction of a neutral mutant allele at a specified point in the pedigree followed by observations on the transmission of the allele through successive generations will allow estimates of $\bar{t}_0$ that are realistic for that population. In addition the effect of the observable population structure on $\bar{t}_0$ can be quite accurately determined. Hitherto, this approach has been limited to pedigrees which are obtained by a Monte Carlo simulation process (Sect. V.1.6.) since complete pedigree data for genetic isolates are not generally available. However extensive pedigrees which are as nearly complete and accurate as such data is likely to be, are available for a small variety of human isolates in different

parts of the world. One such population, characterized by exceptionally thorough pedigree data, is the Tokelau Islands group, located some 300 miles north of Samoa, in the South West Pacific. This Polynesian population can be considered as three distinct isolates, each isolate being located on a single coral atoll. For each of the isolates, extensive pedigree data are available for at least 9 generations.

It is proposed to supplement the studies of rare variants among American Indians with studies on another (recently) primitive population, the Tokelau Islanders.

As well as having extensive and accurate pedigree data the Tokelau Island population is especially suitable for a study of mutation rates in a human isolate for the following reasons:

- 1) Extensive and complete records are available for the majority of the population (3,300 individuals). These records, which incorporate medical, dietary, sociological, demographic and anthropological data, have been accumulated since 1968 as part of the Tokelau Islands Migrant Study (TIMS) based in New Zealand. Ward has been associated with this study as a Consultant¹ for the past year and for the next 18-24 months will have a pro forma position on the TIMS project, allowing access to this body of data.

- 2) Although not a "virgin" tribal society, the Tokelauans resident in the Tokelau Islands (1700) have by reason of their geographic isolation maintained their traditional dietary and subsistence patterns almost in their entirety. The rest of the population (1600 individuals) have emigrated from their atoll environment in the past decade, and most now reside in New Zealand. Thus to an approximation, the observed mutation rate in this population will not have been influenced by the complex of factors associated with Western civilization.

- 3) The three atolls form internal contrasts in that one (Atafu) was founded 9 generations ago by ten individuals and as such forms an example of an expanding population, while the other two (Fakaofu and Nukunonu) have been relatively stable populations during the past 10 generations.

- 4) The extensive genealogical records allow a complete characterization of the migration between the three atolls, while the

geographic isolation of the island group has kept migration from the outside to a minimal level. Thus the population can be visualized as a distinct genetic isolate, subdivided into three demes, one of which (Atafu) is of recent origin. Such a clear-cut situation does not often exist for natural populations, human or otherwise.

As part of a general biomedical survey, nuclear genealogies were collected by two anthropologists. The accuracy of these genealogies which extend back 9-14 generations, is attested by the independent records of visiting explorers, missionaries, etc., from the mid-nineteenth century to the present day. So far the complete genealogical record spanning 9 generations has been reconstructed for one atoll (Atafu) from a founding population of ten individuals to a present day census of 650 individuals. This has been used to document the increase of average inbreeding due to finite population size; the effect of the observed mating pattern compared with random mating, given the same demographic structure of the population, etc. The pedigrees for the other two atolls are currently being reconstructed as part of ongoing research by the TMS project, and should be completed by July-August, 1973. Once completed, a unique set of pedigrees will exist linking three genetic isolates in a very specific way, since the migrants to any one isolate can be specified exactly with respect to their relationship to the isolate from which they came. Two phases concerned with the estimation of mutation rates are planned for this study.

1) An extensive collection of blood samples so that the whole population can be extensively screened for rare variants. The sampling will be carried out in two phases: An initial phase, which will coincide with a biomedical survey by the TMS team to the Tokelau Islands in early to mid-1973. At this time, working in conjunction with the medical team, some 800-1200 samples will be taken (the exact number depending both on the sampling design and the constraints imposed by the screening lab) and shipped to Ann Arbor where they will be frozen and stored for screening at a later date. Following this a secondary sample of Tokelauans currently resident in New Zealand, will be collected by members of the TMS group as follow-up biomedical surveys are conducted in the period May-October 1973. A maximum of 1200 samples can be achieved

in the second stage--the exact number depending on the limitations of the screening laboratory. Although we anticipate that the majority of samples will be tested by October 1973, screening of the samples may well extend beyond this period. In this case an independent application for a grant would be submitted to ensure completion of the screening process. The frequency of rare variants and frequency of mutational events will then be analyzed for each of the three atolls and the results compared with the frequencies obtained from tribal societies in South America (Sect. VI.2.2.7.) and from the Japanese program (Sect. VI.3.).

2) The second phase of this study which will run concurrently with the sampling phase will concern the detailed analysis of the actual pedigrees in terms of generating realistic estimates of $\bar{t}_0$. This will be done by introducing a neutral allele into the actual pedigree under a variety of situations, and documenting its fate (cf. Sect. VI.1.3.). In this way the fate of a neutral allele can be fairly accurately described in this population over a temporal span of 9-14 generations. In addition sample pedigrees may be simulated to extend the temporal range beyond a maximum of 15 generations. In this instance, a unique opportunity exists of checking the simulation process in a diachronic manner, rather than the more usual limitation to comparison of synchronic events. Also certain parameters may be changed in order to derive an estimate of the effect of differential fertility, structured mating patterns, etc., on $\bar{t}_0$ in such a population. Finally the results can be compared with those obtained on a simulated Indian tribal society (Sect. VI.1.3.).

VI.5.4. The Continuing Participation of Bloom in the Cytogenetic Program of the ABCC. - During his proposed visit, in '73-'74 as Consultant to the Department of Clinical Laboratories at ABCC, Dr. Bloom will review with the professional staff of the Cytogenetics Laboratory the following specific studies:

1) The F₁ Study: By December, 1973 (or January, 1974), an additional 400-500 cases will have been examined cytogenetically. The efficacy of the contacting methods will be evaluated, as well as the findings, especially those coming from the banding techniques.

2) The NE-200 Study: Data from the in vitro , and neutron radiation studies will be available at this time, and both acute (in vitro) and delayed (in vivo) dose-response curves will be worked out with Drs. Awa, Sofuni, and their colleagues. Further, the low dose exposure group will be summarized in terms of cytogenetic aberrations, with an eye to publishing a report on the findings, which will be very valuable in the setting of radiation standards.

3) The Lymphocyte Lines: Mr. S. Iida should have established by then a small bank of cell lines from heavily exposed persons with a lot of residual aberrations. We will initiate the cloning experiments, and attempt to begin to generate data at this time on selection within cytogenetically abnormal cell lines.

4) The Immunology Study: While it is unlikely that this study of the immunological competence of heavily exposed persons will have been initiated by this time, further efforts will be made to pave the way for its implementation.

VI.6. CORE SUPPORT

This section as usual is entirely administrative. The necessary budget will be found in the financial section.

VI.7. REFERENCES

- Asher, J. H. Jr. Parthenogenesis and genetic variability. Ph. D. Thesis, University of Michigan, Ann Arbor, Michigan. 1970.
- Asher, J. H. Jr. Parthenogenesis and genetic variability. II. One-locus models for various diploid populations. *Genetics* 66: 369-391. 1970.
- Bodmer, W. F., and Cavalli-Sforza, L. L. A migration model for the study of genetic drift. *Genetics* 59: 565-592. 1968.
- Brewer, G. J., Sing, C. F., and Sears, E. R. Studies of isozyme patterns in nullisomic-tetrasomic combinations of wheat. *Proc. Nat. Acad. Sci.* 64: 1224-1229. 1969.
- Bylund, E. Theoretical considerations regarding the distribution of settlement in inner north Sweden. *Geografiska Annaler* 42: 225-231. 1960.
- Carson, H. L. Selection for parthenogenesis in Drosophila mercatorum. *Genetics* 55: 157-171. 1967.
- Carson, H. L., Wei, I. Y., and Niederkorn, J. A. Jr. Isogenicity in parthenogenetic strains of Drosophila mercatorum. *Genetics* 63: 619-628. 1969.
- Cavalli-Sforza, L. L., and Bodmer, W. F. The genetics of human populations. San Francisco: Freeman and Co. pp. xvi & 965. 1971.
- Cavalli-Sforza, L. L. et al. Paper presented at Conference on Population Structure. Hawaii July 1972.
- Charlesworth, B., and Giesel, J. T. Selection in populations with overlapping generations. II. The relations between gene frequency and demographic variables. *Amer. Natur.* 106: 388-401. 1972.
- Cockerham, C. C. Variance of gene frequencies. *Evolution* 23: 72-84. 1969.
- Cockerham, C. C. Analysis of gene frequencies. *Genetics* (In press).
- Crow, J. F., and Kimura, M. An Introduction to Population Genetics Theory, New York: Harper & Row. 1972.
- Evens, W. J. The sampling theory of selectively neutral alleles. *Theoretical Pop. Biol.* 3: 87-112. 1972.
- Feller, W. On fitness and the cost of natural selection. *Genet. Res.* 9: 1-15. 1967.
- Franklin, I., and Lewontin, R. C. Is the gene the unit of selection? *Genetics* 45: 707-734. 1970.
- Friedlaender, J. S. Population studies in south central Bougainville. Ph.D. Thesis. Harvard University. 1969.

- Frøland, S. S., and Natvig, J. B. Class, subclass and allelic exclusion of membrane-bound Ig of human β lymphocytes. *J. Exp. Med.* 136: 409-414. 1972.
- Goldstein, J. L. et al. Genetics of Hyperlipidemia in Coronary Heart Disease. In: *Transactions of the Association of American Physicians.* 1972.
- Haight, F. A. Some statistical problems in connection with word association data. *J. Math. Psych.* 3: 217-233. 1966.
- Haldane, J. B. S. The cost of natural selection. *J. Genet.* 55: 511-524. 1957.
- Isard, W. Location and space economy. Arnold, New York. 1956.
- King, J. L., and Jukes, T. H. Non-Darwinian evolution. *Science* 164: 788-798. 1969.
- Kirkmann, M. N., Schettini, F., and Pickard, B. M. Mediterranean variant of G6PD. *J. Lab. Clin. Med.* 63: 726. 1964.
- Latter, B. D. H. The island model of population differentiation: a general solution. *Genetics* 73: 147-157. 1973.
- Leslie, P. H. On the use of matrices in certain population mathematics. *Biometrika* 33: 183-212. 1945.
- Leslie, P. H. Some further notes on the use of matrices in population mathematics. *Biometrika* 35: 213-245. 1948.
- Lobb, N., Curtain, C. C., and Kidson, C. The synthesis of immunoglobulin G by cultured human lymphocytes. *Nature* 214: 783-785. 1967.
- Lobb, N. Regulatory genes controlling the synthesis of heavy chains of human immunoglobulin G. *Aust. J. Exp. Med. Sci.* 46: 397-405. 1968.
- Lösch, H. Die räumliche Ordnung der Wirtschaft. Jena. 1940.
- Malécot, G. Quelques schémas probabilistes sur une variabilité des populations naturelles. *Ann. Univ. Lyon Sci.* A13: 37-50. 1950.
- Malécot, G. Identical loci and relationship. *Proc. 5th Berk. Symp. Math. Statist. Probability* 6: 317-332. 1967.
- Morrill, R. L. Simulation of central place patterns over time. *Lund. Studies in Geography, Human Geography* 24: 109-120. 1962.
- Neel, J. V. The detection of increased mutation rates in human populations. *Persp. Biol. Med.* 14: 522-537. 1971.
- Neel, J. V., Tiffany, T. O., and Anderson, N. G. Approaches to monitoring human populations for mutation rates and genetic disease. In: Chemical Mutagens, Vol. 3 (A. Hollaender, ed.) In press.
- Nei, M. Identity of genes and genetic distance between populations. *Genetics* 68: 547. 1971.
- Nei, M. Genetic distance between populations. *Amer. Nat.* 106: 283-292. 1972.

- Nei, M., and Feldman, M. Identity of genes by descent within and between populations under mutation and migration premises. *Theoret. Pop. Biol.* 3: 460-465. 1972.
- Rivat, L., Gilbert, D., and Ropartz, C. The in vitro synthesis of subclasses of human G as demonstrated by a study on allotypes: The genes of the Gm system--are they structural genes? *Immunology* 19: 959-966. 1970.
- Simon, H. A. On a class of skew distribution functions. *Biometrika* 42: 425-440. 1954.
- Sing, C. F., and Brewer, G. J. Isozymes of a polyploid series in wheat. *Genetics* 61: 391-398. 1969.
- Sing, C. F., and Brewer, G. J. Evidence for nonrandom multiplicity of gene products in 22 plant genera. *Biochem. Genet.* 5: 243-251. 1971.
- Stevenson, A. C., and Kerr, L. B. On the distributions of frequencies of mutation to genes determining harmful traits in man. *Mut. Res.* 4: 339-352. 1967.
- Templeton, A. R. Statistical models of parthenogenesis. Ph.D. Thesis, University of Michigan, Ann Arbor, Michigan. 1972.
- Templeton, A. R., and Rothman, E. D. The population genetics of parthenogenetic strains of Drosophila mercatorum. I. One-locus model and statistics. *Theoretical and Applied Genetics*. (In press).
- Turner, J. R. G. Wright's adaptive surface, and some general rules for equilibria in complex polymorphisms. *The American Naturalist* 105: 267-278. 1971.
- Wagener, D. K. An extension of migration matrix analysis to account for differential immigration from the outside world. *Amer. J. Hum. Genet.* 25: 47-56. 1973.
- Ward, R. H., and Neel, J. V. Gene frequencies and microdifferentiation among the Makiritare Indians, IV. Comparison of a genetic network with ethnohistory and migration matrices; a new index of genetic isolation. *Amer. J. Hum. Genet.* 22: 538-561.
- Wills, C. J. et al. A computer model allowing maintenance of large amounts of genetic variability in Mendelian populations. I. Assumptions and results for large populations. *Genetics* 64: 107-123. 1970.
- Yule, G. V. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. *Phil. Trans. Roy. Soc., Lond. 'B'* 213: 21-87. 1923.

VII. FEDERAL SUPPORT FOR DEPARTMENTAL PROGRAMS

	<u>PRINCIPAL INVESTIGATOR</u>	<u>% TIME</u>	<u>TITLE</u>	<u>TOTAL PERIOD OF SUPPORT</u>	<u>TOTAL AMOUNT FOR PERIOD</u>	<u>CURRENT ANNUAL SUPPORT</u>
A. AEC - Cont. at AT(11-1)-1-22)	J. Neel with Gershowitz, Sing, Bloom and Smouse	40	Area Program in Populat- ion Genetics	11/1/68-10/31/73	NEGOTIABLE	\$402,010
B. Nat'l. Sci. Fdn. (thru AEC)	J. Neel			4/1/70-10/31/73	\$537,851	\$162,000
Nat'l. Sci. Fdn.	E.H.Y. Chu	20	Genetic Control of Mam- malian Cell Metabolism	2/1/73-1/31/75	\$ 50,000	\$ 50,000
C. USPHS						
<u>Actual</u>						
Training Grant 5-T01-QM-71-15	J. Neel	5	Graduate Training Program in Genetics	7/1/68-6/30/73	\$1,113,375	\$228,398
<u>Research Grants</u> QM-15419-06	M. Levine with A. Bloom D. Rucknagel D. Shreffler R. Tashian J. Klein		Area Program in Cellular and Biochemical Genetics	10/1/72-9/30/77	\$2,890,763	\$525,851
NHLI-72-291-8 (thru Wayne State University)	G. Brewer	15	Studies of Sickle Cell Disease	6/15/73-6/14/74	NEGOTIABLE	\$ 64,969
NICHD-71-24-1	A. Bloom	15	Nat'l. Registry for Amniocentesis	6/27/72-6/28/74	NEGOTIABLE	\$ 24,757
RCDA-6-K3-H-24980	J. Neel on behalf of D. Shreffler	100	Gen. Studies of Blood Group & Serum Poly- morphisms	7/1/71-6/30/76	NEGOTIABLE	\$ 30,686
<u>Pending</u>						
	E.H.Y. Chu	10	Selection and Propagation of Somatic Cells Having Specific Physiological Mutations	10/1/73-9/30/75	\$ 140,392	
	C.F. Sing	10	Genetic Unit of Selection	9/1/73-8/31/75	\$ 34,646	

FEDERAL SUPPORT

CONT'D. (Pending)

	<u>PRINCIPAL INVESTIGATOR</u>	<u>% TIME</u>	<u>TITLE</u>	<u>TOTAL PERIOD OF SUPPORT</u>	<u>TOTAL AMOUNT FOR PERIOD</u>	<u>CURRENT AN. SUPPORT</u>
	P. Smouse	25	Population Studies of Microbial Ecosystems	9/1/73-8/31/76	\$ 44,042	
D. National Edtn.	A. Bloom with R. Schmickel	10	Clinic for the Diagnosis & Counseling of Birth Defects	1/1/73-12/31/74	\$ 49,782	\$ 24,89
E. U.S. Army Medical Res. & Devel. Command (DADA-17-60-9103)	G. Brewer	15	The Effect of Erythro- cyte Intermediate on Oxygen Dissociation Properties of Hemoglobin	3/1/69-2/28/74	NEGOTIABLE	\$ 86,50

IX. AEC - EQUIPMENT

<u>Date Ordered</u>	<u>Item</u>	<u>Amount</u>
1/18/71	*Nitrogen freezer, refrig. conversion kit	\$ 335.00
2/04/72	Buchler polystaltic pump	277.00
2/28/72	Balance; incubator; tissue culture hood; 2 shaker, incubator	2845.40
2/28/72	Spectrophotometer	880.00
2/29/72	Microzone cell; 2 bridge assemblies; sample applicator	1170.00
2/29/72	2 Microzone cells; 1 bridge assembly; 2 sample applicators	1290.00
3/09/72	Revco storage chest w/alarm system & casters	2243.00
3/13/72	2 Frigidaire refrigerators	480.55
3/13/72	1 Cryo-flask	1960.00
3/22/72	*Sony tape recorder	122.72
5/05/72	Parts for DS-12 disk storage system	890.00
5/05/72	IBM 1133 Multiplex control enclosure	1328.75
5/08/72	Cal Comp DS-12 disk storage system	20000.00
6/06/72	Frigidaire refrigerator	141.00
6/16/72	*6' Extended leg tri-pod	121.75
8/16/72	Refrigerator	230.50
9/01/72	Quebec colony counter	130.00
9/19/72	Autoclave electric sterilizer & 4 stainless steel trays	1122.00
9/19/72	Immunodiffusion camera and extra lens	727.00
9/21/72	Incubator	458.00
10/25/72	*Model 140 Statistician; Model 100 Statistician	1264.00
11/10/72	*Model 2316 Disk pack	241.50
12/04/72	Polystaltic pump	279.25
12/06/72	MSE Ultrasonic disintegrator probe	120.75

*Ordered on NSF 036380

PROPOSED BUDGET

VI.1. The Continuing Program on the American Indian (Nash)

<u>PERSONNEL</u>	<u>1973-74</u>	
Res. Assoc. (R. Ward, 100%)	\$17,490	
Res. Assoc. (F. Li, 100%)	15,900	
Res. Assoc. (R. Spielman, 100%)	15,900	
Res. Assoc. (L. Jensen, 3 Mos.)	4,000	
Lab. Tech. (Open, 8 Mos.)	5,200	
Lab. Tech. (A. Mao, 100%)	9,010	
Lab. Asst. (S. Nevins, 100%)	<u>6,800</u>	
	\$74,300	
Fringe Benefits (12% of S & W)	<u>8,916</u>	
TOTAL PERSONNEL		\$ 83,216
<u>MOVABLE EQUIPMENT</u>		
Small miscellaneous items	\$ 1,500	
TOTAL MOVABLE EQUIPMENT		\$ 1,500
<u>CONSUMABLE SUPPLIES¹</u>		
<u>Laboratory Supplies</u>		
692 Serotypings at \$6.50/specimen	\$ 4,498	
Reagents for enzyme studies (600 Specimens) (See Attached)	4,408	
Liquid Nitrogen	1,749	
<u>Field Work</u>		
(See attached Budget)	<u>12,019</u>	
TOTAL CONSUMABLE SUPPLIES		\$ 22,674
<u>TRAVEL</u>		
Domestic - To attend 1-2 Scientific Meetings at \$250 ea. for 4 individuals	\$ 2,000	
Commuting, Penn State Park, Pa., to Ann Arbor (Dr. N. Chagnon)	<u>600</u>	
TOTAL TRAVEL		\$ 2,600
TOTAL DIRECT COST		\$109,990
INDIRECT COST 62.5% of S & W PLUS FRINGE		<u>52,010</u>
		<u>162,000</u>

FIELD WORK BUDGET - NSF - 1973-74

O-Helix Field Work

CONSUMABLE SUPPLIES

Vacutainers - ACD 20ml, 2400	\$ 329.40
Vacutainers - Empty 20 ml, 2400	357.30
Vacutainer Needles (20 ga., 1 1/2") 2400	179.52
Gauze Sponges, 2 x 2, 3000	51.58
Drugs and Medical Supplies	401.20
Polaroid Film (150 boxes)	500.00
2400 Samples at \$3.00/sample	7,200.00
Contingency	<u>3,000.00</u>
TOTAL COST FOR FIELD WORK	<u>\$12,019.00</u>

¹An Application in the amount of \$14,645 has been submitted to NSF for the transportation expenses of this field work. However, we anticipate an unexpended balance of \$10,000 on the NSF portion of this total budget. If we may carry this over, the request for supplementary funding from NSF can be reduced from \$14,645 to \$6,645.

Detailed Budget for Screening Lab; NSF Component

Estimate of Cost to Process and Type 17 Standard Systems
in 600 Blood Samples

I. Nonchemical

Electrostarch	40 kg at \$55/8 kg	\$ 275.00
Liquid Nitrogen Supplies		
1. Cryovials	10 vials per sample at \$.11 per vial	660.00
2. Canes	600	60.00
3. Liquid Nitrogen	26 cylinders	750.00
4. Marker Pens	15	11.00
Poloroid Film (B and W)	60 rolls	390.00
Glassware		
1. Pasteur pipettes	30 gross	42.00
2. Gel flasks	15	82.50
3. Capillary pipettes		7.00
4. Miscellaneous glassware		200.00
Petrolatum	2 cans	3.00
Masking Tape		
1. 1" x 60 yards	20 rolls	10.00
2. 1/2" x 60 yards	30 rolls	9.00
Saran Wrap	135 rolls	54.00
Filter Paper Sheets	2 boxes	44.00
Subtotal		<u>\$2,597.50</u>

II. Chemicals

Benzidine HCl	30g	\$ 3.00
Sodium Nitroferrocyanide	30g	.50
Sodium Pyrophosphate	500g	1.00
Methanol	6 gal	15.00
Sodium Hydroxide	500	2.00
KCN	6g	.50
Chloroform	600 ml	1.50
Sodium Chloride	6 lbs	1.50
D-Sorbitol	250g	1.50
Glycerine	1 kg	1.50
Toluene	1,200 ml	.50
Acetone	600 ml	1.00
Ionagar #2	150g	5.00
MTT	6g	110.00
PMS	4.5g	14.50
Adenosine	1g	1.00
Amido Black	6g	1.00
Charcoal	2 lb	7.00
Dextrose	5g	.50
Phenolphthaleine Diphosphate	20g	66.00
Dianisidine	50g	2.00
Ethanol	15 liters	12.00
Fructose-6-PO ₄	3g	30.00
Lactate (Na)	300 ml	1.00
MgCl ₂	30g	.50
DL-Maleic Acid	250g	1.50

MnCl ₂	10g	.50
H ₂ O ₂	1 liter	1.50
KI	70g	1.00
N ₂ Arsenate • 7H ₂ O	30g	.50
Na ₃ Isocitrate	6g	30.00
NaH ₂ PO ₄	500g	1.00
Na ₂ HPO ₄	1 kg	3.00
Na Thiosulfate	10g	.50
ADP (Na ₂)	1g	22.00
β-DPN	6g	60.00
TPN	6g	570.00
Glucose-1-PO ₄	50g	60.00
Glucose-6-PD	5 ml	112.00
6-Phosphogluconate	1g	50.00
Hexokinase	3 ml	24.00
Nucleoside Phosphorylase	3 ml	45.00
Peroxidase	3g	120.00
Snake Venom	1.5g	55.00
Xanthin Oxidase	3 ml	50.00
Val-Leu	3g	45.00
Leu-Gly-Gly	3g	110.00
Citric Acid	500g	1.50
Boric Acid	1 kg	4.00
EDTA	0.5 lbs	1.00
Glycine	500g	4.00
DL-Histidine	250 gm	18.00
Lithium Hydroxide	30	1.00
Maleic Acid	400g	7.50

Sodium Acetate	500g	1.00
Sodium Citrate	20 kg	30.00
Tris	3 kg .	87.00
Acetic Acid	3 x 5 lb btcls	6.00
HCl	1 pt	1.00
NH ₄ OH	600 ml	.50
EDTA • Na ₂	300g	6.00
	Subtotal	\$1,811.00
	Grand Total	<u>\$4,408.00</u>

PROPOSED BUDGET

V1.3, Supporting Laboratory Activities (Neel)

<u>PERSONNEL</u>	<u>1973-74</u>	
Princ. Investigator, (R. Gershowitz, 50%)	\$14,530	
Res. Assoc. (R. Tanis, 50%)	10,070	
Res. Assoc. (R. Ferrell, 100%)	15,900	
Japanese Biochemist (100%)	17,500	
Lab. Tech. (M. Morrow, 100%)	10,628	
Lab. Asst. (J. Lynn, 100%)	<u>7,000</u>	
	\$75,628	
Fringe Benefits (12% of S & W)	<u>9,075</u>	
TOTAL PERSONNEL		\$ 84,703
<u>MOVABLE EQUIPMENT</u>		
Automation*		\$ 30,000
<u>CONSUMABLE SUPPLIES</u>		
Reagents for blood typings (400 bloods at \$6.50)	\$ 2,600	
Media, Reagents and Glassware (See Attached)	3,351	
Liquid Nitrogen	1,500	
Transport of blood specimen from Oceania	<u>600</u>	
TOTAL CONSUMABLE SUPPLIES		\$ 8,051
<u>TRAVEL</u>		
<u>International</u> - 5 Roundtrips Hiroshima, Japan at \$1,158 ea. (Drs. Neel, Ferrell, Tanis, Kishimoto and Wife)	\$ 5,790	
150 Days Per Diem at \$20/day	3,000	
<u>Domestic</u> - To attend 1-2 Scientific Meetings at \$400	<u>1,600</u>	
TOTAL TRAVEL		\$ 10,390
TOTAL DIRECT COST		\$133,144
INDIRECT COST 62.5% of S & W PLUS FRINGE		<u>52,939</u>
TOTAL COST		<u>\$186,083</u>

*Based on development of GENSAEC Software at Oak Ridge

Detailed Budget for Screening Lab; AEC Component

Estimate of Cost to Type 6 New Systems in 600 Samples

I. Nonchemical

Electrostarch	16 kg at \$55/8 kg	\$ 110.00
Liquid Nitrogen	52 cylinders	1,500.00
Polaroid Film (B and W)	10 rolls	70.00
Glassware		
1. Pasteur pipettes	5 gross	7.00
2. Gel flasks	5	27.00
3. Miscellaneous glassware		150.00
Petrolatum	2 cans	3.00
Masking Tape		
1. 1" x 60 yards	10 rolls	5.00
2. 1/2" x 60 yards	15 rolls	5.00
Saran Wrap	40 rolls	16.00
Filter Paper Sheets	1 box	22.00

II. Chemicals

1,436.00

\$3,351.00

PROPOSED BUDGET

VI.4. Mathematical and Statistical Aspects of Population Control
(Sing)

<u>PERSONNEL</u>	<u>1973-74</u>	
Assoc. Prof. (C. Sing, 50%)	\$11,660	
Res. Assoc. (R. Griffith, 100%)	19,080	
Assistant (E. Rothman, 25%)	5,000	
Asst. in Res. (M. Horikawa, 100%)	10,998	
Asst. in Res. (B. Hsiao, 50%)	<u>6,095</u>	
	\$52,833	
Fringe Benefits (12% of S & W)	<u>6,340</u>	
TOTAL PERSONNEL		\$ 59,172
<u>MOVABLE EQUIPMENT</u>		0
<u>CONSUMABLE SUPPLIES</u>		
Machine Time IBM 360-67	\$ 3,500	
Service on Calculators	350	
Supplies (Paper and IBM cards)	<u>1,948</u>	
TOTAL CONSUMABLE SUPPLIES		\$ 5,798
<u>RENTAL AND OTHER SERVICE CHARGES</u>		
Service contract on 1130 system	\$ 4,242	
Service contract on IBM 029-A22 Key Punch	369	
Service contract on IBM 029-C22 Key Punch	369	
Rental of IBM Card Sorter 082/yr.	1,600	
Rental on Data Set to connect 1130 system to University of Michigan IBM 360-67	942	
Maintenance DS-12 Disk Drive	<u>1,680</u>	
TOTAL RENTAL AND OTHER SERVICE CHARGES		\$ 9,202
<u>TRAVEL</u>		
Domestic - To attend 1-2 Scientific Meetings (2 ea. for 3 investigators at \$250)		<u>\$ 1,500</u>
TOTAL DIRECT COST		\$ 75,672
INDIRECT COST 62.5% of S & W PLUS FRINGE		<u>36,982</u>
TOTAL COST		\$112,654

PROPOSED BUDGET

VI.5.1. The Genetic Unit of Selection (Sing)

<u>PERSONNEL</u>	<u>1973-74</u>	
Asst. in Res. (B. Thirtle, 100%)	\$10,610	
Fringe Benefits (12% of S & W)	<u>1,273</u>	
TOTAL PERSONNEL		\$11,883
 <u>MOVABLE EQUIPMENT</u>		
Small miscellaneous items		100
 <u>CONSUMABLE SUPPLIES</u>		
Supplies for starch gel electrophoreses (Starch, TPN, TPNH, DPNH, Glucose-6-phosphate, and numerous other substrates, coenzymes, and a number of enzymes used in staining the starch gel for specific and supplies for fly laboratory)		\$ 2,500
 <u>TRAVEL</u>		
Domestic - To attend 1-2 Scientific Meetings (Dr. Sing)		\$ 600
TOTAL DIRECT COST		<u>\$15,083</u>
INDIRECT COST 62.5% of S & W PLUS FRINGE		<u>7,426</u>
TOTAL COST		<u>\$22,509</u>

PROPOSED BUDGET

VI.5.2. Genetic, Chemical, and Serological Studies of Human
Immunoglobulins (Gershowitz)

<u>PERSONNEL</u>	<u>1973-74</u>	
Res. Asst. (W. Niezgoda, 100%)	\$13,310	
Fringe Benefits (12% of S & W)	<u>1,597</u>	
TOTAL PERSONNEL		\$14,907
 <u>MOVABLE EQUIPMENT</u>		
1 NAPCO - CO ₂ Incubator		\$ 2,200
 <u>CONSUMABLE SUPPLIES</u>		
Antisera	\$ 800	
Chemicals	600	
Glassware	600	
Blood from donors	750	
Animal purchase and maintenance	<u>2,500</u>	
TOTAL CONSUMABLE SUPPLIES		\$ 5,250
 <u>TRAVEL</u>		
Domestic - Blood collections	\$ 400	
To attend 1-2 Scientific Meetings	<u>800</u>	
TOTAL TRAVEL		<u>\$ 1,200</u>
 TOTAL DIRECT COST		\$23,557
INDIRECT COST 62.5% of S & W PLUS FRINGE		<u>9,316</u>
TOTAL COST		<u>\$32,873</u>

PROPOSED BUDGET - AEC - 1973-74

VI.5.4 Continuing Cytogenetic Studies in Japan (Bloom)

<u>TRAVEL</u>	<u>1973-74</u>
International - 1 Roundtrip Airfare - Detroit- Hiroshima, Japan (January or February 1974)	\$1,158
21 Days Per Diem at \$20/day	420
Miscellaneous expenses	<u>22</u>
TOTAL COST	<u>\$1,600</u>

PROPOSED BUDGET - ACC - 1973-74

VI.6. Core Section

<u>PERSONNEL</u>	<u>1973-74</u>	
Admin. Asst. (J. Starks, 50%)	\$ 5,035	
Princ. Clerk (K. Hilden, 50%)	4,466	
Secretary (A. Preston, 50%)	4,000	
Secretary (R. Jones, 100%)	9,000	
Secretary (M. Kellogg, 50%)	3,335	
Engr. Tech. (T. Bell, 50%)	5,077	
Student Res. Assistantship (50%)	<u>4,000</u>	
	34,913	
Fringe Benefits (12" of S & W)	<u>4,189</u>	
TOTAL PERSONNEL		\$39,102
 <u>CONSUMABLE SUPPLIES</u>		
Office Supplies	\$ 2,000	
Telephone Communications and toll charges	<u>1,019</u>	
TOTAL CONSUMABLE SUPPLIES		\$ <u>3,019</u>
TOTAL DIRECT COST		42,121
INDIRECT COST 62.5% of S & W PLUS FRINGE		<u>24,438</u>
TOTAL COST		\$66,559

XII. PUBLICATIONS SINCE LAST ANNUAL REPORT

Bloom, A. D. Induced chromosomal aberrations in man. In: Advances in Human Genetics, H. Harris, and K. Hirschhorn (eds). New York: Plenum Press. pp. 99-172. 1972.

Bloom, A. D. Induced chromosomal aberrations: Biological and clinical significance. *J. Pediat.* 81: 1-8. 1972.

Heuch, I. The effect of a fusion of subpopulations on the total fixation index. *Theoret. Applied. Genet.* 42: 327-330. 1972.

Neel, J. V. The detection of increased mutation rates in human populations. In: Mutagenic Effects of Environmental Contaminants, H. E. Sutton and M. I. Harris (eds). New York: Academic Press. pp. 99-119. 1972.

Neel, J. V., and Schull, W. J. Differential fertility and human evolution. In: Evolutionary Biology, Vol. 6, T. Dobzhansky, M. K. Hecht, and W. C. Steere (eds). New York: Appleton-Century-Crofts. pp. 363-378. 1972.

Neel, J. V., and Ward, R. H. The genetic structure of a tribal population, the Yanomama Indians. VI. Analysis by F-statistics (including a comparison with the Makiritare and Xavante). *Genetics* 72: 639-666. 1972.

Neel, J., Arends, T., Brewer, C., Chagnon, N., Gershowitz, H., Layrisse, M., Layrisse, Z., MacCluer, J., Migliazza, E., Oliver, W., Salzano, F., Spielman, R., Ward, R., and Weitkamp, L. Studies on the Yanomama Indians. In: Human Genetics, Proc., IV Int. Cong. Hum. Genet., Paris, Sept. 1971. Amsterdam: Excerpta Medica. pp. 96-111. 1972.

Rothhammer, F., and Spielman, R. S. Anthropometric variation in the Aymará: Genetic, geographic, and topographic contributions. *Amer. J. Hum. Genet.* 24: 371-380. 1972.

Rothhammer, F., Neel, J. V., Da Rocha, F., and Sundling, G. Y. The genetic structure of a tribal population, the Yanomama Indians. VIII. Dermatoglyphic differences among villages. *Amer. J. Hum. Genet.* 25: 152-166. 1973.

Salzano, F. M., Gershowitz, H., Junqueira, P. C., Woodall, J. P., Black, F. L., and Hierholzer, W. Blood groups and H-Le^a salivary secretion of Brazilian Cayapo Indians. *Am. J. Phys. Anthropol.* 36: 417-426. 1972.

Salzano, F. M., Neel, J. V., Weitkamp, L. R., and Woodall, J. P. Serum proteins, hemoglobins and erythrocyte enzymes of Brazilian Cayapo Indians. *Human Biol.* 44: 443-458. 1972.

Schanfield, M. S., Gershowitz, H., Ohkura, K., and Blackwell, R. Q. Studies on the immunoglobulin allotypes of Asiatic populations. II. Gm and Inv allotypes in Chinese. *Human Hered.* 22: 178-193. 1972.

Schanfield, M. S., Gershowitz, H., Hong, K-Ja, and Shim, B-Sop. Studies on the immunoglobulin allotypes of Asiatic populations. III. Gm and Inv allotypes among random Koreans. *Human Hered.* 22: 144-148. 1972.

Scherer, C. A., Dorschel, C. A., Cook, J. M., and LeQuesne, P. W. A model iron-catalyzed biomimetic cyclization of a cyclic tryptamine N-oxide. *J. Org. Chem.* 37: 1083-1085. 1972.

Schull, W. J. Genetic implications of population breeding structure. In: The Structure of Human Populations, G. A. Harrison and A. J. Boyce (eds). Oxford: Clarendon Press. pp. 146-164. 1972.

Schull, W. J., and Neel, J. V. The effects of parental consanguinity and inbreeding in Hirado, Japan. V. Summary and interpretation. *Amer. J. Hum. Genet.* 24: 425-453. 1972.

Sinnock, P., and Sing, C. F. Analysis of multilocus genetic systems in Tecumseh, Michigan. I. Definition of the data set and tests for goodness-of-fit to expectations based on gene, gamete, and single-locus phenotype frequencies. *Amer. J. Hum. Genet.* 24: 381-392. 1972a.

Sinnock, P., and Sing, C. F. Analysis of multilocus genetic systems in Tecumseh, Michigan. II. Consideration of the correlation between nonalleles in gametes. *Amer. J. Hum. Genet.* 24: 393-415. 1972b.

Spielman, R. S., Da Rocha, F. J., Weitkamp, L. R., Ward, R. H., Neel, J. V., and Chagnon, N. A. The genetic structure of a tribal population, the Yanomama Indians. VII. Anthropometric differences among Yanomama villages. *Am. J. Phys. Anthropol.* 37: 345-356. 1972.

Ward, R. H. The genetic structure of a tribal population, the Yanomama Indians. V. Comparisons of a series of genetic networks. *Ann. Hum. Genet.*, Lond 36: 21-43. 1972.

Ward, R. H. Some aspects of genetic structure in the Yanomama and Makiritare, two tribes of Southern Venezuela. In: Methods and Theory in Anthropological Genetics, M. H. Crawford, and P. L. Workman (eds). University of New Mexico Press. 1973.

Weiss, K. M. Demographic models for anthropology. *Society for American Archeology, Memoirs* No. 27. (*American Antiquity* 38: no. 2, part 2).

Weiss, K. M. A general measure of human population growth regulation. *Am. J. Phys. Anthropol.* 37: 337-343. 1972.

Weiss, K. M. On the systematic bias in skeletal sexing. *Am. J. Phys. Anthropol.* 37: 239-249. 1972.

Weitkamp, L. R., and Neel, J. V. The genetic structure of a tribal population, the Yanomama Indians. IV. Eleven erythrocyte enzyme and serum of proteins. *Ann. Hum. Genet., Lond.* 35: 423-441. 1971.

XIII. PAPERS CURRENTLY IN PRESS

- Adams, J. P., and Ward, R. H. Admixture studies and the detection of selection. *Science*, in press.
- Bergold, G. H., Sequera, C. L., Saturno, A., and Neel, J. V. Encuesta serologica para los grupos arbovirus entre Indios Venezolanos (Yanomamos, Maquiritare, Bare). *Acta Cientifica Venezolana*, in press.
- Bloom, A. D., Neel, J. V., Tsuchimoto, T., and Meilinger, K. Chromosomal breakage in leukocytes of South American Indians. *Cytogenetics*, in press.
- Dunston, G., and Gershowitz, H. Further studies of Xh, a serum protein antigen in man. *Vox Sanguinis*, in press.
- Glenn, K., Schull, W. J., and Sing, C. F. Linkage among 12 genetic markers in the Tecumseh, Michigan population. *Amer. J. Hum. Genet.*, in press.
- Layrisse, Z., Layrisse, M., Malavé, I., Terasaki, P., Ward, R. H., and Neel, J. V. Histocompatibility antigens in a genetically isolated American Indian tribe. *Amer. J. Hum. Genet.*, in press.
- Li, F. H. F., and Neel, J. V. A simulation of the fate of a mutant gene of neutral selective value in a primitive population. In: *Proceedings of Conference on the Uses of Computer Simulation in Human Population Studies*, The Pennsylvania State University, in press.
- Neel, J. V. A note on congenital defects in two unacculturated Indian tribes. *Congenital Defects; New Horizons*, in press.
- Neel, J. V. Diversity within and between South American Indian tribes. *Sheba Symposium*, in press.
- Sing, C. F., Chamberlain, M. A., and Eggleston, B. K. An analysis of variance of gene frequencies in a human population. In: *Human Population Structure*, N. Morton (ed). Honolulu: University of Hawaii Press, in press.
- Sing, C. F., Brewer, G. J., and Thirtle, B. Inherited biochemical variation in *Drosophila melanogaster*: Noise or signal? I. Single locus analyses. *Genetics*, in press.
- Spielman, R. S. Do the natives all look alike? Size and shape components of anthropometric differences among Yanomama Indian villages. *Am. Naturalist*, in press.
- Spielman, R. S. Differences among Yanomama Indian villages: Do the patterns of allele frequencies, anthropometrics and map locations correspond? *Am. J. Phys. Anthropol.*, in press.

Szathmary, E. V., Cox, D. W., Gershowitz, H., Rucknagel, D. L., and Schanfield, M. S. The Northern and Southwestern Ojibwa: Serum proteins and red cell enzyme systems. *Am. J. Phys. Anthropol.*, in press.

Tanis, R. J., Neel, J. V., Dovey, H., and Morrow, M. The genetic structure of a tribal population, the Yanomama Indians. IX. Serum protein and enzyme differences between the Yanomama and 5 neighboring tribes. *Amer. J. Hum. Genet.*, in press.

Templeton, A. R. Density dependent selection in parthenogenetic and self-mating populations. *Theoret. Pop. Biol.*, in press.

Templeton, A. R., and Rothman, E. D. The population genetics of parthenogenetic strains of Drosophila mercatorum. I. One locus model and statistics. *Theoret. Applied Genet.*, in press.

Weiss, K. M. A method for approximating age-specific fertility in the construction of life tables for anthropological populations. *Hum. Biol.*, in press.

Weitkamp, L. R., Salzano, F. M., Neel, J. V., Porto, P., Geerdink, R. A., and Tárnoky, A. L. Human serum albumin: Twenty-three genetic variants and their population distribution. *Ann. Hum. Genet.*, in press.