RECEIVED
APR 04 2000
OSTI

DE-FG02-94ER61883
"Preparation of PAC Libraries"

Pieter J. de Jong

FINAL TECHNICAL REPORT

Established goals of project:

To create P1 Artificial Chromosome (PAC) cloning vectors and use these vectors to generate, characterize, and distribute both human and mouse genomic PAC libraries to the scientific community.

Project achievements:

1. Development of PAC and BAC cloning vectors: Efforts to overcome the limitations of cosmids and YACs have resulted in alternative cloning approaches using a bacteriophage P1-derived vector and P1-derived packaging extracts or based on the *E. coli* fertility plasmid (F-factor). The P1-derived vector has been used with a bacteriophage T4 *in vitro* packaging system, enabling the recovery of recombinants with inserts up to 122 kb in size. However, the efficiency of these procedures to make slightly larger P1-clones is very low and had not yet resulted in complex recombinant DNA libraries.

The bacteriophage P1 cloning system had the convenience of being commercially available. However, the insert size in recombinant clones is limited to the 70-90 kb range due to the P1 packaging mechanism. Furthermore, the choice of bacterial hosts is restricted to *E. coli* strains expressing the P1 site-specific recombinase gene, *cre*. This enzyme is needed to circularize the linear DNA injected by the P1 phage particles. Attractive properties of the P1 system include the ability to select for recombinants over non-recombinants and the presence of two P1-derived replication mechanisms. The single-copy number replicon is used mostly for stable propagation of the clones. The multi-copy number replication mechanism is suppressed by the bacterial lac-repressor and can be induced using IPTG. The higher copy number is important for DNA isolation and should also facilitate the detection of clones by colony-hybridization procedures. The main disadvantages of this system are the modest increase in the size of the inserts compared to cosmids, the requirement for an elaborate *in vitro* packaging system and the restriction to the use of host-cells expressing *cre* enzyme. The alternative T4 packaging system may allow some further increase in insert size but does not mitigate the other limitations.

The F factor-based vector utilizes the well-studied *E. coli* Fertility plasmid for the construction of Bacterial Artificial Chromosomes (BACs). The DNA is introduced into bacterial cells by electroporation of circular ligation products, thus avoiding elaborate packaging extracts and *in vivo* site-specific recombination. Inserts up to 300 kb have been cloned in bacterial cells using this vector. However, unlike the P1 system, there is no positive selection for recombinants, requiring the identification of recombinant clones by colony hybridization with human repetitive DNA probe. Also, DNA recovery from the clones is low as a result of the replication of F-factor at 1-2 copies per cell.

We have developed a hybrid approach which combines the desirable aspects of the P1 vector (e.g. positive selection for recombinants by saccharose using the SacB gene) with the electroschock transformation protocols of the BACs. This approach makes it possible to create libraries with low-levels of cloning artifacts, with low levels of non-recombinants and increased insert size compared to conventional P1 cloning. However, the stringent restriction of P1-particle size has been replaced by a strong size bias in favor of clones with smaller inserts. This bias is mostly the result of the transformation approach. Clones with large or very large inserts have been obtained as rare clones from our first generation human library. These clones are stable and generate mostly the same size subclones upon retransformation into *E. coli*. The largest clone obtained so far from our human PAC library is 400 kb.

Clones with larger inserts are desirable to reduce mapping complexity, to decrease DNA sequencing redundancy and to facilitate gene isolation procedures (e.g. exon trapping). In addition, there is an increasing need within the scientific community [outside the genome mapping world], with respect to reintroduction of large-insert clones to generate transgenic animals or modify the phenotype of somatic cells or tissues. We

DISCLAIMER

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

DISCLAIMER

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

have therefore developed a PAC shuttle vector, pPAC4, which permits transfer of clones from bacteria to mammalian cells. PAC DNA replicate using the Epstein Barr virus replicon (oriP) if the EBV EBNA1 replication protein is provided in trans. Transfected cells can be selected using a dominant marker gene: the blasticidin-S-methylase gene under SV40 promoter control. We have used the pPAC4 vector to generate a 10.6-fold redundant PAC library of 129 strain mouse. In addition to the new PAC vector, we have also constructed a modular BAC vector (pBACe3.6) with new cloning enzyme choices and the option to select for recombinant clones using sucrose in the medium. The use of different vectors (PAC versus BAC) and different cloning enzymes (MboI versus EcoRI) will permit the generation of BAC libraries complementary to the PAC libraries previously constructed.

PAC Cloning Vectors:

BAC Cloning Vector:

2. Construction and characterization of genomic human and mouse PAC libraries: We have used the pCYPAC2 and pPAC4 cloning vectors to generate a number of human and mouse genomic DNA PAC libraries. These libraries have been extensively characterized as to average clone insert size, clone size distribution, level of non-recombinant clones, level of chimeric clones, and calculated genomic representation. The human PAC libraries are summarized below:

<u>Name</u>	<u>Source</u>	<u>Vector</u>	<u>Total 384 Well Plates</u>	<u>Total Clones</u>	<u>Total Non-Recombinants</u>	<u>Average Insert Size</u>	<u>Genomic Coverage</u>
RPCI-1	Male	pCYPAC2	321	115,200	30,000	110 kbp	3X
RPCI-3	Male	pCYPAC2	207	75,513	6,500	115 kbp	3X
RPCI-4	Male	pCYPAC2	288	105,251	<1,000	116 kbp	4X
RPCI-5	Male	pCYPAC2	384	142,773	<2,000	115 KBP	6X
RPCI-6	Female	pPAC4	240	87,897	<1,700	135 kbp	4X

The RPCI-21 mouse PAC library constructed using funds awarded for this project is summarized below:

<u>Library Segment</u>	<u>Cloning Vector</u>	<u>DNA</u>	<u>Plate Numbers</u>	<u>Total Plates</u>	<u>Total Clones</u>	<u>Empty Wells (Total)</u>			
RPCI21 (Total)	pPAC4	129/SvEvTac Mouse Spleen	1-672	672	254,217	6,780			
RPCI21 (Segment 1)	pPAC4	129/SvEvTac	1-336	336	125,318	3,868			
RPCI21 (Segment 2)	pPAC4	Mouse Spleen	337-672	336	128,899	2,912			
	<u>Empty Wells (%)</u>	<u>Non-Recombinant Clones (Total)</u>		<u>Non-Recombinant Clones (%)</u>	<u>Insert Size (average)</u>	<u>Genomic Coverage</u>			
RPCI21 (Total)	2.5	1312		0.5	137 Kbp	11.6X			
RPCI21 (Segment 1)	3.0	650		0.5	127 Kbp	5.3X			
RPCI21 (Segment 2)	2.2	662		0.5	147 Kbp	6.3X			

3. Generation of a genomic C57BL/6J mouse BAC library: Using partial funding from this award, we have generated a large insert C57BL/6J mouse library in the pBACe3.6 cloning vector, extensively characterized and distributed this clone collection to the scientific community. The RPCI-23 mouse BAC library is summarized on the next page.

The RPCI-23 C576J/BJ mouse BAC Library:

Segment	Cloning Vector	DNA	Plate Numbers	Total Plates	Total Clones	Empty Wells (Total)
1	pBACe3.6	Female (kidney/brain)	1-240	240	80,661	1,717
2	pBACe3.6	Female (kidney/brain)	241-480	240	89,973	2,207
Total Library				480	170,634	3,924

Segment	Empty Wells (%)	Non-Recombinant Clones (Total)	Non-Recombinant Clones (%)	Insert Size (average)	Genomic Coverage
1	1.9	approx. 10,523	11.3	198 Kbp	5.3X
2	2.4	approx. 1,211	1.3	197 Kbp	5.9X
Total Library	2.1	11,734	6.8	197 Kbp	11.2X

4. Distribution of PAC and BAC library resources: Distribution of PAC and BAC library resources that were generated under this award occurs via three broad mechanisms.

A.) Distribution of library culture plate arrays: By taking care of the library plate distribution from our laboratory, we have been able to facilitate rapid transfer of the library copies and maintain optimal library quality. We have been able to replicate library copies within one to four weeks after receipt of the request. The RPCI human PAC libraries have now been distributed to over forty separate genome centers/laboratories and/or commercial biotech companies (see table below). Many of these centers are major genome sequencing centers (eg.: Washington University, St. Louis; Whitehead Institute for Genome Research; and The Sanger Centre, Cambridge) that have been using the PAC clones in assembling high throughput human genome sequence. In addition, we have distributed the human PAC libraries to two US companies involved in providing screening service to the user community (Genome Systems, RPCI-1; Research Genetics, RPCI-1,-3,-4,-5 and -6). In Europe, copies have been distributed to resource centers (eg. UK Resource Centre, Hinxton, UK; and the Reference Library System (Max Planck Institute, Berlin) to serve the European users. Copies of the RPCI-21 mouse PAC library have been distributed to six mouse genome centers as well as the Library Reference System (resource center) and UK Resource Centre in Europe and Research Genetics in the USA (see table below).

B.) Distribution of high density hybridization filters: We are providing users with library screening filters that have been produced in our laboratory on the "Q-Bot" and the "BioGrid" robotic gridding systems. The clones are gridded onto the filters in a 4 x4 clone duplicate array such that a positive clone will appear as two distinct spots in one of eight possible "vector" patterns on the autorad image of the probed filter. Each filter contains over 36,000 clones (>18,000 individual clones). We have found that the end users of these filters have no difficulty interpreting the data from this type of filter grid and locating the positive clones in the original library plates. To date, over 5,000 of the high-density filters of the funded PAC and BAC libraries have been produced and distributed to laboratories worldwide.

C.) Distribution of individual positive clones: We also distribute individual clones from the funded libraries to small scale users who do not have direct access to the library plates. The clones are provided on a cost recovery basis. Users have obtained high-density hybridization filters from our lab and perform the screening in their own laboratories. Upon reading the screening results, they contact us with the locations of their positives clones and we pick, culture and ship these clones to them either as glycerol stocks (US labs) or LB stabs (overseas shipment). We have been able to ship these clones within two to four days after receipt of the request. To date, we have supplied over 5000 RPCI human clones directly to the end user. We have found that this is an effective method to increase the use of our libraries while providing access to laboratories that due to budget constraints, would not be able to obtain the complete clone collection.

Center/Institute Name	Contact	Library ID	Replicate Date
RPCI-1:			
1. Max-Planck Institute, Germany	Lehrach	R4;R5	Mar. 94
2. Sanger Centre, Cambridge, U.K.	Bentley	R6	Mar. 94
3. Genome Systems, St. Louis, MO	Smoller	R1;R12	July 94
4. Hospital for Sick Children, Toronto, Canada	Tsui	R8-6	Nov. 94
5. Welcome Trust Centre, Oxford, U.K.	Cox/Monaco	R8-5	Nov. 94
6. Human Genome Mapping Resource Centre, U.K.	Nyberg	R9-10	Mar. 95
7. Bios, Inc., New Haven, CT	Kouri	R9-8	July 95
8. Yale University, CT	Ward	R8-1	Mar. 95
9. University of California, San Francisco, CA	Gray	R7-2	Oct. 94
10. Lawrence Livermore National Laboratory, CA	Carrano	R7-1	Oct. 94
11. Los Alamos National Laboratory, NM	Deaven	R7-5	Oct. 94
12. Nat. Center for Human Genome Research, NIH	Collins	R7-4	Oct. 94
13. Baylor College of Medicine, TX	Nelson	R7-8	Oct. 94
14. Lawrence Berkeley Laboratory, CA	Rubin	R7-3	Oct. 94
15. Genome Therapeutics, CA	Mao	R7-7	Oct. 94
16. Eleanor Roosevelt Institute, CO	Patterson	R7-9	Oct. 94
17. Washington University, St. Louis, MO	Brownstein	R10	Oct. 94
18. Albert Einstein College of Medicine, NY	Kucherlapati	R7-10	Oct. 94
19. Stanford University, CA	Myers/Cox	R7-6	Oct. 94
20. Boston University, MA	Amemiya	R8-8	Nov. 94
21. University of Wurzburg, Germany	Weber	R8-10	Nov. 94
22. Childrens Hospital of Philadelphia, PA	Emmanuel	R8-7	Nov. 94
23. Univ. of TX, SW Medical Center, TX	Evans	R8-4	Nov. 94
24. Duke University Medical School, NC	Marchuk	R8-9	Nov. 94
25. National Public Health Institute, Finland	Aaltonen	R8-11	Nov. 94
26. CEPH, France	Neri	R9-9	Mar. 95
27. Univ. of Mainz, Childrens Hospital, Germany	Zabel	R8-12	Mar. 95
28. Columbia University, NY	Fisher	R9-4	Mar. 95
29. University of California, Irvine, CA	Wasmuth	R9-3	Mar. 95
30. Univ. of TX, Health Science Center, TX	Naylor	R9-5	Mar. 95
31. Riken, Tsukuba Life Science Center, Japan	Soeda	R9-7	Mar. 95
32. State University of Leiden, The Netherlands	van Ommen	R9-6	Mar. 95
33. Cedar-Sinai Medical Center, CA	Korenberg	R8-14	Mar. 95
34. Glaxo, NC	Shephard	R9-1	Dec. 95
35. Whitehead Institute for Genome Research	Lander	R8-3	Mar. 96
36. Sequana Therapeutics, La Jolla, CA	Wadkins	R9-14	Oct. 96
37. Research Genetics, Huntsville, AL	Hudson	R9-16	June 96
38. SmithKline Beecham, UK	Browne	R9-15	Nov. 96

Center/Institute Name	Contact	Library ID	Replicate Date
<u>RPCI-3:</u>			
1. Sanger Centre, Cambridge, U.K.	Bentley	R1-3	Oct. 95
2. Univ. of TX, SW Medical Center, TX	Evans	R1-4	Nov. 95
3. Albert Einstein College of Medicine, NY	Kucherlapati	R1-5	Nov. 95
4. Welcome Trust Centre, Oxford, U.K.	Cox/Monaco	R1-6	Nov. 95
5. Glaxo, NC	Shephard	R1-7	Dec. 95
6. Washington University, St. Louis, MO	Brownstein	R1-8	Jan. 96
7. Whitehead Institute for Genome Research	Lander	R2-2	Mar. 96
8. Max-Planck Institute, Germany	Lehrach	R2-3	May 96
9. RIKEN, Tsukuba Life Science Center, Japan	Soeda	R1-9	Feb. 96
10. Research Genetics, Huntsville, AL	Hudson	R2-1	June.96
11. Sequana Therapeutics, La Jolla, CA	Wadkins	R2-5	Oct. 96
12. Columbia University, NY	Fischer	R2-6	Sept.96
13. Millennium Pharmaceuticals, Cambridge, MA	Duyk	R2-4	Sept.96
14. Baylor College of Medicine, Houston	Gibbs	R2-8	June 97
15. Chinese Univ. of Hong Kong	Waye	R2-8	July 97
<u>RPCI-4:</u>			
1. Washington University, St. Louis, MO	Brownstein	R1-3	Jan. 96
2. Sanger Centre, Cambridge, U.K.	Bentley	R1-4	Oct. 96
3. Whitehead Institute for Genome Research	Lander	R2-2	Mar. 96
4. Max-Planck Institute, Germany	Lehrach	R2-3	May 96
5. Univ. of TX, SW Medical Center, TX	Evans	R1-5	Feb. 96
6. Research Genetics, Huntsville, AL	Hudson	R2-1	June.96
7. Millennium Pharmaceuticals, Cambridge, MA	Duyk	R2-4	Sept.96
8. Sequana Therapeutics, La Jolla, CA	Wadkins	R2-5	Oct. 96
9. Karolinska Hospital, Sweden	Dumanski	R2-6	Aug. 96
10. RIKEN, Tsukuba Life Science Center, Japan	Soeda	R2-8	Nov. 96
11. Albert Einstein College of Medicine, NY	Kucherlapati	R2-9	Nov. 96
12. Baylor College of Medicine, Houston	Gibbs	R2-10	June 97
13. Univ. of Texas Health Science Center, San Antonio	Garcia	R2-11	July 97
<u>RPCI-5:</u>			
1. Washington University, St. Louis, MO	Brownstein	R1-3	Jan. 96
2. Sanger Centre, Cambridge, U.K.	Bentley	R1-4	Oct. 96
3. Whitehead Institute for Genome Research	Lander	R2-2	Mar. 96
4. Max-Planck Institute, Germany	Lehrach	R2-3	May 96
5. Univ. of TX, SW Medical Center, TX	Evans	R1-5	Feb. 96
6. Research Genetics, Huntsville, AL	Hudson	R2-1	June.96
7. Whitehead Institute for Genome Research	Lander	R2-2	Mar. 96
8. Millennium Pharmaceuticals, Cambridge, MA	Duyk	R2-4	Sept.96
9. Sequana Therapeutics, La Jolla, CA	Wadkins	R2-5	Oct. 96
10. Karolinska Hospital, Sweden	Dumanski	R2-6	Aug. 96
11. RIKEN, Tsukuba Life Science Center, Japan	Soeda	R2-9	Nov. 96
12. Albert Einstein College of Medicine, NY	Kucherlapati	R2-10	Nov. 95
13. The Institute of Cancer Research, UK	Cooper	R2-8	Oct. 96
14. Institute of Genetics and Biochemical Evol., Italy	Sala	R2-11	Jan. 97
15. Baylor College of Medicine, Houston	Gibbs	R2-10	June 97
16. Univ. of Texas Health Science Center, San Antonio	Garcia	R2-11	July 97

Center/Institute Name	Contact	Library ID	Replicate Date
<u>RPCI-6:</u>			
1. Millennium Pharmaceuticals, Cambridge, MA	Duyk	R3	Sept. 96
2. Columbia University, NY	Fischer	R4	Sept. 96
3. Research Genetics, Huntsville, AL	Hudson	R5	June 96
4. Karolinska Hospital, Sweden	Dumanski	R6	Aug. 96
5. Max-Planck Institute, Germany	Lehrach	R7	Jan. 97
6. Sequana Therapeutics, La Jolla, CA	Wadkins	R8	Oct. 96
7. RIKEN, Tsukuba Life Science Center, Japan	Soeda	R2-1	Nov. 96
8. The Biotechnology Centre of Oslo, Norway	Prydz	R2-2	Nov. 96
9. Leiden University, The Netherlands	den Dunnen	R2-3	Nov. 96
10. Sanger Centre, Cambridge, U.K.	Bentley	R2-4	Dec. 96
11. Washington University, St. Louis, MO	Brownstein	R2-5	Dec. 96
12. Baylor College of Medicine, TX	Gibbs	R2-6	Feb. 97
13. Albert Einstein College of Medicine, NY	Kucherlapati	R10	April 97
14. Baylor College of Medicine, Houston	Gibbs	R2-10	June 97

<u>RPCI-21:</u>			
1. RIKEN, Tsukuba Life Science Center, Japan	Hayashizaki	R6 Seg 1&2	April 97
2. Sanger Centre, Cambridge, U.K.	Bentley	R5 Seg 1&2	April 97
3. Leiden University, The Netherlands	den Dunnen	R6 Seg. 2	April 97
4. Max-Planck Institute, Germany	Lehrach	R7 Seg. 1&2	May 97
5. Baylor College of Medicine, TX	Nidamarthy	R8 Seg. 1&2	June 97
6. Research Genetics	Hudson	R9 Seg. 1&2	Sept. 97
7. UK Resource Centre	Gorick	R10 Seg. 2	Sept. 97