2001

A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration

Scott C. Fahrenkrug
University of Minnesota, SCOTT-F@biosci.cbs.umn.edu

Gary A. Rohrer
USDA-ARS

Brad A. Freking
USDA-ARS, brad.freking@ars.usda.gov

Timothy P. L. Smith
USDA-ARS, tim.smith@ars.usda.gov

Kazutoyo Oswoegawa
Children's Hospital Oakland Research Institute

See next page for additional authors

Mammalian Genome 12, 472–474 (2001).
DOI: 10.1007/s003350020015

Mammalian Genome
Incorporating Mouse Genome

A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration

Scott C. Fahrenkrug,[1,*] Gary A. Rohrer,[1] Brad A. Freking,[1] Timothy P.L. Smith,[1] Kazutoyo Osoegawa,[2,] Chung Li Shu,[2,**] Joseph J. Catanese,[2,**] Pieter J. de Jong[2,***]**

[1]USDA, ARS, U.S. Meat Animal Research Center, PO Box 166, Spur 18D, Clay Center, Nebraska 68933-0166, USA
[2]Roswell Park Cancer Institute, Dept. of Cancer Genetics, Elm and Carlton St., Buffalo, New York 14263, USA

Received: 16 November 2000 / Accepted: 01 February 2001

Recent advances in porcine genomics have identified quantitative trait loci (QTL) that influence pork production traits such as carcass traits, meat quality, and reproductive efficiency (Rohrer 2000; Cassady 1999). The low resolution to which most of these loci have been defined precludes the accurate application of marker-assisted selection (MAS) strategies for increasing production efficiency. Large-insert genomic libraries are an excellent resource for marker development aimed at increasing the resolution of QTL and for the development of contiguous physical maps (contigs) of the chromosomal regions containing them.

To facilitate applications requiring genomic clones, several porcine yeast artificial chromosome (YAC) libraries have been developed representing onefold (Leeb et al. 1995), threefold (Rogel-Gaillard et al. 1997), and 5.5-fold (Alexander et al. 1997) coverage of the pig genome. These YAC libraries have been valuable resource owing to their large insert size. However, marker isolation from YAC libraries is confounded by the equimolar representation of the complex yeast genome and YAC DNA, susceptibility to insert rearrangement, and a relatively high degree of chimerism. The development of bacterial artificial chromosome (BAC) libraries (Shizuya et al. 1992; Ioannou et al. 1994) represents a compromise between insert size, stability, and ease of clone DNA isolation. BACs are capable of stably maintaining insert sizes exceeding 200 kb and can be easily isolated by standard alkaline lysis from bacterial genomic DNA by virtue of their closed-circular conformation (Shizuya et al. 1992; Ioannou et al. 1994). The ease of BAC DNA isolation allows for efficient restriction analysis, subcloning, and direct BAC DNA sequencing, procedures invaluable for marker isolation and the development of contiguous physical maps.

Recently BAC libraries representing livestock species, including cattle, sheep, and pigs (Cai et al. 1995; Warren et al. 2000; Vaiman et al. 1999; Suzuki et al. 2000; Rogel-Gaillard et al. 1999) have been developed. Among these are two porcine BAC libraries representing fourfold (Suzuki et al. 2000) and fivefold (Rogel-Gaillard et al. 1999) coverage of the porcine genome. We undertook the development of a publicly available pig BAC library, RPCI-44, representing at least a tenfold coverage of the pig genome as a resource for physical map development and pig genome sequencing.

* *Present address:* University of Minnesota, Department of Animal Science, 495 An Sci/Vet Med, 1988 Fitch Ave., St. Paul, MN 55108, USA.

** *Present address:* Children's Hospital Oakland Research Institute, 747 Fifty-second St., Oakland, CA 94609-1809, USA.

*** *Present address:* Pfizer Global Research and Development, Alameda Laboratories, 1501 Harbor Bay Parkway, Alameda, CA 94502, USA.

Correspondence to: S.C. Fahrenkrug; E-mail: SCOTT-F@biosci.cbs.umn.edu

High-molecular-weight genomic DNA from four crossbred male pigs (breed composition: 37.5% Yorkshire, 37.5% Landrace, and 25% Meishan) was isolated as previously described (Osoegawa et al. 1998), partially digested with *Eco*RI (New England Biolabs, Beverly, MA) and *Eco*RI methylase (New England Biolabs), and fractionated by clamped homogeneous electrical field (CHEF; Chu et al. 1986) electrophoresis (BioRad, Hercules, CA). Digested genomic DNA in the range of 150–200 kpb was gel-isolated by electroelution and ligated to *Eco*RI-cut pTARBAC2 (Wang et al., unpublished; http://www.chori.org/bacpac/vectorframe.htm). Ligated DNA was drop-dialyzed on floating membranes, first against sterile water, followed by 0.5 × TE containing 30% PEG8000, and introduced by electroporation into DH1OB electrocompetent cells (Life Technologies, Rockville, MD). Transformed bacterial cells were allowed to recover in SOC medium for 1 h at 37°C, and were spread on LB agar plates containing 20 μg/ml chloramphenicol and 5% sucrose. Individual colonies were picked with an automatic colony-picking robot (Q-bot; Genetix, Stony Brook, NY) and arrayed into 528 individual 384-well microtiter dishes (Genetix) with LB medium containing 7.5% glycerol and 20 μg/ml chloramphenicol. The library was divided into two segments and gridded onto 11 high-density replica filters, each containing up to 18,432 independent clones that have been spotted in duplicate. To serve as anchor spots, an end-sequenced BAC clone (1A1) from a *Caenorhabditis brigsea* BAC library (RPCI-94), containing DNA apparently unique to nematode genomes, was gridded on the corner of each of the six fields that comprise the high-density filters. The inclusion of a probe directed against the anchor spots can facilitate orientation of hybridized filters.

The predicted coverage of RPCI-44 is a function of the total number of insert-containing clones and the size of the genomic DNA fragments they contain. From a total of 202,752 clones picked, 1,764 (segment 1, 1.9%) and 1,519 (segment 2, 1.4%) clones failed to grow, leaving a total of 199,469 clones.

Non-recombinant clones containing the original vector were identified by high-density colony filter hybridization with the vector as a probe. The probing identifies two classes of clones with strong versus weak hybridization signals corresponding to the presence or absence of the original vector stuffer fragment ("pUC-link"), which contains a high-copy-number replicon. Recombinant clones lack the stuffer fragment, resulting in much weaker hybridization signals. In total, 119 clones (0.06%) were identified as containing intact vector by this approach. The plate and well addresses of these are available at *http://www.chori.org/framenonrec.htm.* To evaluate the average insert size and insert size distribution in the library, randomly selected clones from each segment were subjected to restriction analysis with *Not*1, insert size determination by CHEF electrophoresis, and comparison with molecular weight standards (Osoegawa et al. 2000). Segment 1 was composed of 90,396 clones containing inserts averaging 157 kpb

Table 1. Characteristics of the RPCI-44 are summarized in this table. The library was divided into two segments. Non-insert clones were identified as a single smaller vector band after digestion with *Not*1 and analyzed with pulsed-field gel electrophoresis. The ratio of non-insert clones and average insert size were estimated by analyzing 144 clones from segment 1 and 148 clones from segment 2.

Segment	Number of plates	Non-insert clones (%)	Avg. insert size (%)	Number of recombinant clones	Genome-fold
1	240	7	157	84,007	4.9
2	288	6.9	171	101,489	6.4
Total	528	7	165	185,496	11.3

(Table 1, Fig. 1a). Segment 2 was composed of 109,073 clones containing inserts averaging 171 kpb (Table 1, Fig. 1a). The average insert size for the entire library was 165 kpb. Approximately 7% of the clones from each segment of the library did not appear to contain insert when assessed by restriction analysis (Table 1). These clones do not contain the intact vector since they fail to hybridize to the pUC19 stuffer fragment. Instead, as evidenced by reduced vector size upon *Not*1 restriction digestion and pulsed-field gel electrophoretic analysis, these clones correspond to vector molecules that have deletions. We postulate that these clones arise because of restriction enzyme star activity, resulting in nonspecific digestion in the vector and sacB inactivation. The deleted vectors can thus give rise to colonies able to propagate on sucrose selection media (Ioannou et al. 1994). Excluding these clones from the total results in a final tally of 185,496 clones with an average insert size of 165 kpb. Therefore, with a porcine genome size of 2.7×10^9 bp (Schmitz et al. 1992), the library contains enough clones to represent the genome with an 11.3-fold average redundancy.

Another method for determining the coverage of the BAC library is to determine the number of clones that contain selected markers. Because this type of determination also reveals the abundance of particular genomic segments, we chose to assess the coverage of the library on a per chromosome basis, using a collection of previously mapped microsatellite markers (Table 2). Overlapping oligonucleotides (overgo oligos; McPherson, 1999) were designed from sequences flanking microsatellite markers to screen the library by hybridization. Two 24-base oligonucleotides that have 8-pb complementary sequences at their 3′ ends were synthesized and labeled by primer extension in the presence of [α-^{32}P]dCTP and [α-^{32}P]dATP and hybridized to the high-density filters. After the primary library screening with a pool containing all of the overgo oligo probes, all positive clones were picked from the library and re-arrayed.

Secondary hybridization filters were gridded from the arrayed positive clones and screened with individual overgo oligo probes. The detailed procedure was essentially described previously (Osoegawa et al. 2000). The sequence of the overgo oligos and the clone identification numbers are available at http://borg.marc.usda.gov/pig_bac_library. The presence of the microsatellites on the identified BAC clones was verified by PCR amplification with marker-specific primers (http://sol.marc.usda.gov/genome/swine/htmls/chromosome list.html) essentially as described (Rohrer et al. 1994) by using approximately 1 ng of BAC DNA. The representation of each marker is presented in Table 2.

The average representation considering all these data is 11.5 ± 7.9 clones/marker. The number of clones identified did not follow a normal distribution, with the abundance of several clones inflating the mean and deviation. To address this, clones corresponding to the markers *SW274, SW2160, S0385, SW957, SW1056,* and *SW472* were further analyzed by *Eco*RI DNA fingerprinting (Marra et al. 1997). The predicted contigs were analyzed carefully to identify clonal rearrangements as described previously (Osoegawa et al. 2000). Analyses revealed that clones representing markers *SW274, S0385, SW957,* and *SW1056* do indeed correspond to single-copy loci with the number of clones in the contigs

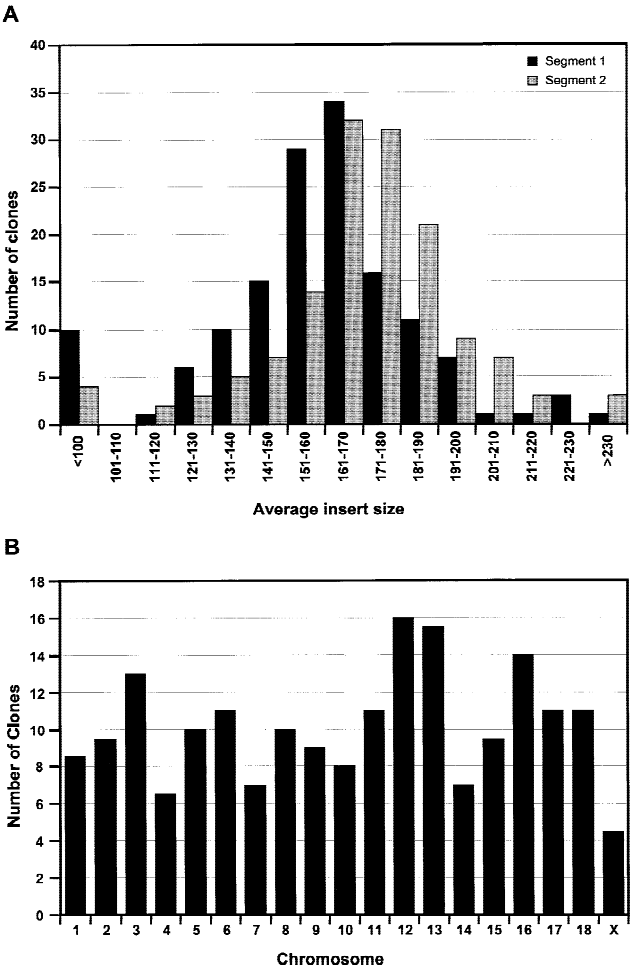

Fig. 1. A. In total, 144 clones from segment 1 (black) and 148 clones from segment 2 (gray) were randomly selected and analyzed by pulsed-field gel electrophoresis. The vertical axis indicates the number of clones in each size range. The horizontal axis refers to the size range of insert DNA. **B.** Assessment of the coverage of RPCI-44 on a per chromosome basis. The average number of clones isolated by using chromosome-specific markers is indicated. Two contigs corresponding to *SW2160* were identified. Though this marker has been genetically mapped to SSC8, which contig corresponds to the mapped DNA segment has not been determined. Therefore, the representation of SSC8 was predicted by the average of the number of clones in these contigs. Note that the number of clones expected for the X Chr is expected to be 1/2 of the autosome representation since the library is derived from male DNA.

indicated in Table 2. Analysis of clones identified as containing *SW2160* were found to belong to two non-overlapping contigs, with 14 and 6 clones, respectively, suggesting that a sequence homologous to *SW2160* is represented at least twice in the pig genome. Fingerprint analysis of the 47 clones corresponding to marker *SW472* was complicated by the presence of an *Eco*RI restriction-fragment (either 3.6 kb or 4 kb) that was present at a very disproportionate concentration relative to the other restriction fragments. Subcloning and sequencing of these restriction-fragments revealed them to contain sequence corresponding to *SW472.* This indicates that the *SW472* marker is included in a simple sequence direct repeat, possibly part of alpha-satellite elements within the porcine genome. In agreement with this proposal, PCR amplification of *SW472*-positive clones using *SW472*-specific primers and analysis by denaturing PAGE revealed the amplification of multiple products from each BAC (data not shown), indicating *SW472*-related sequence is not uniquely represented even within individual BAC clones. Genetic mapping studies relying on the use of

Table 2. Microsatellite markers used to screen RPCI-44 to determine chromosomal representation. The marker name, chromosome of origin, relative genetic position, and the number of clones isolated is indicated.

Marker	Chromosome	Position (cM)	Clones
SW1430	1	58.5	9
SW780	1	81.0	8
SW1650	2	27.3	10
SW1026	2	60.6	9
SW274	3	0	17
SW902	3	58.4	9
SW489	4	8.0	8
SW2454	4	51.2	5
SW1482	5	39.9	12
SW1468	5	97.5	8
SW322	6	149.8	11
SW1369	7	48.2	7
SW472	7	58.9	47
SW2160	8	80.1	20
SW2093	9	103.6	10
SW497	10	39.3	9
SW951	10	96.0	7
S0385	11	0	18
SW1632	11	16.6	4
SW957	12	33.4	16
SW935	13	25.4	15
SW1056	13	96.1	16
SO356	14	8.6	10
SW1081	14	72.1	4
SW1562	15	29.5	12
SW1401	15	62.5	7
SW2411	16	16.7	14
SW335	17	0	11
SW1808	18	0	12
SW1984	18	29.4	10
SW1608	X	101.9	8
SW2534	X	30.2	1

SW2160 and *SW472* should take into account the presence of multiple targets within the pig genome.

Recalculating the number of clones corresponding to each marker, excluding *SW472*, and independently counting the clones from the two contigs isolated with *SW2160*, results in a mean of 10 ± 4 clones/marker, in close agreement with the representation predicted by BAC clone number and insert size. Consideration of the chromosomal representation as indicated by average marker content, with the SSCX expected to be haploid in a a male-derived library, is also consistent with a ten-fold coverage of the porcine genome.

We have constructed a porcine BAC library, RPCI-44, from four male crossbred pigs. Partially redundant sequencing of RPCI-44 will result in an elevated detection of polymorphisms compared with libraries developed from individual or isogenic animals, without hindering clone isolation. Given the role that polymorphic loci play in meiotic mapping, the genetic diversity captured by this library is especially valuable for the isolation of markers for physical and genetic map integration. Successful fingerprint analysis conducted in the course of this work, as well as the generation of genome-wide physical contigs using this library (Warren et al. 2001), demonstrates that the gain in polymorphism content does not decrease the utility of RPCI-44 for physical mapping. The value of the RPCI-44 porcine BAC library for isolating clones containing specific porcine genes has been demonstrated by the successful isolation of clones corresponding to more than 4000 pig genes (Warren et al. 2001). Its utility in providing markers suitable for genetic mapping has been shown via the development of more than 15 gene-linked markers polymorphic in the U.S. Meat Animal Research Center (MARC) porcine reference population (Fahrenkrug et al.; Smith et al.; Campbell et al.; Rohrer et al. unpublished data). The 10- to 11-fold genomic coverage represented by RPCI-44, in conjunction with the other porcine libraries mentioned here, will be of great value for the isolation of genetic and physical markers, the generation of comprehensive physical maps, and the construction of minimum tiling paths for region-specific or whole-genome sequencing. Information regarding the public distribution of RPCI-44 at cost can be found at http://www.chori.org/bacpac/.

Acknowledgments. Thanks to Sherry Kluver for expert secretarial assistance. Thanks to Jeni Kuszak for technical support. K. Osoegawa has been supported with a Postdoctoral Fellowship for Research Abroad from the Japan Society for the Promotion of Science (JSPS).

References

Alexander LJ, Smith TP, Beattie CW, Broom MF (1997) Construction and characterization of a large insert porcine YAC library. Mamm Genome 8, 50–51

Cai L, Taylor JF, Wing RA, Gallagher DS, Woo SS (1995) Construction and characterization of a bovine bacterial artificial chromosome library. Genomics 29, 413–425

Cassady J (1999) Identification of quantitative trait loci affecting reproduction in pigs, Ph.D. Dissertation, University of Nebraska, Lincoln, NE

Chu G, Vollrath D, Davis RW (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science 234, 1582–1585

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H et al. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat Genet 6, 84–89

Leeb T, Reftenberger G, Hameister H, Brem G, Brenig B (1995) Construction of a porcine YAC library and mapping of the cardiac muscle ryanodine receptor gene to chromosome 14q22-q23. Mamm Genome 6, 37–41

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B et al. (1997) High throughput fingerprint analysis of large-insert clones. Genome Res 7, 1072–1084

Osoegawa K, Woon PY, Zhao B, Frengen E, Tateno M et al. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 52, 1–8

Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, Hayashizaki Y, de Jong PJ (2000) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. Genome Res 10(1), 116–128

Rogel-Gaillard C, Bourgeaux N, Save JC, Renard C, Coullin P et al. (1997) Construction of a swine YAC library allowing an efficient recovery of unique and centromeric repeated sequences. Mamm Genome 8, 186–192

Rogel-Gaillard C, Bourgeaux N, Billault A, Vaiman M, Chardon P (1999) Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements. Cytogenet Cell Genet 85, 205–211

Rohrer G (2000) *Genetics of the Pig*, In: *Biology of the Domestic Pig*, 2nd ed. Mersmann H, Pond W, eds. 122–149

Rohrer GA, Alexander LJ, Keele JW, Smith TP, Beattie CW (1994) A microsatellite linkage map of the porcine genome. Genetics 136, 231–245

Schmitz A, Chaput B, Fouchet P, Guilly MN, Frelat G, Vaiman M (1992) Swine chromosomal DNA quantification by bivariate flow karyotyping and karyotype interpretation. Cytometry 13, 703–710

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T et al. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA 89, 8794–8797

Suzuki K, Asakawa S, Iida M, Shimanuki S, Fujishima N et al. (2000) Construction and evaluation of a porcine bacterial artificial chromosome library. Anim Genet 31, 8–12

Vaiman D, Billault A, Tabet-Aoul K, Schibler L, Vilefte D et al. (1999) Construction and characterization of a sheep BAC library of three genome equivalents. Mamm Genome 10, 585–587

Vollrath D (1999) DNA markers for physical mapping. In: *Genome Analysis: A Laboratory Manual*, vol. 4. B Birren, ED Green, P Hieter, S Klapholz, RM Myers et al., (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press), pp 207–213

Warren W, Smith TPL, Rexroad CE, Fahrenkrug SC, Allison T et al. (2000) Construction and characterization of a new bovine bacterial artificial chromosome library with 10 genome-equivalent coverage. Mamm Genome 11, 662–663

Warren W, Tao N, Barbazuk B, Allison T, Landewe T et al. (2001) A BAC based physical map of the *Sus Scrofa* genome. Plant and Animal Genome IX Conference, San Diego, CA, P589