

BAC Isolation, Mapping, and Sequencing

BAC Isolation and Mapping

BAC clones were isolated from the following libraries:

Species	BAC Library	Redundancy	Reference
Chimpanzee (<i>Pan troglodytes</i>)	RPCI-43 (BACPAC Resource)	~3.5-fold	
Baboon (<i>Papio cynocephalus anubis</i>)	RPCI-41 (BACPAC Resource)	~10.4-fold	
Cat (<i>Felis catus</i>)	RPCI-86 (BACPAC Resource)	~10.6-fold	
Dog (<i>Canis familiaris</i>)	RPCI-81 (BACPAC Resource)	~8.1-fold	Ref. 1
Cow (<i>Bos taurus</i>)	RPCI-42 (BACPAC Resource)	~11.9-fold	Ref. 2
Pig (<i>Sus scrofa</i>)	RPCI-44 (BACPAC Resource)	~10.2-fold	Ref. 3
Rat (<i>Rattus norvegicus</i>)	RPCI-31 (BACPAC Resource)	~10.5-fold	Ref. 4
Mouse (<i>Mus musculus</i>)	RPCI-23 (BACPAC Resource)	~11.2-fold	Ref. 5
Chicken (<i>Gallus gallus</i>)	White Leghorn JFID256 Red Jungle Fowl 31-JF256-BI (TAMU)	~5.4-fold ~5.2-fold	Refs. 6 & 7
Fugu (<i>Fugu rubripes</i>)	BAC/Fugu (Genome Systems, Inc.)	~10.0-fold	
Tetraodon (<i>Tetraodon nigroviridis</i>)	GSTNA	~7.0-fold	Ref. 8
Zebrafish (<i>Danio rerio</i>)	BAC/Zebrafish (Genome Systems, Inc.)	~3.0-fold	

Note that the initial 12 species selected for sequencing included two non-human primates (chimpanzee and baboon), two carnivores (cat and dog), two artiodactyls (cow and pig), two rodents (rat and mouse), one bird (chicken), and three fish (two pufferfish, fugu and tetraodon, and zebrafish). This group simply reflects species for which BAC libraries were available at the time we initiated our program.

BACs were obtained from the BACPAC Resource (see www.chori.org/bacpac) for the first eight libraries, from the Texas A & M GENEfinder Genomic Resource (see BACserver.tamu.edu) for the chicken library, from Genome Systems, Inc. for the fugu and zebrafish libraries, and from Drs. J. Weissenbach and H. Roest Crolius (Genoscope, France) for the tetraodon library.

Clone isolation from the mouse BAC library was performed as described⁹. Clone isolation and characterization from the rat BAC library were described previously¹⁰. For the other mammals, ‘universal’ hybridization probes¹¹ were used for clone isolation and nascent contig construction, when necessary supplemented with probes derived from orthologous species-specific or BAC-end sequences. Fugu clones were isolated by use of available orthologous genomic sequence¹² and subsequent walking with BAC-end sequences. Isolation of zebrafish and chicken clones involved generating probes specific for identified orthologous sequences, when available, and subsequent contig extension by walking with BAC-end sequences. Tetraodon clones were selected computationally based

on the identification of tetraodon BAC-end sequences⁸ that aligned to orthologous fugu sequence.

BAC Sequencing

BACs were sequenced using a standard shotgun-sequencing strategy^{13,14}. In brief, purified BAC DNA (see genome.wustl.edu/tools/protocols) was kinetically sheared with a Hydroshear instrument (GeneMachines, San Carlos, CA), and the resulting fragments were end repaired with T4 DNA polymerase and Klenow. BstXI/EcoRI linkers (Invitrogen, Carlsbad, CA) were ligated to the end-repaired fragments, and the ligated DNA was then size selected (1.5-3.0 kb) by agarose gel electrophoresis and subcloned into the plasmid pOTWI3. Sequence reads were generated from both insert ends of randomly selected subclones using BigDye dye-terminator chemistry and model 3700 automated DNA sequencing instruments (Applied Biosystems, Foster City, CA). Following the generation of an estimated ~10-fold sequence redundancy (based on the measured insert size of each starting BAC), sequences were assembled and edited using the Phred/Phrap/Consed suite of programs¹⁵⁻¹⁷ (see www.phrap.org). Manual inspection of the assembled sequences allowed readily apparent errors, artifacts, and misassemblies to be corrected. Contigs were ordered and oriented based on read-pair associations of gap-spanning subclones and sequence overlaps between neighboring clones. For virtually all of the sequence reported here, finishing was then performed to the standards established for sequencing the human genome (see genome.wustl.edu/Overview/finrulesname.php?G16=1), which includes the absence of gaps and an error rate of less than 1 in 10,000 bp. Additional details and protocols are available on request.

A summary of the generated sequence data set is provided below (note that this is a more detailed version of Table 1 of the main paper):

Sequence generated from the greater *CFTR* region in 12 non-human vertebrates

Species	Non-redundant Total bp	No. BACs	Finished BACs*	Clone Gaps†	Sequence Gaps§
Chimpanzee (<i>Pan troglodytes</i>)	1,317,858	11	11	2	0
Baboon (<i>Papio cynocephalus anubis</i>)	1,508,413	11	10	0	2
Cat (<i>Felis catus</i>)	1,357,338	12	11	2	3
Dog (<i>Canis familiaris</i>)	1,195,669	8	6	4	10
Cow (<i>Bos taurus</i>)	1,480,745	11	11	0	0
Pig (<i>Sus scrofa</i>)	1,077,879	8	8	1	0
Rat (<i>Rattus norvegicus</i>)	1,600,751	15	14	0	3
Mouse (<i>Mus musculus</i>)	1,486,509	10	10	0	0
Chicken (<i>Gallus gallus</i>)	415,528	5	5	0	0
Fugu (<i>Fugu rubripes</i>)‡	273,624	4	4	0	0
Tetraodon (<i>Tetraodon nigroviridis</i>)	257,833	3	3	0	0
Zebrafish (<i>Danio rerio</i>)¶	162,514	3	3	1	0

* Number of BACs sequenced to the standards for finishing human genomic sequence. For the remaining few BACs, the order and orientation of all sequence contigs were established.

† Number of gaps between non-overlapping BACs within the clone tiling path.

§ Number of sequence gaps within the sequenced BACs.

|| Includes ~150 kb of sequence that we generated previously (GenBank No. AF162137) and reported by Ellsworth *et al.*¹⁸.

‡ Includes ~36 kb of sequence generated previously (GenBank No. AJ009961.1) and reported by Cottage *et al.*¹⁹

¶ Includes some non-orthologous sequence.

Below are listed each of the sequenced BACs, sorted by species:

SPECIES	BAC NAME	GENBANK NO.
baboon	RP41-479B1	AC084730.2
baboon	RP41-12N1	AC084729.2
baboon	RP41-150K7	AC087214.4
baboon	RP41-455I14	AC087215.3
baboon	RP41-265N24	AC087252.3
baboon	RP41-416C2	AC087598.3
baboon	RP41-63I4	AC087555.3
baboon	RP41-384C5	AC089990.3
baboon	RP41-468N20	AC091381.3
baboon	RP41-187H19	AC087806.3
baboon	RP41-353B7	AC087103.3
cat	RP86-177C24	AC087421.2
cat	RP86-294B21	AC087807.2
cat	RP86-49M22	AC087861.2
cat	RP86-459O8	AC087731.2
cat	RP86-523H23	AC090033.2
cat	RP86-439N5	AC087514.2
cat	RP86-181L4	AC089994.2
cat	RP86-238N12	AC117940.3
cat	RP86-87J17	AC091436.3
cat	RP86-264L24	AC091382.3
cat	RP86-317G18	AC091094.2
cat	RP86-565O15	AC117941.3
chicken	XXbac-68C5	AC091708.2
chicken	WAG-65N20	AC084760.2
chicken	WAG-105M15	AC091725.2
chicken	WAG-55C14	AC087179.2
chicken	WAG-69H2	AC084761.2
chimp	RP43-144M20	AC087512.2
chimp	RP43-22H1	AC087265.2
chimp	RP43-150N20	AC087253.2
chimp	RP43-150N19	AC087264.3
chimp	RP43-143F17	AC087729.2
chimp	RP43-126A24	AC087777.2
chimp	RP43-176G10	AC091637.1
chimp	RP43-89H17	AC087730.2
chimp	RP43-124I17	AC087834.2
chimp	RP43-75L6	AC090112.2
chimp	RP43-93P21	AC087835.2
cow	RP42-400M23	AC090976.2
cow	RP42-552E5	AC090961.2
cow	RP42-550C12	AC087843.2
cow	RP42-135P2	AC087860.2
cow	RP42-194K3	AC089991.2
cow	RP42-341K3	AC090031.2

Thomas *et al.* Supplementary Information #1 - 5

cow	RP42-332L24	AC122112.2
cow	RP42-99M11	AC091728.2
cow	RP42-210F11	AC089992.2
cow	RP42-553M7	AC089993.2
cow	RP42-324C24	AC091252.3
dog	RP81-391L22	AC090972.5
dog	RP81-55A4	AC090445.2
dog	RP81-386E20	AC090032.2
dog	RP81-196D8	AC090890.3
dog	RP81-142A6	AC091119.2
dog	RP81-256F7	AC092249.3
dog	RP81-156O18	AC090889.3
dog	RP81-182F3	AC117937.3
fugu	242D16	AC090119.1
fugu	NA	AJ009961.1
fugu	250A8	AC091727.2
fugu	244C11	AC087333.2
mouse	RP23-11G22	AC023173.3
mouse	RP23-208N9	AC024950.3
mouse	RP23-73G15	AC092495.2
mouse	RP23-40I5	AC027285.4
mouse	RP23-86H7	AC068561.2
mouse	NA	AF162137.2
mouse	RP23-219K14	AC027654.3
mouse	RP23-230F22	AC090712.4
pig	RP44-56M21	AC087424.2
pig	RP44-55G2	AC087423.2
pig	RP44-254M15	AC104718.1
pig	RP44-338P10	AC087180.2
pig	RP44-418J7	AC087160.2
pig	RP44-519O7	AC096884.2
pig	RP44-360A14	AC092478.2
pig	RP44-30C22	AC092497.2
rat	RP31-331N5	AC087102.3
rat	RP31-425H6	AC121515.2
rat	RP31-314L24	AC087042.2
rat	RP31-230B14	AC087101.2
rat	RP31-188L2	AC087041.2
rat	RP31-40H10	AC087067.3
rat	RP31-194D8	AC087066.3
rat	RP31-162L19	AC087112.2
rat	RP31-422M21	AC087213.2
rat	RP31-210K23	AC090444.2
rat	RP31-536L14	AC091268.2
rat	RP31-464J4	AC087775.2
rat	RP31-186N2	AC087251.2
rat	RP31-52E16	AC087776.2
rat	RP31-249D7	AC087218.2

tetraodon	GSTNA-28E1	AC113580.3
tetraodon	GSTNA-45F12	AC113581.3
tetraodon	GSTNA-27O10	AC113579.3
zebrafish	66I19	AC087105.2
zebrafish	91J1	AC087254.2
zebrafish	20A7	AC087104.2

References

1. Li, R. *et al.* Construction and characterization of an eightfold redundant dog genomic bacterial artificial chromosome library. *Genomics* **58**, 9-17 (1999).
2. Warren, W. *et al.* Construction and characterization of a new bovine bacterial artificial chromosome library with 10 genome-equivalent coverage. *Mamm. Genome* **11**, 662-663 (2000).
3. Fahrenkrug, S.C. *et al.* A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration. *Mamm. Genome* **12**, 472-474 (2001).
4. Woon, P.Y. *et al.* Construction and characterization of a 10-fold genome equivalent rat P1-derived artificial chromosome library. *Genomics* **50**, 306-316 (1998).
5. Osoegawa, K. *et al.* Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10**, 116-128 (2000).
6. Crooijmans, R.P., Vrebalov, J., Dijkhof, R.J., van der Poel, J.J. & Groenen, M.A. Two-dimensional screening of the Wageningen chicken BAC library. *Mamm. Genome* **11**, 360-363 (2000).
7. Lee, M.-K. *et al.* Construction and characterization of three BAC libraries for analysis of the chicken genome. *Animal Genetics* **34**, 151-152 (2003).
8. Roest Crollius, H. *et al.* Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome Res.* **10**, 939-949 (2000).
9. Thomas, J.W. *et al.* Comparative genome mapping in the sequence-based era: early experience with human chromosome 7. *Genome Res.* **10**, 624-633 (2000).
10. Summers, T.J. *et al.* Comparative physical mapping of targeted regions of the rat genome. *Mamm. Genome* **12**, 508-512 (2001).
11. Thomas, J.W. *et al.* Parallel construction of orthologous sequence-ready clone contig maps in multiple species. *Genome Res.* **12**, 1277-1285 (2002).

12. Elgar, G. *et al.* Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning. *Genome Res.* **9**, 960-971 (1999).
13. Wilson, R.K. & Mardis, E.R. Genome analysis: a laboratory manual. Vol. 1 Analyzing DNA. Birren, B., Green, E.D., Klapholz, S., Myers, R.M. & Roskams, J. (eds.), pp. 397-454 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1997).
14. Green, E.D. Strategies for the systematic sequencing of complex genomes. *Nat. Rev. Genet.* **2**, 573-583 (2001).
15. Ewing, B., Hillier, L., Wendl, M.C. & Green, P. Base-calling of automated sequencer traces using *Phred*. I. accuracy assessment. *Genome Res.* **8**, 175-185 (1998).
16. Ewing, B. & Green, P. Base-calling of automated sequencer traces using *Phred*. II. error probabilities. *Genome Res.* **8**, 186-194 (1998).
17. Gordon, D., Abajian, C. & Green, P. *Consed*: A graphical tool for sequence finishing. *Genome Res.* **8**, 195-202 (1998).
18. Ellsworth, R.E. *et al.* Comparative genomic sequence analysis of the human and mouse cystic fibrosis transmembrane conductance regulator genes. *Proc. Natl. Acad. Sci.* **97**, 1172-1177 (2000).
19. Cottage, A. *et al.* Three receptor genes for plasminogen related growth factors in the genome of the puffer fish *Fugu rubripes*. *FEBS Lett.* **443**, 370-374 (1999).