Journal of Biochemistry and Molecular Biology, Vol. 39, No. 4, July 2006, pp. 464-467

JOURNAL OF
Biochemistry and Molecular Biology

Short communication

One Step Cloning of Defined DNA Fragments from Large Genomic Clones

Christian Scholz, Gabriele Döderlein and Horst H. Simon*
Interdisciplinary Center for Neuroscience
Department of Neuroanatomy, University of Heidelberg, Im Neuenheimer Feld 307, 69120 Heidelberg, Germany

Received 8 December 2005, Accepted 13 March 2006

Recently, the nucleotide sequences of entire genomes became available. This information combined with older sequencing data discloses the exact chromosomal location of millions of nucleotide markers stored in the databases at NCBI, EMBO or DDBJ. Despite having resolved the intron/exon structures of all described genes within these genomes with a stroke of a pen, the sequencing data opens up other interesting possibilities. For example, the genomic mapping of the end sequences of the human, murine and rat BAC libraries generated at The Institute for Genomic Research (TIGR), reveals now the entire encompassed sequence of the inserts for more than a million of these clones. Since these clones are individually stored, they are now an invaluable source for experiments which depend on genomic DNA. Isolation of smaller fragments from such clones with standard methods is a time consuming process. We describe here a reliable one-step cloning technique to obtain a DNA fragment with a defined size and sequence from larger genomic clones in less than 48 hours using a standard vector with a multiple cloning site, and common restriction enzymes and equipment. The only prerequisites are the sequences of ends of the insert and of the underlying genome.

Keywords: BAC end sequences, Bacterial artificial chromosome, Gene targeting, Genomic BLAST

Introduction

Before the sequencing data of the human and murine genomes became publicly available (Venter *et al.,* 2001; Gregory *et al.,* 2002), large scale BAC (Bacterial artificial chromosome) end-sequencing projects were carried out at The Institute for Genomic Research (TIGR, http://www.tigr.org) in order to obtain random markers on the human, murine and rat genomes (Zhao, 2000; Zhao, 2001; Zhao *et al.,* 2001). As a result, at least one human (RPCI-11), two murine (RPCI-23 and -24), and one rat (CHORI-230) BAC library, each covering the respective genomes more than 10 fold (Osoegawa *et al*., 2000), are stored as individual clones and can be purchased for a nominal fee from the Children's Hospital Oakland Research Institute (CHORI) (http://bacpac. chori.org/). Each of these BACs has a unique identifier that is kept as an annotation with the two corresponding end sequences of the insert. These end sequences were submitted to the nucleotide databases at Genbank, EMBO and DDBJ and are accessible by text or BLAST searches (Altschul *et al.,* 1990). When the genomic sequences of mouse and man were published, due to the average length of 400 bases of these end sequences, it became possible to determine the exact genomic location for most of them. Furthermore, if both ends are mapped, as a consequence, the entire sequence of the insert is revealed.

BAC libraries have been extensively used to isolate genomic DNA for promoter/enhancer analyses or for the generation of homologous recombinant mutant mice. The isolation of smaller genomic fragments from larger genomic clones like BACs or others is a time consuming process and requires a significant amount of resources. It normally involves screening of libraries either by PCR or hybridization (Campbell and Choy, 2002), followed by a fragmentation process and rescreening until a plasmid is isolated with the targeted DNA.

We demonstrate here that combining genome and BES (BAC end sequence) data substantially reduces the time necessary to isolate defined genomic fragments from BAC clones of species with a fully sequenced genome. Our method reduces the process to a single cloning step requiring only common reagents and equipment. Using standard restriction enzymes, we cloned DNA fragments of 1 to 10 kb in size from a given BAC in less than 48 hours.

*To whom correspondence should be addressed.
Tel: 49 6221 548342; Fax: 49 6221 545605
E-mail: horst.simon@urz.uni-heidelberg.de

Materials and Methods

Databases. Mouse Genome Blast at http://www.ncbi.nlm.nih.gov/genome/seq/MmBlast.html, default settings - Database = genome, Program = blastn, use MegaBLAST, Expect = 0.01, Filter = default, Descriptions = 100, Alignments = 100. In some cases, the MegaBLAST option was removed and the Filter option set to "none" to determine the location of BAC end sequences. The accession numbers for the mouse genes used for the database searches: *Deleted in Colorectal Carcinoma* (DCC) NM_007831, *Dopamine Receptor 2* (Drd2) NM_010077, *Engrailed-1* (En1) NM_010133, *Foxa1* NM_008259, Nerve Growth Factor Receptor 1 (Ngfr or p75) NM_033217 and *Pbx1a* NM_183355.

Programs. Gene Construction Kit 2.5 from SciQuest, Inc. Research Triangle Park, NC 27709-2156, USA for generation of the restriction maps and the virtual assembly of vectors and inserts before designing primers. Oligo 6.8 from Molecular Biology Insights, Inc, Cascade, CO 80809, USA, to search for appropriate primers. The search for primer pairs was always performed at highest stringency. All of the chosen primer pairs had a suggested annealing temperature of 55°C or higher.

PCR conditions and restriction digests. The BAC clones were all purchased for a nominal fee from the Children's Hospital Oakland Research Institute (http://bacpac.chori.org). All used restriction enzymes are commercially available. The ligations were performed with the TaKaRa Ligation Kit Version 2 (BioWhittaker, Germany) at 16°C for at least 30 min. The PCR reactions were all done with standard solutions (PCR buffer with NH_4SO_4 (Fermentas); 12.5 mM dNTPs (Roche Diagnostics); 1.5 mM $MgCl_2$ (Fermentas)); PCR settings were one cycle of 5 min at 94°C, followed by 20-25 cycles

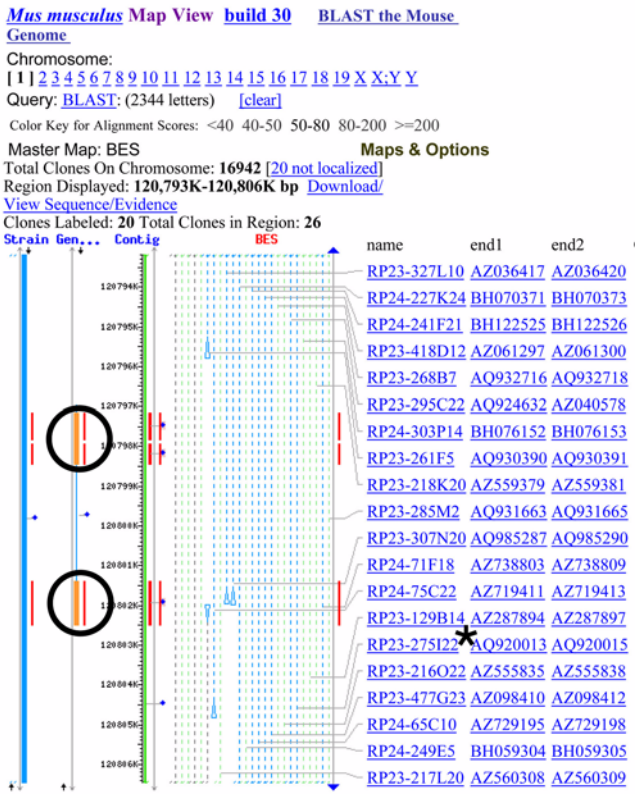

Fig. 1. Graphical Output of BLAST Result on Mouse Genome at http://www.ncbi.nlm.nih.gov/genome/seq/MmBlast.html. The mouse engrailed-1 gene was used for a BLAST search on the mouse genome. Graphical output reveals exon/intron structure of the gene on one contig and the chromosome location. Superimposed on MGSCv3 Map is the library of all mouse BAC end sequences (BES), showing several BACs which cover the entire gene. Database: MGSCV3, Program: blastn, use MegaBLAST, Expect: 0.01 Filter: default, Descriptions: 100, Alignment: 100. The asterisk delineates the BAC clone used for experiment in Tab.1, the circles mark the exons of the engrailed-1 gene.

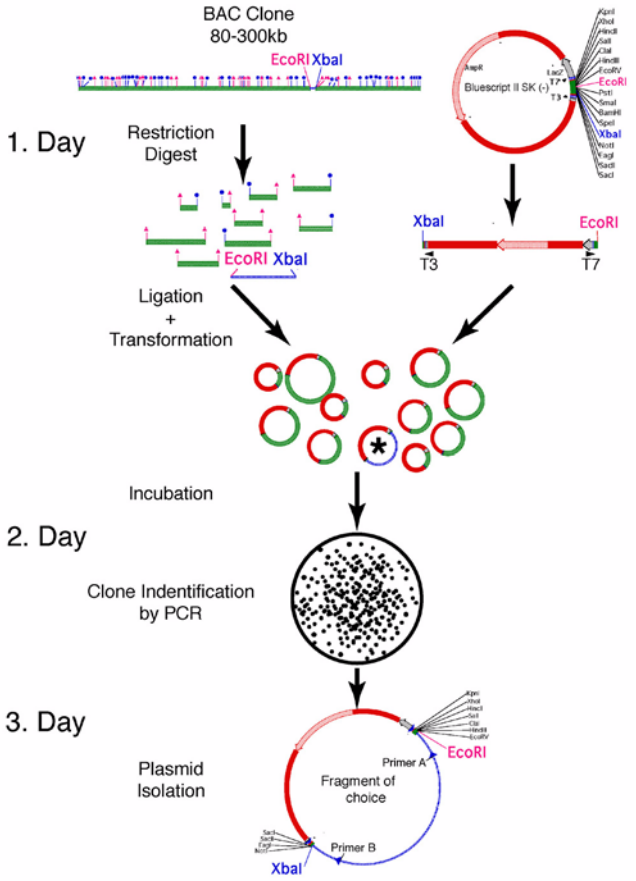

Fig. 2. Outline of the strategy to obtain a defined genomic fragment from a BAC clone. The outline is based on the isolation of a genomic fragment containing exon 2, 3 and 4 of the dopamine receptor 2 from the BAC clone RP23-416F20. Day One: The BAC clone is cut with restriction enzymes flanking the fragment of interest (XbaI and EcoRI). All resulting fragments are subcloned into a vector (Bluescript II SK(-)) which was cut with the same enzymes. After ligation and transformation, the bacteria are plated out and grown overnight. Second Day: The individual clones are picked and subjected to PCR with one primer specific for the targeted DNA fragment, the other for the vector (T7 + Primer A and T3 + Primer B). The positive clones are grown overnight and the plasmid DNA is isolated at the third day.

with 30' at 94°C, 60' at 55°C or higher (determined by primer set used), 45' at 72°C and a final step of 7 min at 72°C.

Results and Discussion

To clone a distinct genomic fragment of defined size and sequence, BAC clones containing the targeted DNA were first identified. For this purpose, we performed a BLAST search on the entire genome using cDNA sequences. The result is in most cases, as expected, one supercontig containing the entire cDNA on one chromosome (Fig. 1). The "Map View" option reveals exon/intron structure and BAC clones stored at TIGR containing the genomic region corresponding to the employed cDNA. To isolate smaller fragments from the BAC clones, we performed another genome wide BLAST, this time using both BESs as query templates. This revealed their positions in the genome and made it possible to retrieve the sequence of the entire insert. With this sequence, we generated a restriction map, which we applied to design a one-step cloning strategy to isolate fragments of 1 to 10kb straight from the BAC DNA. These restriction maps were generated with a commercial software (Gene Construction Kit 2.5 from SciQuest, Inc.) using a limited number of restriction enzymes producing all non-degenerate overhangs compatible with the multiple cloning site of a standard vector. In our case, we used Bluescript II SK (-). If a very large DNA fragment was required, we sometimes made use of a modified Bluescript vector, which contained several 8-base restriction sites in the multiple cloning site and increased the pool of enzymes accordingly. As the next step, we chose one or two restriction enzymes, which had sites flanking but not occurring within the DNA fragment of interest (for examples see Table 1). Then we digested approximately 10 µg of the BAC DNA with the chosen enzyme(s) and subcloned the entire pool of fragments. On the next day, the clones that contained the targeted DNA were identified by PCR colony picking. For this purpose, we constructed a virtual Bluescript vector containing the presumed sequence of the targeted genomic fragment. With this sequence information, we identified a primer (18-22 bases) on the genomic DNA fragment compatible to the T7 or T3 site of the vector using a commercial primer design program (Oligo 6.8, Molecular Biology Insights, Inc). We picked the bacterial clones, placed them first into the PCR reaction mix and then into bacterial medium using 96 well plates. After the PCR reaction, the positive clones were placed in 1-2ml of medium and incubated over night. Next day, the plasmids were purified and analyzed by restriction digests (see Fig. 2 for overview). Following this procedure, it required the picking of 25 to 50 individual clones to identify a

Table 1. Examples of isolated mouse genomic DNA fragments

Gene	BAC Clone ID (TIGR)	Enzymes	#	Positions	Size kb	Primer	pBs
Drd2	RP23-416F20	EcoRI	38	119219 to 124483	5.2	TGGGAAAGGGCTACAGCAT	T7
		EcoRI/XbaI	52	115249 to 118123	2.9	GAGGATCATGGGAGGGGAC	T7
		XbaI/AvrII	52	118123 to 119796	1.7	GCCTGAAACTCCTGTTGCT	T3
Foxa1	RP23-91E17	EagI/AscI	3	177512 to 180868	3.3	ATTGATCTTGGGGAAAGGTT	T3
		ClaI/XhoI	4	180678 to 183368	2.7	CACACCCGAGGGCTCA	T3
		HindIII	67	173869 to 181355	7.5	CGCCTCAGTCCACTCCA	T3
		BamHI	36	185555 to 188790	3.2	GGATCTAGCCTGCCGAATC	T3
		EcoRI/EagI	58	107638 to 112173	4.5	CCCGTGTTGGCGTAGGACA	T7
Pbx1	RP24-293G19	XhoI/EcoRI	16	372141 to 380000	7.9	TTTAAATTTGGGTAAGATTC	T7
Dcc	RP23-4L9	BamHI	56	80066 to 87352	7.2	AATACCTGGGCTTGGAAC	T7
		BamHI/NheI	81	80066 to 87352	7.2	AATACCTGGGCTTGGAAC	T7
		XbaI	58	56496 to 62426	5.7	TGGTTAAAGAAGGTGGCT	T7
p75	RP24-347P1	XhoI	10	69790-73585	3.6	GCTAAGTGGGCTGCTCAG	T7
		SmaI	19	18896-24615	5.7	GTTATGGTTGGCTAGGGA	T7
En1	RP23-2751222	EcoRI/XhoI	18	6468 to 10215	3.9	TGGGCAAGGGAATCAGC	T7
		XhoI/XbaI	23	10215 to 14529	4.3	GCGCATCCTCCAAGAGACT	T7
		XbaI/NarI	15	14529 to 19904	5.4	GCTGACCTCTTACGCTTCTT	T7

The table contains the gene names, the identifier for each TIGR BAC clone, the restrictions enzymes which were used for the digests, the number of fragments produced by each digestion (#), the exact numeric position of the isolated fragment in the BAC clone, their sizes in kilobases the primers used for the PCR reactions together with the T7 or T3 primers (sequence below table). En1 = engrailed-1 Foxa1 = forkhead box A1 transcription factor DDC = deleted in colorectal carcinoma DRD2 = dopamine receptor 2 Pbx1 = pre-B-cell leukemia transcription factor 1 pBs = Bluescript II SK (-).

positive plasmid, no matter if one or two enzymes were used for the digestion of the BAC clones. The number of clones we had to analyze by PCR amplification to obtain the targeted DNA was lower than expected, if one assumes that all fragments generated by the restriction digests are equally likely to be inserted into the vector (see Table 1). We think there may be a bias for insertion of fragments of 1kb and larger, however, we did not test this further. In 2 out of 24 experiments, even extensive screening of more than 300 colonies yielded no result, however a repetition employing different restriction enzymes was in both cases successful. It is likely that the sequence information was inaccurate, but other reason cannot be excluded.

The method we present here is quickly performed and does not depend on any reagents and equipment that are not standard and readily commercially available. This method should be applicable to similar sources of subcloned genomic DNA like PACs (Sternberg, 1992), YACs (Burke *et al.,* 1987) and others, if sequences of the ends as well as of the underlying genome are known as in the cases of the here described BACs. The used Bluescript vector can be substituted for any vector with a comparable number of specific restriction sites within its multiple cloning site. This method works with very high fidelity compared to another widely used alternative procedure, long-range PCR, which is notoriously unreliable, even on low-complexity templates, e.g. BAC clones. The strategy presented here may become the method of choice to isolate defined genomic DNA out of larger genomic clones.

Acknowledgments This work was supported by a grant from the Federal Secretary for Education and Research. We thank Bill Wisden, Lavinia Bhatt and Paola Sgadó for discussion and help preparing the manuscript. This work has been carried out in compliance with the current laws of the Federal Republic of Germany governing genetic experimentation.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Burke, D. T., Carle, G. F. and Olson, M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-812.

Campbell, T. N. and Choy, F. Y. (2002) Approaches to library screening. *J. Mol. Microbiol. Biotechnol.* **4**, 551-554.

Gregory, S. G., Sekhon, M., Schein, J., Zhao, S., Osoegawa, K., Scott, C. E., Evans, R. S., Burridge, P. W., Cox, T. V., Fox, C. A. et al. (2002) A physical map of the mouse genome. *Nature* **418**, 743-750.

Osoegawa, K., Tateno, M., Woon, P. Y., Frengen, E., Mammoser, A. G., Catanese, J. J., Hayashizaki, Y. and de Jong, P. J. (2000) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10**, 116-128.

Sternberg, N. L. (1992) Cloning high molecular weight DNA fragments by the bacteriophage P1 system. *Trends Genet.* **8**, 11-16.

Venter, J. C. Adams, M. D. Myers, E. W. Li, P. W. Mural, R. J. Sutton, G. G. Smith, H. O. Yandell, M. Evans, C. A. Holt, R. A. et al. (2001) The sequence of the human genome. *Science* **291**, 1304-1351.

Zhao, S. (2000) Human BAC ends. *Nucleic Acids Res.* **28**, 129-132.

Zhao, S. (2001) A comprehensive BAC resource. *Nucleic Acids Res.* **29**, 141-143.

Zhao, S., Shatsman, S., Ayodeji, B., Geer, K., Tsegaye, G., Krol, M., Gebregeorgis, E., Shvartsbeyn, A., Russell, D., Overton, L., Jiang, L, Dimitrov, G. Tran, K., Shetty, J. Malek, J. A. Feldblyum, T. Nierman, W. C. and Fraser, C. M. (2001) Mouse BAC ends quality assessment and sequence analyses. *Genome Res.* **11**, 1736-1745.