Technique for Cloning and Sequencing the Ends of Bacterial Artificial Chromosome Inserts

BioTechniques 29:271-276 (August 2000)

ABSTRACT

Bacterial artificial chromosome (BAC) libraries are an important tool for positional cloning, gene analysis and physical mapping. During studies using BAC clones, it is often necessary to organize them into contiguous sequences (contigs). To finalize, join and extend the contigs, both cloning and sequencing of the ends of the inserts are required. Here, we describe a low-cost, accessible, fast and powerful method for the routine isolation of BAC ends. This method allows the isolation of 20 BAC clone ends in

one day. The analysis of the ends reveals fragment sizes compatible with sequencing, and the structure of these clones allows the sequencing of both ends using the same plasmid. Moreover, long end fragments can be sequenced in both directions.

INTRODUCTION

The use of high molecular weight DNA insert libraries (YAC or BAC) is a central part of gene isolation by positional cloning, genome analysis or physical mapping (6,9). After the clones containing the regions of interest have been isolated, a contig is assembled by fingerprinting. Once the contig is generated, the second step is to isolate the ends of the inserts from each clone to confirm the contig and find suitable probes to extend it if necessary. Thus, the isolation of insert ends is an important part of strategies that use high mol-

ecular weight DNA libraries.

Four main techniques are used to carry out the isolation of BAC ends: (i) direct sequencing of the ends of the BAC insert (3,8)—these methods provide a fast way of obtaining sequence but require a reasonable amount of DNA at a concentration suitable for sequencing; (ii) PCR to amplify the ends (8)—these methods need only a few nanograms of DNA but are limited by the quality of the PCR and the number of fragments actually generated; (iii) end rescue (8), which can isolate the ends from a few nanograms of DNA but is limited to the isolation of the end next to the chloramphenicol resistance gene or the isolation of both ends joined together, without knowing the site of union. In this technique, the ends are cloned into a single-copy vector, resulting in low DNA yields. The fourth method (iv) is the directional cloning of the ends (1), which produces ends in a

Short Technical Reports

multiple-copy vector but depends on complex digestions and screening to find the right clones.

For these reasons, a new technique has been developed to clone and subsequently sequence both BAC ends simultaneously from a single clone. This technique based on plasmid rescue is designed to overcome the problems encountered with end rescue and allows the generation of longer sequences than direct sequencing of the BAC ends. This can also be achieved with no need to design internal primers.

The principle here is to digest the BAC DNA with enzymes that cut between the two ends of the insert but do not cut the BAC vector itself (Figure 1). A selectable linker DNA containing a high copy number bacterial origin of replication (*ori*) is then inserted between the two BAC ends. We describe here the isolation of insert ends from randomly picked maize BAC clones with an average insert size of 150 kb in the vector, pBeloBAC11 (5). This technique proved to be a quick, easy and low-cost way to

clone BAC ends in a form that can easily be purified and sequenced.

MATERIALS AND METHODS

pTRAPends Construction

The selectable linker DNA (pTRAPends) containing a high copy number bacterial *ori* was constructed from pUC18 (10). The vector was digested using the enzymes *Aat*II and *Bsp*II to

delete all the sequences located around the multiple cloning sites (MCS). When these sequences were removed, the deleted vector (1811 bp) was self-ligated with the two complementary oligonucleotides, 5'-OHGCTAGCTCTAGACCTAGG-3' and 5'-OHCTAGCCTAGGTCTAGAGCTAGCACGT-3'. This created a new polylinker containing an *Xba*I site between *Avr*II and *Nhe*I sites and generated the vector pTRAPends (Figure 1). The three sites in this MCS have compatible cohesive ends.

Figure 1. Flow diagram of the technique. BAC DNA is first digested with restriction endonucleases *Nhe*I and *Avr*II. Then these fragments are ligated with the linearized vector pTRAPends, and the product of the ligation is transformed into bacteria. The pBends vector is then selected because of antibiotic resistance.

Short Technical Reports

Ends Cloning

BAC DNA extractions were carried out by the alkaline lysis miniprep method from 5-mL cultures for BAC clones and 1.5-mL cultures for plasmids. The DNA were resuspended in 40 μ L (BAC) or 50 μ L (plasmid) 10 mM Tris, pH 8.0, and 0.5 mM EDTA.

Approximately 200 ng (25 μ L) BAC DNA were digested for 3 h at 37°C with 5 U *AvrII* and 5 U *NheI* (New England Biolabs, Beverly, MA, USA) in a final volume of 50 μ L (1 $\times$ NEbuffer 2) containing 100 μ g/mL BSA. The digested DNA was extracted with phenol and chloroform and 25 μ L (50 ng) of the extracted digestion separated on a 0.9% agarose gel.

The pTRAPends vector was digested with an excess of *XbaI* for 3 h at 37°C and gel-purified after migration on a 1.2% agarose gel.

To clone the BAC ends, 20 ng digested BAC DNA were ligated in T4 DNA ligase buffer (Roche Molecular Biochemicals, Indianapolis, IN, USA) for 3 h at room temperature with 10 ng pTRAPends in a final volume of 10 μ L, with 1 Weiss unit of T4 DNA ligase.

Preparation of Electrocompetent Bacteria and Transformation

DH10BTM bacteria (Life Technologies, Rockville, MD, USA) were used to prepare electrocompetent cells (12).

To transform the bacteria, ligation mixtures were drop dialyzed on an MF 0.025 membrane (Millipore, Bedford,

MA, USA) against 10 mM Tris, pH 8.0, and 0.5 mM EDTA for 1 h. Between 1 and 5 μ L dialyzed ligation mixture were added to 27 μ L bacteria and electroporated in a 1-mm cuvette at 1.5 kV, 25 μ F and 200 Ω using a gene pulser electroporator (Bio-Rad Laboratories, Hercules, CA, USA).

To select the clones containing the BAC ends, the transformation mixtures were spread onto LB agar plates containing 100 μ g/mL ampicillin and 12.5 μ g/mL chloramphenicol and incubated at 37°C for 20 h.

Sequencing of the Ends

Sequencing reactions were carried out using the BigDyeTM automatic fluorescent sequencing kit (PE Biosystems, Foster City, CA, USA). Sequencing reactions were carried out in a final volume of 10 μ L with 500 ng DNA, 4 μ L BigDye mixture and 1.6 pmol primer. The sequencing reactions were carried out using 25 cycles: 96°C for 30 s, 50°C for 15 s and 60°C for 3 min on a model 9600 GeneAmp[®] PCR system (PE Biosystems). The sequence reactions were analyzed using an ABI PRISMTM 377 DNA sequencer according to the manufacturer's recommendations (PE Biosystems).

RESULTS

To validate this new technique, 10 BAC clones with an average insert size of 150 kb were randomly picked from a

maize BAC library (constructed in pBeloBAC11) (7). The restriction endonuclease enzymes *AvrII* and *NheI* were chosen to digest the mini-prepped BAC DNA. These two enzymes were chosen because they both generate compatible cohesive ends with themselves and with the enzyme *XbaI* and because these enzyme sites are present in the vector pTRAPends and absent in pBeloBAC11. A vector carrying both ampicillin and chloramphenicol resistance is only created when the BAC ends, and therefore pBeloBAC11, are cloned. The other *AvrII* and *NheI* fragments will also be cloned but are selected against because they do not carry chloramphenicol resistance. Moreover, as the vector pTRAPends has an MCS containing an *XbaI* site between the *AvrII* and *NheI* sites, one BAC end can be excised with a *NotI/AvrII* digestion and the other by a *NotI/NheI* digestion.

Twelve clones that resulted from the ligation of each cut BAC DNA with the vector pTRAPends were analyzed. The DNA was first digested with *NotI* to verify the presence of pBeloBAC11. Clones consisting of pBeloBAC11 and a second fragment (representing the BAC ends plus the plasmid pTRAPends) were analyzed to determine the size of the inserts. For this process, four digestions were carried out (Figure 2): (i) digestion with *AvrII/NheI* to verify if the ends result from a total digestion and to verify the presence of pTRAPends; (ii) digestion with *NotI/AvrII* to estimate the size of the *AvrII* fragment; (iii) digestion with *NotI/NheI* to estimate the size of the *NheI* fragment; and (iv) a digestion with *NotI/AvrII/NheI* to visualize the four fragments, pBeloBAC11 (7.5 kb), pTRAPends (1.8 kb), the *AvrII* fragment and the *NheI* fragment.

The ends of all 10 BAC clones were successfully isolated and their sizes estimated on an agarose gel. The average sizes were 1.87 kb, and a statistical analysis (4) predicted that 99.9% ($P = 0.001$) of the ends should have a size between 2.6 and 0.8 kb. The average size found is close to the theoretical size expected for a double digestion using restriction enzymes having a 6-bp recognition site ($4^{6/2} = 2048$ bp). The difference between the theory and the observed size can be explained by the fact that the ends are the result of a dou-

Figure 2. Size determination of ends for 6 of 12 minipreps from one cloning experiment. Legend: 1 kb: 1-kb ladder molecular weight marker (Promega, Madison, WI, USA); TEn: pTRAPends digested by *NotI* (non-digested); BeIn pBeloBAC11 and pBluescript[®] (pBS) linearized by *NotI*; TEx: pTRAPends (pTE) linearized by *XbaI*; NA, NAN and NN are the *NotI* + *AvrII*, *NotI* + *AvrII* + *NheI* and *NotI* + *NheI* digestions, respectively, for each of the six minipreps.

Short Technical Reports

ble digestion with *NheI* and *AvrII* and a partial digestion using *HindIII*.

Each clone could be sequenced from four different sites (Figure 1). The *NheI* and the *AvrII* side of pTRAPends used the primers pTN (5'-CTCATGAGCG-GATACATATTTGA-3') and pTA (5'-CTATGGAAAAACGCCAGCAAC-3'). In addition, the universal reverse primer for the chloramphenicol side of pBeloBAC11 and the M13 primer for the other side of pBeloBAC11 can also be used to generate sequences. Three clones were sequenced, generating the entire sequence for four ends and giving the first 600 bases in either direction from the other two ends.

DISCUSSION

We have shown that the isolation and sequencing of the ends of BAC clones can be carried out in a routine manner. Indeed, a common way to use BAC clones in the laboratory is to prepare DNA when the clones are generat-

ed to verify the insert size. The technique described earlier allows the use of part of the remaining DNA to quickly (one day) generate clones carrying both BAC ends. Once these clones are generated, they can be stocked or analyzed the following day. Thus, a clone containing both ends of each BAC clone can be isolated within two days, giving this technique the speed required for routine use. The technique also results in the generation of BAC ends of a suitable size for sequencing (statistical analysis). Moreover, the characteristics of these clones are such that the plasmid is present as a high copy number plasmid, and both ends can be sequenced in either direction with standard primers to generate four sequences. These four sequences can generate twice the amount of sequence information if they do not overlap or may generate the entire end sequences if they do. This could be important to better avoid repetitive sequences in making PCR-based probes.

In summary, this technique makes it possible to clone BAC extremities in a short time using small amounts of DNA. The clones generated contain the ends in a high copy number vector facilitating DNA isolation. Finally, the structure of the plasmids make it possible to obtain either the entire sequence of the ends or to double the sequence generated (from ends that are too big to be sequenced in only one run) without designing new primers.

Compared to other techniques that allow large-scale direct sequence analysis (2) or the isolation of separated ends (7), this new technique provides a fast and reliable method to isolate the ends of BAC clones. Moreover, this method enables the use of already made DNA and premade primer, which implies a lower cost to clone the ends and generate an amount of sequence double that of other techniques.

REFERENCES

1. Chen, C., S. Huang, P. Ling, C. Yu, Z. Deng and F.G.J. Gmitter. 1999. A novel method to clone BAC insert ends based on double digestion. Plant and animal genome VII conference. Poster 117.
2. Hanahan, D., J. Jessee and F.R. Bloom. 1991. Plasmid transformation of *Escherichia coli* and other Bacteria. Methods Enzymol.

204:63-113.

3. Kelley, J.K., C.E. Fields, M. Brook Craven, D. Bocskai, U.J. Kim, S.D. Rounsley and M.D. Adams. 1999. High throughput direct end sequencing of BAC clones. Nucleic Acids Res. 27:1539-1546.
4. Kempthorne, O. 1952. The design and analysis of experiments. John Wiley and Sons, New York.
5. Kim, U.J., B. Birenn, V. Mancino, T. Slepak, C. Boysen, H.L. Kang, M.I. Simon and H. Shizuya. 1996. Construction and characterization of human bacterial artificial chromosome library. Genomics 34:213-218.
6. Meyers, B.C., D.B. Chin, K.A. Shen, S. Sivaramakrishnan, D.O. Lavelle, Z. Zhang and R.W. Michelmore. 1998. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. The Plant Cell 10:1817-1832.
7. O'Sullivan, D., P.J. Ripoll and K.J. Edwards. 1999. A maize BAC library from the European flint line "F2". The international conference on the status of plant and animal genome research. Poster 96.
8. Woo, S.S., J. Jiang, B.S. Gill, A.H. Paterson and R.A. Wing. 1994. Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor. Nucleic Acids Res. 7:4922-4931.
9. Yang, D., A. Sanchez, G.S. Klush, Y. Zhu and N. Huang. 1998. Construction of a BAC contig containing the xa5 locus in rice. Theor. Appl. Genet. 97:1120-1124.
10. Yanisch-Perron, C., J. Vieira and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.

P.J.R. and D.O.S. were funded under the EU Framework IV Programme, MAP MAIZE (BIO4-CT97-2312). IACR-Long Ashton is funded by the Biotechnology and Biological Sciences Research Council of the United Kingdom. Address correspondence to Dr. Pierre-Jean Ripoll, IACR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol, BS41 9AF, UK. e-mail: pierre-jean.ripoll@bbsrc.ac.uk

Received 6 December 1999; accepted 3 April 2000.

Pierre-Jean Ripoll, Donal M. O'Sullivan, Keith J. Edwards and Matthew Rodgers¹
*University of Bristol
Long Ashton, Bristol
¹Rhône-Poulenc Agricultural
Limited,
Ongar Essex, UK*