Short Technical Reports

Long PCR Facilitates Concise Cloning and Sequencing with a Minimal Tiling Set of Templates

BioTechniques 23:1070-1075 (December 1997)

ABSTRACT

We demonstrate a rapid cloning and sequencing strategy for kilobase-size DNA segments using DNase I and long PCR. In a single-tube protocol, deletions were formed in a plasmid insert by two enzymatic cuts, one at a fixed site and one at random. The doubly cut molecules were recircularized to generate a library of plasmids carrying deletions of various sizes and transformed into *E. coli*. The plasmid inserts were directly amplified from transformant colonies by long PCR and sized on a high-resolution agarose gel. A minimal tiling set, selected from the amplified material, was used directly as templates for long-read sequencing. The system is useful for inserts up to about 3.5 kb for *de novo* sequencing (both strands) or 6 kb for confirmatory sequencing (one strand).

INTRODUCTION

Sequence analysis of small (2–10 kb) cloned DNA fragments, is usually accomplished by subcloning strategies that generate “nested” or overlapping deletions from a fixed point. All of these strategies create new sequencing start points by bringing different parts of the fragment adjacent to the sequencing primer site through various methods of deletions. Examples of these deletion methods include the use of the exonucleases ExoIII (6,11,13) or BAL31 (3,8,9), transposition of engineered $\gamma\delta$ in the pDUAL system (12) and endonucleolytic cleavage by DNase I (2,5,7). In the method we describe here, we chose the DNase I technique, which, in the presence of manganese, creates random single cuts in double-stranded plasmid molecules. The only prior knowledge required is the absence of at least two restriction sites in the cloned insert.

Now that polymerase chain reaction

(PCR) can be applied to targets of several kilobases (1), it is the tool of choice for streamlining the screening of subclones and the preparation of DNA templates for sequencing. Indeed, commercial products exist for this purpose (e.g., GeneSCOUT™ from Genosys and ScreenTest™ from Stratagene). In this study, we used the combined power of long PCR and DNase I protocols to provide a concise sequencing strategy for cloned fragments. Deletion subclones generated by DNase I were analyzed for size using long PCR, after which the PCR products were sequenced directly. One set of commercially available primers was used for both the PCR and the sequencing reaction. No other primers were required. The use of Long Ranger Singel™ packs for sequencing gels reduced the number of clones to be sequenced compared with a standard sequencing gel. Up to 700 bases of good sequence data were obtained from each long PCR product. These techniques vastly reduced the amount of time and effort required by more traditional protocols.

MATERIALS AND METHODS

Target Clone Construction

Genomic template DNA was prepared from *E. coli* cells embedded in InCert® Agarose (FMC BioProducts, Rockland, ME, USA). The DNA plugs were melted at 70°C and diluted 1:1 with water for use as PCR template. From this preparation, an 8.2-kb sequence in the *E. coli* genome was amplified by long PCR with the LA Taq II Kit from Takara (PanVera, Madison, WI, USA). The amplimer was tested for cleavage by restriction enzymes; no sites for *NotI* and *BamHI* were found. The amplimer was, therefore, cloned into *SmaI* site of pBluescript® II SK(-) (Stratagene, La Jolla, CA, USA) so that the *NotI* and *BamHI* sites of the plasmid were placed between the sequencing primer binding site and the inserted fragment (Figure 1). The ligated plasmid was transformed into SURE® 2 Supercompetent Cells (Stratagene), and plasmid clones containing both orientations of the insert were obtained.

DNase I Digestion

DNase I (Life Technologies, Gaithersburg, MD, USA) was stored at 1 mg/mL in 0.01 N HCl at -80°C. Ten micrograms plasmid DNA were brought to 50 μ L with 50 mM Tris, pH 7.6 and held at 37°C. DNase I was freshly diluted 1:10⁵ in 100 mM Tris-HCl, pH 7.6, 1 mM EDTA, 20 mM manganese chloride, 200 μ g/mL bovine serum albumin (BSA). Diluted enzyme (50 μ L) was added to the pre-warmed DNA, and 25- μ L samples were withdrawn after 2-, 3-, 4- and 5-min incubations at 37°C and added to 5 μ L 50 mM EDTA to stop the digestion. Samples (2 μ L) were analyzed on a 1% SeaKem® Gold Reliant® Gel (FMC BioProducts). Samples with maximum amounts of linear molecules and minimal degradation were pooled and extracted with phenol/chloroform, precipitated with ethanol, dried and resuspended in water. The DNA was digested with *NotI*, the ends were blunted and the DNA was ligated under dilute conditions as specified for recircularization in the DNA Ligation Kit from Takara (PanVera). Finally, the DNA was digested with *BamHI*, precipitated with ethanol and transformed into SURE 2 cells, selecting for ampicillin resistance. Restriction enzymes were obtained from New England Biolabs (Beverly, MA, USA).

Screening the Deletion Library for Insert Size

Colonies were picked and dispersed in water. At this point, they can also be archived (e.g., by streaking on fresh agar). The suspended bacteria were lysed at 96°C for 5 min to release the plasmid DNA to be used as the PCR template. Primers, -47 Forward and -48 Reverse (24-mers; New England Biolabs designs) were used to amplify the inserts. Long PCR was performed by the TaKaRa ExTaq System (PanVera), which generated high yields of product useful for sequencing as well as sizing. Portions (2 μ L) of the amplimers were analyzed on Reliant® 4% NuSieve® 3:1 Precast Gels (FMC BioProducts). Successfully amplified samples were then accurately sized on hand-cast, large-format gels (4) of 0.8% high-gel-strength agarose SeaKem Gold (FMC BioProducts).

Short Technical Reports

Sequencing

Long PCR products (yield about 5 µg) were purified on QIAquick™ Spin Columns (Qiagen, Chatsworth, CA, USA) to remove unused PCR primers. DNA sequencing was performed according to manufacturer's protocols with PRISM® Ready Reaction Dye Primer or Dye Terminator Reagents (PE Applied Biosystems, Foster City, CA, USA), and data were collected on a Model 377 Automated Sequencer (PE Applied Biosystems). The gel matrix for sequencing was Long Ranger Single packs (FMC BioProducts).

RESULTS AND DISCUSSION

Figure 1 shows the deletion procedure and plasmid structures. When plasmid DNA was digested by DNase I, two types of molecules were produced in approximately equal amounts, one being the linearized plasmid and the other being the nicked (relaxed) circular DNA. Further digestion decreased the amount of unwanted circular material, but also induced multiple cuts on the plasmid, resulting in progressive degradation. By monitoring the digestion on an agarose gel, time-point samples with maximum linear molecules and minimal degradation

were identified. Pooling two or three of the time-course samples yielded adequate amounts of linear DNA to complete the procedure.

Digestion with enzyme R1 (Figure 1), in this case *NotI*, created the fixed end of each deletion. Ligation under dilute conditions allowed linear molecules to recircularize, forming the library of deleted plasmids. Several hundred ampicillin-resistant transformants were obtained from 10 µg starting DNA. All of these steps, up to transformation, were carried out in a single tube. The speed and ease of this method is made evident in Figure 2 by comparison with a traditional deletion protocol.

In two experiments, 60 and 120 clones in each orientation were screened. Since 50% of the ampicillin-resistant clones failed to produce amplimers, it was useful to prescreen the PCR products on a pre-cast gel. Figure 3 shows the accurate sizing of a series of amplimers selected after prescreening. Sizes up to about 4 kb were readily found, larger ones rarely, although the full 8.2-kb insert amplified easily. Some of the largest amplimers also showed a smaller second band (lanes 9–13). Statistically, there was no difference among the number of clones that had insert sizes from 300 bp to about 3.5 kb (plasmid sizes from 3.2–6.4 kb).

Figure 1. Construction of the library of deletions. A DNA fragment was inserted into the vector between two primer-binding sites. Between the cloned insert and the sequencing (or reverse) primer site are unique restriction cleavage sites for enzymes R1 and R2. In the example discussed, R1 and R2 were *NotI* and *BamHI*, respectively. When DNase I introduces a cut within the inserted fragment, subsequent cleavage with R1 will eliminate a portion of the molecule that contains the R2 site and the adjacent piece of the insert. Therefore, when these plasmids are recircularized, various lengths of the insert will be deleted, and they will be resistant to R2 cleavage. On the other hand, molecules that escaped cleavage by DNase I or R1 will be cleaved by R2 and will be eliminated from the system through the transformation process. If DNase I cleavage occurs within the vector sequence, the transformants from these plasmids will either not be viable (due to loss of the ampicillin-resistance gene or the replication origin) or will not produce PCR products (due to loss of one of the primer binding sites).

Short Technical Reports

However, there was a sharp drop in frequency of clones that had larger inserts. The size of the insert may have limited both the circularization and the transformation efficiencies, since larger DNA molecules transform less efficiently than smaller ones, where molecules of different sizes are competing. The probability of circularization in a dilute ligation decreases as the distance between the ends (i.e., the length of the DNA) increases, but the probability-length relationship is nonlinear (10), which is consistent with our observations. While we expected to find some viable clones in which the initial DNase cut was in the vector part of the plasmid, examination of Bluescript's structure shows only 858 bases between the cloning site and the ampicillin-resistance promoter, where such a cut could result in loss of a priming site. These non-amplifying plasmids transform very efficiently due to their small size, resulting in their overrepresentation in the library.

From the accurate sizing gel, amplicons were selected to provide a minimal tiling set for sequencing, covering a span of about 4 kb. Only 0.5–1 μ L from a 50- μ L reaction was needed for this gel. The remaining DNA from samples in lanes 1–9, 12 and 14 were sequenced on a Long Ranger gel to obtain longest possible reads. Up to 700 bases of useful sequence were obtained from each amplicon. Seven of the amplicon sequences assembled into a single contig with the expected pattern of overlap. The amplicons in lanes 10, 11, 12 and 14, which also showed smaller bands on the sizing gel, did sequence cleanly but contributed only short novel sequences consistent with the size of the smaller fragments seen on the gel, with the rest of the sequence reading into the vector beyond the distal primer site. We conclude that the larger bands are artifacts arising during PCR screening, where the polymerase, optimized for kilobase amplicons, replicated around the plasmid.

We have demonstrated that a simple one-tube system can work very efficiently with Long PCR to streamline the production of overlapping clones and their sequencing templates. In this protocol, there is no size fractionation—only one ligation and transfor-

mation—and no requirement for specific restriction site ends. This is in contrast with the ExoIII deletion strategy in which one 3' overhang and one 5' overhang enzyme sites are required to create unidirectional deletions. Finding these sites is not always easy. Long PCR produces templates of sufficient length, quality and quantity for the longest readouts obtainable on the currently used sequencing instruments

(Models 373A and 377 [PE Applied Biosystems], Model 4000L [LI-COR] or ALFTM [Pharmacia Biotech]). With the high accuracy attainable by the newer sequencing chemistries, long reads enable the user to decrease the number of clones to be sequenced. For example, when read length is increased from 500 to 700 bases, the sequence of a 4-kb fragment can be determined from 6–7 subclones instead of 9–10

Figure 2. Comparison of protocols. The presented protocol is compared with the traditional ExoIII deletion strategy. There is at least three days difference in turn-around time between the two strategies.

Figure 3. Accurate size estimation of amplicons. To determine the size of inserts, PCR samples were obtained from ampicillin-resistant colonies using forward and reverse primers and analyzed on a 0.8% high-gel-strength agarose in 1 $\times$ TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0). Gels were electrophoresed for 1 h, 45 min at 5 V/cm and stained with ethidium bromide. Lanes 1–14 were PCR samples from individual clones (1 μ L each); lanes M1 and M2 were 1-kb and 0.5-kb size ladders, respectively. The 4-kb band in lane M2 migrated 8.8 cm from the well. Note the additional small bands in lanes 9–13.

(for each strand, assuming a 50–100-base overlap). For re-sequencing purposes, such as in mutation detection where sequencing one strand may be sufficient, this protocol readily covers at least 6 kb.

Although the system described here was limited to inserts of a few kb for de novo sequencing, it still compares favorably with the ExoIII deletion strategy. On large targets (over 2.5 kb) with digestion times of ten minutes or more, ExoIII is known to create internal nicks, so that size fractionation or screening the deletions for size is required before sequencing (13). Long PCR is again the perfect tool for this task, and in fact it may be used as a downstream process with any nested deletion strategy and could easily be automated. It would add most effectively to a protocol with no size-constraining steps such as the pDUAL system, where the deletions are created *in vivo* by transposition.

The present system could be improved in several ways. Engineering a vector totally comprised of essential DNA would increase the capacity for larger inserts and eliminate non-amplifying clones. Then amplimers could be accurately sized immediately without prescreening. An upstream transcription terminator and a low copy number replicon would both help to prevent toxic effects of insert genes. A drug-resistance marker that gives cleaner selection than ampicillin would reduce the chance of picking mixed colonies for amplification, particularly important if clones are not archived. The cloning site and the two or three rare restriction sites could be chosen so that most enzymatic reactions can be done in one buffer. This will eliminate DNA purification between steps. Lastly, an enzyme that cuts only once at a random site in a plasmid molecule would dramatically enhance the efficiency of creating random deletions compared to DNase I.

REFERENCES

1. **Barnes, W.M.** 1994. PCR amplification of up to 35 kb DNA with high fidelity and high yield from 1 bacteriophage templates. *Proc. Natl. Acad. Sci. USA* 91:2216-2220.
2. **Barnes, W.M., M. Bevan and P.H. Son.** 1983. Kilo-Sequencing: creation of an ordered nest of asymmetric deletions across a large target sequence carried on phage M13. *Methods Enzymol.* 101:98-122.

3. **Breault, D.T., M.L. Stover, M. Li, A.C. Lichtler and D.W. Rowe.** 1995. Novel use for BAL31 nuclease-generated nested deletions: sequencing from the inside out. *BioTechniques* 18:614-617.
4. **Burland, V., F.P. Curtis and N. Kusakawa.** 1996. Agarose gel analysis of 15-40-kb PCR amplimers. *BioTechniques* 21:142-144.
5. **Eberle, J.** 1993. Generation of bidirectional deletions using DNase I. *BioTechniques* 14:408-410.
6. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28:351-359.
7. **Hong, G.F.** 1982. A systematic DNA sequencing method. *J. Mol. Biol.* 158:539-549.
8. **Poncz, M., D. Solowiejczyk, M. Ballantine, E. Schwartz and S. Surrey.** 1982. "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. *Proc. Natl. Acad. Sci. USA* 79:4298-4302.
9. **Sambrook, J., E.F. Fritsch and T. Maniatis.** 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed., p. 5.73-5.75. CSH Laboratory Press, Cold Spring Harbor, NY.
10. **Shore, D., J. Langowski and R.L. Baldwin.** 1981. DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. USA* 78:4833-4837.
11. **Slatko, B., P. Heinrich, B.T. Nixon and D. Voytas.** 1991. Constructing nested deletions for use in DNA sequencing, p. 7.2.1-7.2.20. *In* F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.E. Seidman, J.A. Smith and K. Struhl (Eds.), *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.
12. **Wang, G., R.W. Blakesley, D.E. Berg and C.M. Berg.** 1993. pDUAL: a transposon-based cosmid cloning vector for generating nested deletions and DNA sequencing templates *in vivo*. *Proc. Natl. Acad. Sci. USA* 90:7874-7878.
13. **Zhu, K.Y. and J.M. Clark.** 1995. Rapid construction of nested deletions of recombinant plasmid DNA for dideoxy sequencing. *BioTechniques* 18:222-224.

We thank Wendy Baubonis and Foner Curtis for help with sequencing; also members of the E. coli Genome Project at the University of Wisconsin-Madison for E. coli data and a gift of E. coli DNA. Address correspondence to Dr. Noriko Kusakawa, FMC BioProducts, 191 Thomaston Street, Rockland ME 04841, USA. Internet: noriko_kusakawa@fmc.com

Received 14 February 1997; accepted 20 June 1997.

Valerie Burland and Noriko Kusakawa
FMC BioProducts
Rockland, ME, USA