Single-Primer PCR Procedure for Rapid Identification of Transposon Insertion Sites

BioTechniques 28:1078-1082 (June 2000)

Transposons have been used in many laboratories worldwide as a powerful tool for insertional mutagenesis to investigate gene function in bacteria (8). The technique has recently been adapted in which a transposon, such as mini-Tn5, delivers a unique identifying sequence of 40 nucleotides (1). This technique, termed signature-tagged mutagenesis (STM), is widely used in the search for virulence-associated genes via negative selection and allows the large-scale analysis of pooled mutants.

Different procedures have been developed for identification of the transposon insertion regions. Single-primer PCR (4) requires restriction endonuclease digestion, ligation, PCR and cloning steps. It also depends on the availability of appropriate restriction sites. The rapid amplification of transposon ends (RATE) technique (6) involves several steps including ligation of DNA fragments to linkers, single-strand amplification using biotin-labeled primers, isolation of the amplified products and another PCR before the sequencing. A single-primer PCR technique was originally developed by Parks et al. (5) and modified by Spector et al. (7). However, both techniques required intermediate cloning and selection steps before sequencing.

Here, we report a technique that does not require cloning and allows rapid identification of transposon insertion sites using a single PCR primer and a uniquely designed PCR procedure. To test this method, we used mini-Tn5 insertion mutants from a recently constructed *Yersinia pseudotuberculosis* STM library (3).

Figure 1 outlines the strategy for sequencing the transposon insertion regions. The procedure involves just one transposon-specific primer and a single PCR consisting of three rounds of amplification. The first round is designed for linear amplification of single-stranded, transposon-specific templates from the transposon that are to be used in the following steps and is carried out at a standard annealing temperature (50°C). The second round involves a low annealing temperature (30°C) for amplification of the region adjacent to the transposon insertion site. Amplification of this region is possible because, at low annealing temperatures, the primer will bind to sites of limited sequence com-

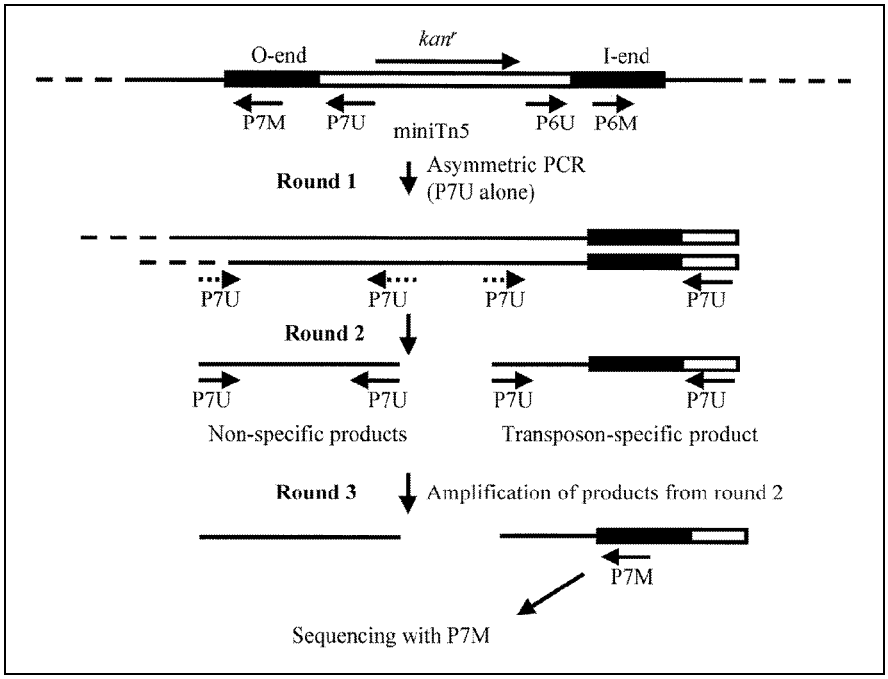

Figure 1. An outline of the single-primer PCR procedure for the identification of transposon insertion sites. A strategy for generating insertion sequences with the use of the P7U PCR primer and the P7M sequencing primer is shown as an example.

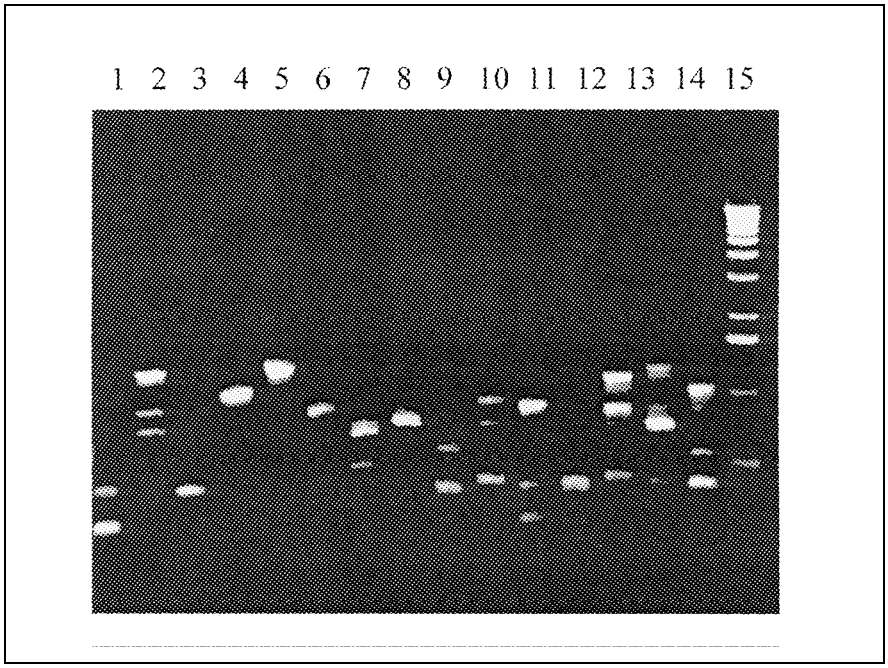

Figure 2. Consistent results of single-primer PCR with different transposon derivatives. Lanes 1–14, PCR products obtained after amplification of DNA from 14 arbitrarily selected derivatives; 15, 1 kb ladder (Life Technologies).

Table 1. Summary of the Sequencing Results

Sequence No.	Sequence name	Length	Readability, %	BLASTx, nrdb	BLASTn, *Y. pestis*
1	1-P6M	LS	-	-	-
2	1-P7M	LS	-	-	-
3	2-P6M	380	100	NS	NS
4	2-P7M	350	86	NS	NS
5	3-P6M	424	100	NS	NS
6	3-P7M	425	100	NS	NS
7	4-P6M	230	99.5	HmuV (90%) 44-85 Q56993.sp, *Y. pestis*	Contig664 (92%) 15,772-15,942
8	4-P7M	656	99.7	HmuV (96%) 93-266 Q56993.sp, *Y. pestis*	Contig664 (97%) 15,952-16,607
9	5-P6M	LS	-	-	-
10	5-P7M	HE	-	-	-
11	6-P6M	558	100	NS	Contig647 (99%) 122,046-122,533
12	6-P7M	477	91.2	ClpA (77%) 569-607 P15716, *E. coli*	Contig647 (79%) 122,571-122,853
13	7-P6M	614	99.8	InsA (82%) 1-90 P03829.sp, *Sh. dysenteriae*	Contig653 (98%) 3,884-4,512
14	7-P7M	372	99.7	NS	Contig653 (97%) 4,531-4,890
15	8-P6M	526	100	MoaC (93%) 30-158 P30747.sp, *E. coli*	Contig660 (99%) 75,734-76,260
16	8-P7M	383	100	TraC (93%) 827-842 p27189.sp, *E. coli*	NS
17	9-P6M	646	99.5	IS element (transposase) 100%, 1-33, CAB54947.1 *Y. pestis*	Contig3 (98%) 49,627-50,171
18	9-P7M	656	97.4	YopP (95%) 119-169 AF053946, *Y. pestis*	Contig3 (97%) 50,560-50,723
19 (98%)	10-P6M	422	100	PotF (86%) 1-76 P3113.sp, *E. coli*	Contig665.0.0 8,978-9,397
20 (95%)	10-P7M	566	97.5	PotF (79%) 110-276 P31133.sp, *E. coli*	Contig665.0.0 9,441-9,999

Readability indicates the ratio of the number of unambiguous bases to the total number in the sequence. BLASTx and BLASTn are similarity search results against the nonredundant protein database (http://www.ncbi.nlm.nih.gov/BLAST) and the nucleotide sequence database of the unfinished *Y. pestis* genome (http://www.sanger.ac.uk/Projects/Y_pestis/). Current (11/15/99) contig number and location are indicated in the last case. Similarity is shown in brackets. Intervals in the BLAST results refer to the region of similarity in the target sequence. LS, low signal; HE, high error; NS, no similarity. P6M and P7M in the sequence names refer to the primers used in the sequencing of P6U and P7U derived PCR products, respectively.

plementarity that are on the opposite strand from the specific primer binding site and close enough to the transposon insertion site for PCR to work. The third round is designed for further amplification of the products from round two and uses a standard annealing temperature.

Crude cell lysates were prepared essentially as described (2). Briefly, 1 mm agar culture was resuspended in 20 µL lysis buffer [1 mg/mL lysozyme, 10 mM Tris-HCl (pH 8.0), 10 mM EDTA and 0.02 % Triton® X-100), incubated at 37°C for 2 min, followed by 15 s at 100°C, and diluted 10× with TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA].

The lysates were used in single-primer PCR with P6U (5′-CGAGCTCGAATTCGGCCTAG-3′) or P7U (5′-CTGCAGGCATGCAAGCTTCG-3′) primers with the following conditions: 1 min at 94°C, 20 cycles of 94°C for 30 s), 50°C for 30 s and 72°C for 3 min; 30 cycles of 94°C for 30 s, 30°C for 30 and 72°C for 2 min; 30 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for 2

min, followed by a 7-min extension at 72°C. The reaction was performed using *Taq* DNA polymerase (Life Technologies, Paisley, UK), standard PCR mixture supplied by the manufacturer and an Omnigene thermocycler (Hybaid, Teddington, UK).

We expected the final products to consist of a mixture of transposon-specific fragments and the products derived from genomic regions unrelated to the transposon insertion site (Figure 1). These nonspecific products would be common to the wild-type and almost all transposon mutants. We presumed, however, that the use of a transposon-specific primer and enrichment for an initial linear stage would favor amplification of the insertion site-related regions. This was confirmed experimentally. As shown in Figure 2, 14 arbitrarily selected derivatives generated different fragment patterns, indicating that the amplification resulted in predominantly Tn-specific PCR products.

To test the validity of the procedure and the reliability of the generated sequencing information, 10 randomly selected Tn derivatives were investigated in a separate experiment. The mixtures of PCR products from each reaction mixture were passed through a MicroSpin™ S-300 HR Column (Bio-Rad Laboratories, Hercules, CA, USA) and sequenced with an internal primer as outlined in Figure 1. DNA sequencing was performed on an ABI 377 automatic sequencer using an ABI PRISM™ dye terminator cycle sequencing kit (PE Biosystems, Foster City, CA, USA) using P7M primer (5′-GCCGAACTTGTGTATAAGAGTC-3′) for P7U-generated products and P6M primer (5′-GCCAGATCTGATCAAGAGAC-3′) for P6U-generated products. The results are presented in Table 1. In the majority of cases, the procedure allowed the generation of low-error sequences of approximately 500 bp. Note that even with the sequence having the highest sequencing error rate (Table 1, 6-P6M), the disrupted gene can be determined unambiguously. Sequencing of the region adjacent to another Tn end (7-P7M) gives confirmatory information as it corresponds to the same region of the *Y. pestis* contig. In general, the results obtained with a second pair of primers (Table 1) clearly demonstrate that the derived sequences belong to the same genome locus.

Insertion sites in only two out of 10 mutants could not be determined. We presume that in some cases failure could be a result of the lack of an appropriate nonspecific priming site for an individual primer in the regions adjacent to particular insertion sites. An average sequence length in the 16 successful samples was 480 bases with an average error rate of 1.3%.

To summarize, a novel, rapid and reproducible technique for establishing transposon insertion sites has been developed. In 16 out of 20 cases, the sequence adjacent to the insertion was determined unambiguously. Note that for other tansposons and bacteria, the optimal annealing temperature may vary. However, as appropriate primers can be designed for each transposon/bacteria combination, the method described here may be applicable to different systems. This technique is suitable for scaling up and for screening different libraries generated by transposon mutagenesis and could be particularly useful for STM.

REFERENCES

1. **Hensel, M., J.E. Shea, C. Gleeson, M.D. Jones, E. Dalton and D.W. Holden.** 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science *269*:400-403.
2. **Karlyshev, A.V. and S. MacIntyre.** 1996. Study of the intergenic *exeF-exeG* region and its application as a simple preliminary test for *Aeromonas* spp. FEMS Microbiology Letters *137*:37-44.
3. **Karlyshev, A.V., P.C. Oyston, K. Williams, R.W. Titball, R. Davis, E.A. Winzeler and B.W. Wren.** Discovery of novel *Yersinia pseudotuberculosis* virulence-associated genes using high-density array based signature-tagged mutagenesis. (In press).
4. **Novak, J., L. Novak, G.R. Shah, W.A. Woodruff and P.W. Caufield.** 1997. Transposon mutagenesis: cloning of chromosomal DNA from the site of Tn916 insertion using polymerase chain reaction. Biotechnol. Techniques *11*:51-54.
5. **Parks, C.L., L.-S. Chang and T. Shenk.** 1991. A polymerase chain reaction mediated by a single primer: cloning of genomic sequences adjacent to a serotonin receptor coding region. Nucleic Acids Res. *19*:7155-7160.
6. **Ribot, E.M.** 1998. Rapid amplification of transposon ends for the isolation, cloning and sequencing of transposon-disrupted chromosomal genes. BioTechniques *24*:16-22.
7. **Spector, M.P., C.C. DiRusso, M.J. Pallen, F.G. del Portillo, G. Dougan and B.B. Finlay.** 1999. The medium-/long-chain fatty acyl-CoA dehydrogenase (*fadF*) gene of *Salmonella typhimurium* is a phase 1 starvation-stress responce (SSR) locus. Microbiology *145*:15-31.
8. **Voelker, L.L. and K. Dybvig.** 1998. Transposon mutagenesis. Methods Mol. Biol. *104*:235-238.

This work was supported by The Wellcome Trust, UK. *Address correspondence to Dr. Andrey V. Karlyshev, Department of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, University of London, Keppel Street, London, WC1E 7HT, UK. Internet: andrey.karlyshev@lshtm.ac.uk*

Received 15 December 1999; accepted 23 February 2000.

Andrey V. Karlyshev, Mark J. Pallen[1] and Brendan W. Wren
University of London
London
[1]The Queen's University of Belfast
Belfast, UK