LINE-1 insertion dimorphisms identification by PCR

Wichai Pornthanakasem and Apiwat Mutirangura
Chulalongkorn University, Bangkok, Thailand


Differentiating between recent LINE-1 (L1) insertion dimorphisms (LIDs) is predominantly useful for studying not only the transposition mechanism but also population genetics or evolution based on polymorphic markers where the ancestral state is known (i.e., absence of the insertion) (1–5). Previous techniques such as L1 display (1) or ATLAS (2) have identified new LIDs from distinct populations. To improve both efficiency and simplicity, we designed a PCR-based approach, LID identification by PCR (LIDSIP), by combining and modifying ligation-mediated PCR (LMPCR) (6) and interspersed repetitive sequence PCR (IRSPCR) (7). This method requires only a small amount of DNA, a limited and specific number of PCRs to provide a genome-wide scan within a PCR range of an appropriate restriction site, conventional molecular genetic techniques such as agarose gel electrophoresis devoid of radioactive label, and in addition, this method yields clear, easily distinguishable, specific results.

The aim of applying LIDSIP was to globally map active L1 in the human genome by amplifying the 3′ untranslated region (UTR) end of the L1-Ta subset up to its next specific restriction enzyme recognition sequences. First, 500 ng human genomic DNA were digested with BstYI (New England Biolabs, Beverly, MA, USA) and purified with phenol-chloroform extraction and ethanol precipitation. Second, the DNA was ligated to 20 pmol each of LIDSIP-LINK (5′-AGG-TAACGAGTCAAGACCGACTC-GTGGAGCT-3′) and BstYI-LINK (5′-GATCAGTCCAGCAGAG-3′) using T4 DNA ligase (New England Biolabs) at 16°C overnight. Finally 50 ng of the purified ligated DNA were subjected to nested PCR. The first PCR (50 µL total volume) contained 200 µM each of the four dNTPs, 1x PCR buffer (contains 1.5 mM MgCl2, Qiagen, Valencia, CA, USA), 0.1 U HotStarTaq (Qiagen), 0.4 µM of L1-ACA (5′-GAGATCT-TACCTAATGCTAGATGACACA-3′) (1), and 5′ linker (5′-AGTGAAC-GAGTGAGACCACCGA-3′). The PCR amplification was performed as follows: initial denaturation at 95°C for 15 min, followed by 20 cycles of denaturation at 95°C for 45 s, annealing at 57°C for 45 s, extension at 72°C for 2 min, and a final extension at 72°C for 7 min. To divide the number of product types into subsets distinguishable by electrophoresis, 20 cycles of nested PCR were performed using 32 different chimeric primers. This step helps fractionate the repetitive products similar to Msrl3+ primers in sequence-specific amplification polymorphism (S-SAP), which was developed to display plant retrotransposons (8). Each reaction consisted of 20 amplification cycles containing 1 µL primary PCR product, one of 32 chimeric primers [5′-GACTCGGACGACGTGAC(T/C)XX-3′, where (C/T) was C or T, and XX was 2 bp of randomly selected sequences; e.g., TGA chimeric is 5′-GACTCGGACGACGTGAC(TGA-3′)], GCNP (1), and the L1-nested-primer, GCGACCAGCATGGCACA. This is further illustrated by the schematic diagram (Figure 1) and an example of LIDSIP (Figure 2a), respectively. When LIDSIP products were compared among 14 Thai individuals, 37 candidate LIDs could be identified. On average, each nested PCR yielded 10 out of a total of 300 products per sample. Five candidate LIDs were selected, cloned to the pGEM-T easy cloning vector (Promega, Madison, WI, USA), and transformed into Escherichia coli DH5α. DNA sequencing was performed on an ABI 3100 DNA sequencer (Applied Biosystems, Foster City, CA, USA), and results were analyzed by Basic Local Alignment Search Tool (BLAST) program (www.ncbi.nlm.nih.gov/BLAST). All clones revealed upstream L1, integration site, single copy, and chimeric primer sequences. Nevertheless, the BLAST results were devoid of L1 upstream of the flanking sequence, and only the preintegration sites were actually present in the human genome database (GenBank accession nos. AC015547 at 86,101 bp, AC087307 at 13,330 bp, AL392087 at 11,869 bp, AL583842 at 37,981 bp, and AP001996 at 37,981 bp), suggesting recent retrotranspositions. LID PCRs (1,2) were performed, and polymorphisms of all markers were confirmed (Figure 2b).

Applying LIDSIP has improved both simplicity and efficacy compared with the two previous PCR techniques. The first, L1 display (1), screens for LIDs by PCR using arbitrary primers, Southern blot analysis, and hybridization for identification. Each PCR yields a

---

**Figure 1. LINE-1 insertion dimorphisms identification by PCR (LIDSIP) diagram.** A modified ligation-mediated PCR (LMPCR) linker (black line) was ligated to genomic DNA (grey line) at the BstYI site. The L1-Ta subset is illustrated as a black triangle. The dashed lines represent the first and nested products. The arrows are primers. L1-ACA and GCNP are 3′ sequences of L1. The linker primer is the 5′ sequence of the linker. The chimeric primer is the 3′ sequence of the linker and the restriction site plus a randomly unique sequence of two nucleotides, illustrated as X.
limited number of products. Therefore, to cover the whole human genome, the technique requires both a significant amount of DNA and number of PCRs. Both ATLAS and LIDSIP apply the same principle of single-site PCR and yield more sequences depending on restriction enzyme sites next to L1. Because of the large number of size-variable templates, with ATLAS, those products have to be labeled with radioactive [$^{32}$P]ATP and separated by denaturing long-range polyacrylamide gel electrophoresis. With LIDSIP, radioactive labeling is not required, and gel recovery is greatly facilitated. Hence, by dividing L1 products into several subsets with nested PCR using chimeric primers, size differentiation by agarose gel electrophoresis becomes feasible. Additionally, LIDSIP is able to limit utilization of the original DNA by the nested PCR when reproducing LID recovery is needed. The simplicity of LIDSIP will not only globally expand the discovery rate, but also, by adjusting restriction enzymes and nested chimeric primers, the methodological principle of S-SAP can be modified to identify polymorphic genome insertions of other mobile elements (8).

ACKNOWLEDGMENTS

This study has been supported by the Royal Golden Jubilee Ph.D. program, Molecular Biology and Genetics of Cancer Development Research Unit, Chulalongkorn University, and the Thailand Research Funds. We thank Ms. Petra Hirsch for her critical review of the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare no conflicts of interest.

REFERENCES


Received 3 June 2004; accepted 28 June 2004.

Address correspondence to Apiwat Mutirangura, Genetics Unit, Department of Anatomy, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand. e-mail: mapiwat@chula.ac.th

Protocol for amplification of GC-rich sequences from Pseudomonas aeruginosa

Aniruddha Raychaudhuri and Peter A. Tipton
University of Missouri-Columbia, Columbia, MO, USA

Pseudomonas aeruginosa is an opportunistic, ubiquitous pathogen that causes life-threatening infections in cystic fibrosis, burn, and cancer patients (1). A complicating factor in P. aeruginosa infections is that the bacteria secrete an exopolysaccharide called alginate, which is a constituent of the biofilm that contributes to antibiotic resistance and protects against the defenses of the host immune system. The Pseudomonas genome project has yielded the sequence of all of the genes in the genome (2), but detailed study of the functions of their encoded proteins requires a facile means for amplification of the genes. However, a major problem encountered in amplifying full-length genes in P. aeruginosa stems from the high GC content of its genome, which means that stable secondary structures often form in the DNA that reduce or halt the progression of DNA polymerase during amplification. The mean GC content in the P. aeruginosa genome is 67.2%, with even higher GC content

Figure 2. LINE-1 insertion dimorphisms identification by PCR (LIDSIP) and LINE-1 insertion dimorphism (LID) PCR. (A) LIDSIP products from four individuals, lanes 1–4. Each was identified by two different chimeric primers. TGA chimeric and TCG chimeric. Candidate LIDs were indicated at the arrows. M is the M24 100-bp + 1.5-kb DNA ladder (SibEnzyme, Novosibirsk, Russia). (B) Two LID PCRs demonstrated polymorphisms among six different individuals, lanes 1–6. M is the M24 100-bp + 1.5-kb DNA ladder.