Self-Formed Adaptor PCR: a Simple and Efficient Method for Chromosome Walking

Shiming Wang, Jian He, Zhongli Cui, and Shunpeng Li

Key Laboratory of Microbiological Engineering of Agricultural Environment, Department of Microbiology, Nanjing Agricultural University, Nanjing, People’s Republic of China, and Department of Bioengineering, School of Chemical Engineering, Nanjing University of Science and Technology, Nanjing, People’s Republic of China

Received 22 December 2006/Accepted 1 May 2007

We developed a self-formed adaptor PCR (termed SEFA PCR) which can be used for chromosome walking. Most of the amplified flanking sequences were longer than 2.0 kb, and some were as long as 6.0 kb. SEFA PCR is simple and efficient and should have broad applications in the isolation of unknown sequences in complex genomes.

The PCR-based methods developed to clone the flanking DNA sequences of a known region can be divided into three types: inverse PCR (1, 6, 8, 12), ligation-mediated PCR (2, 4, 5, 9, 14, 19), and randomly primed PCR (3, 7, 17, 18). The first two types require enzyme digestion and ligation, making them relatively expensive and time-consuming. Thermal asymmetric interlaced PCR (TAIL PCR) (3), a representative of the third type, has gained popularity for its simplicity. It needs no prior DNA manipulation (such as enzyme digestion and ligation) or further manipulation after PCR (such as cloning the PCR products for screening). It can be extended to large numbers of reactions and complex genetic systems (10, 11, 13, 16), so it is widely used for chromosome walking. However, the cloned sequence length from TAIL PCR is usually less than 1.0 kb.

SiteFinding-PCR (15) is another newly developed randomly primed PCR. It is quite efficient, but the PCR products still require cloning and further screening to get a high specificity.

Here, we present a new PCR method for chromosome walking, i.e., self-formed adaptor PCR (SEFA PCR). It combines the advantages of ligation-mediated PCR in its specificity and that of TAIL PCR in its simplicity. The principle behind SEFA PCR is illustrated in Fig. 1. The four primers are located sequentially on the known DNA sequences. SP1, SP2, and SP4 are specific primers and have relatively high annealing temperatures (e.g., 70°C). SP3, e.g., 5’-TACCCAAAGAAGCAGGA ANNNNNNNNTGAAGAAA-3’), is partially degenerate primer, plays a key role in the process. Its specific parts are taken from the known target DNA sequence (Fig. 1a). First, a single cycle of PCR was carried out at a low annealing temperature (e.g., 35°C) with only primer SP3. At this low annealing temperature, SP3 can prime and elongate at many positions on the DNA template (Fig. 1b). A position probably exists somewhere downstream of the known DNA sequence where SP3 primes and extends, thus creating a nascent single strand which has a binding site for SP1. After a single cycle of PCR, the annealing temperature is increased to the point (e.g., 70°C) corresponding to the annealing temperature of SP1. Then, SP1 is added to the reaction mixture. At this high annealing temperature, only SP1 can prime the target site efficiently, thus creating a pool of single-stranded DNA with the SP1 sequence at the 5’ end and the SP3 complementary sequence at the 3’ end (Fig. 1c).

During the performance of the above-described PCR, SP1 could still possibly prime some other sites besides the correct target site, even at the high annealing temperature. This is also the case for SP3. Therefore, final wrong products fall into three categories based on combinations of the two primers: (i) those primed only by SP1, (ii) those primed by SP1 and SP3, and (iii) those primed only by SP3. The first two kinds would be diluted out during the second round of PCR primed by the inner single primer SP2. Only the third kind would cause problems, because it is possible to take the correct products as templates and extend from both ends, creating adaptors at both ends. Fortunately, all three products are amplified in a decreasing order of efficiency at high temperature, so the correct target sequences primed by SP1 and SP3 are amplified much more efficiently than the false ones primed only by SP3.

The detailed procedures were as follows. The long and accurate Taq buffer, and deoxynucleoside triphosphates were purchased from TaKaRa Biotechnology Co., Ltd. The PCR mixture included 15 μl of 2× GC buffer I, 5 μl of 2.5 mM deoxynucleoside triphosphates, 1.5 U of long and accurate Taq enzyme, and about 50 ng (for bacteria) or 1 μg (for plants and the fungus) of template genomic DNA, with deionized water added to 30 μl. All PCRs were run on a PTC-200 Peltier thermal cycler. The detailed thermal cycling conditions for SEFA PCR are listed in Table 1.

After SEFA PCR, a second round of nested PCR was run with the single primer SP2. The reaction mixture was the same as that for SEFA PCR except for the template and primer: 0.1 μl of the above SEFA PCR product (or diluted 10 times) and...
3 μl of 5 μM SP2 were added to the reaction mixture. About 20 cycles (for bacteria) or 30 cycles (for plants) of normal PCR protocol were run, namely, (i) denaturing at 94°C for 30 s and (ii) annealing and extension at 70°C for 5.5 min. If needed, a third round of thermally asymmetric PCR was run to improve the specificity with primer SP4 (e.g., annealing at 70°C) and the other short primer, SP5 (e.g., annealing at 60°C), positioned between SP2 and SP3. The reaction mixture was also the same as that for SEFA PCR except for the template and primer: 0.1 μl of the second-round PCR product, 3 μl of 5 μM SP4, and 0.3 μl of 5 μM SP5 were added to the reaction mixture. About 10 cycles of the thermally asymmetric protocol listed in Table 1 were run.

Using the above protocol, we cloned the sequences involved.

### TABLE 1. Thermal cycling conditions for SEFA PCR

<table>
<thead>
<tr>
<th>Primer addition</th>
<th>No. of cycles</th>
<th>Cycling conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add 1 μl of 5 μM SP3</td>
<td>1</td>
<td>(i) Denaturing at 94°C for 90 s; (ii) annealing at 35°C to 45°C for 3 min and then ramping to 70°C at 0.2°C per s; and (iii) extension at 70°C for 5 min</td>
</tr>
<tr>
<td>Add 3 μl of 5 μM SP1</td>
<td>25</td>
<td>(i) Denaturing at 94°C for 30 s and (ii) annealing and extension at 70°C for 5.5 min</td>
</tr>
<tr>
<td>Add no primers</td>
<td>5 to 10 cycles of thermally asymmetric protocol</td>
<td>(i) Two cycles of denaturing at 94°C for 30 s and annealing and extension at 70°C for 5 min; (ii) one cycle of denaturing at 94°C for 30 s, annealing at 55°C to 60°C for 30 s, and extension at 70°C for 5 min</td>
</tr>
</tbody>
</table>
in p-nitrophenol degradation from *Pseudomonas putida* DLL-1 and the promoter sequence of the squalene synthase gene from *Ganoderma lucidum* (Fig. 2). By testing five plant genomes, we found that chromosome walking also works well on complex genomic systems (Fig. 3). The known DNA sequences used for chromosomal walking and the primers for each walking are listed in the supplemental material. A clear main DNA band (indicated by the arrows in Fig. 2 and 3) appeared at the end of the second round of nested PCR, except for lane 3 of the P3 strain DNA. To narrow down the correct band of the exception, we ran a third round of thermally asymmetric PCR with a long inner primer, SP4 (melting temperature \( T_{m} \), 72°C), and the other short one, positioned between SP2 and SP3 \( (T_{m}, 62^\circ \text{C}) \). A clear band appeared (Fig. 2, lane 3). All final main DNA bands were recovered and sequenced by using SP4 directly whenever possible or cloned into a T vector (pMD18-T; TaKaRa Company) for sequencing. The sequenced results (see the supplemental material) showed that all the recovered sequences, except in the case of lane 5 of the rice DNA, had part of the known DNA sequences at their 5' ends, thus proving the specificity of the method.

From the 17 walking tests, we obtained 15 correct target sequences after two rounds of PCR and 1 correct target sequence after three rounds of PCR (lane 3). Of all 16 correctly cloned target sequences, 4 were about 6 kb, 10 were between 2 and 6 kb, and only 2 were below 1 kb.

The following aspects of SEFA PCR that are different from normal PCR should be noted: (i) the DNA template concentration should be high to facilitate the creation of the adaptors; (ii) the concentrations of the pairs of primers in SEFA PCR and the possible third round of thermally asymmetric PCR should be asymmetric, and the concentration of primers SP1 and SP4 should be high, while the concentration of primers SP3 and SP5 should be low to increase the specificity; and (iii) SP1 should be added at a temperature above its annealing temperature to improve its specificity.

This work was supported by the National Natural Science Foundation program (grant 40471073) and the Agricultural Technology Transfer program (grant 2004.514).

We thank Zhengqiang Ma for providing the plant genomic DNA samples and Mingwen Zhao for providing the mushroom genomic DNA samples.

REFERENCES


FIG. 2. Chromosome walking of bacterial and fungal genomic DNA. P1, P2, and P3 are three transposon insertion mutants of *P. putida* DLL-1. Lane 1 is the product of the first round of SEFA PCR (indicated in Fig. 1). Lane 5 and lane 3 are the products of the second round of nested PCR (primed by only SP2), walking into the 5' and 3' end of the known DNA sequence, respectively. Lane 3+ is the product of the final third round of thermally asymmetric PCR, walking into the 3' end of the known DNA sequence. Lane M1: λ-HindIII digest marker; from top to bottom, 23.13, 9.416, 6.557, 4.361, 2.322, and 2.027 kb.

FIG. 3. Chromosome walking of five plant genomes. The lane numbers that are the same as in Fig. 2 have the same meanings as given in the Fig. 2 legend. Lane M2, λ-HindIII and EcoRI digest markers of (from top to bottom) 21.227, 5.148, 4.296, 3.530, 2.027, 1.581, 1.375, 0.941, 0.831, and 0.564 kb.