

Restriction Site-dependent PCR: An Efficient Technique for Fast Cloning of New Genes of Microorganisms

Yu JIANG^{1,2}, Jianjun PEI¹, Xin SONG¹ and Weilan SHAO^{1,*}

Jiangsu Key Laboratory for Biodiversity and Bio-Resources, Nanjing Normal University, Nanjing, Jiangsu 210046, People's Republic of China¹ and Laboratory of Molecular Microbiology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, People's Republic of China²

(Received 14 September 2007; accepted 14 November 2007; published online 17 December 2007)

Abstract

New bioactive proteins need to be screened from various microorganisms for the increasing need for industrial and pharmaceutical peptide, proteins, or enzymes. A novel polymerase chain reaction (PCR) method, restriction site-dependent PCR (RSD-PCR), was designed for rapid new genes cloning from genomic DNA. RSD-PCR strategy is based on these principles: (i) restriction sites disperse throughout genomes are candidacy for universal pairing; (ii) a universal primer is a combination of a 3'-end of selected restriction sites, and a 5'-end of degenerated sequence. A two-round PCR protocol was designed and optimized for the RSD-PCR: amplify the single strand target template from genomic DNA by a specific primer and amplify the target gene by using the specific primer and one of the universal RSD-primers. The optimized RSD-PCR was successfully applied in chromosome walking using specific internal primers, and cloning of new genes using degenerated primers derived from NH₂-terminal amino acid sequence of protein.

Key words: degenerated primer; new gene cloning; polymerase chain reaction

1. Introduction

Bioactive proteins from various microorganisms have become attractive alternative to conventional catalysts for production of fine chemicals and pharmaceutical intermediates. However, the properties of the current available proteins do not always meet the criteria of the applications of interest. Directed evolution is a powerful tool for adapting the characteristics of a bioactive protein, but the selection of the evolved variants is always a bottle-neck step. Novel genes and proteins still need to be screened from a wide variety of microorganisms to meet the increasing requirements of various industrial and pharmaceutical peptide, proteins, or enzymes. Thus the current challenge is focus on the development of efficient rapid screening

methods of related genes. However, classical gene cloning is a tedious and time-consuming work which usually requires constructing and screening a library, and delicate work often does not assure successful cloning. Required steps often include probe labeling, suitable restriction digestion, and Southern hybridization followed by construction and screening of a library.

Polymerase chain reaction (PCR) has been employed to amplify DNA fragments for molecular cloning, and is especially useful in cDNA cloning using reverse transcriptase PCR (RT-PCR). The core point of PCR technique is that the sequences for short oligonucleotide primers annealing must be available for DNA polymerization. In most cases, a problem arises from the shortage of information when using PCR technique to clone a new gene or desired DNA fragment, because the sequence at the other end of the target DNA is unknown. The success of RT-PCR relies on a universal primer complementary to the polyA tails of mRNA, which is combined with a specific primer of the known sequence.

Edited by Dr Osamu Ohara

* To whom correspondence should be addressed. Tel. +86 25-85891836. E-mail: wlshao@jmail.com.cn

© The Author 2007. Kazusa DNA Research Institute.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

PCR techniques employing various methods to yield efficient universal primers for chromosome walking on an unknown DNA segment are classified into three types: inverse PCR,¹⁻⁴ ligation-mediate PCR,⁵⁻¹⁷ and random primer PCR.¹⁸⁻²⁵ In general, the current PCR technique for chromosome walking suffers from these limitations: (i) they either face the high-complexity of whole genomes and require delicate manipulations, or (ii) give contaminated products, and (iii) few of them are actually suitable for new gene cloning using degenerated primers based on amino acid sequence or conserved gene locus. Thus, PCR methods are commonly performed in combination with the traditional cloning methods, such as library construction, probe labeling, Southern hybridization, and so on.

We report here a simplified and effective PCR method, restriction site-dependent PCR (RSD-PCR), which used the natural restriction sites in the genomic DNA to design universal primers, and was applied with a two-round PCR protocol without nested process. The RSD-PCR protocols were optimized, and successfully used for chromosome walking and new gene cloning via degenerated primers.

2. Materials and methods

2.1. Cell cultivation and genomic DNA isolation

Chromosomal DNA of *Thermoanaerobacter ethanolicus* JW200, *Thermotoga maritima*, and *Schizosaccharomyces pombe* YHL6403 was gifts from Dr J. Wiegel and Dr M. Adams (University of Georgia, USA) and Dr Ying Huang (Nanjing Normal University, China). *Bacillus pumilus* ARA was isolated in the laboratory, and grown at 50°C on liquid medium containing (w/v): 1.5% xylan; 0.3% K₂HPO₄; 0.2% yeast extract. *Escherichia coli* JM109 and their recombinants were routinely grown aerobically in Luria-Bertani (LB) medium at 37°C. Plasmids or DNA fragments were isolated from *E. coli* cells or agarose gels with QIAGEN Plasmid Mini Kit, or QIAquick Gel Extraction Kit, respectively.

2.2. Obtaining an amino acid sequence of a novel arabinosidase

Arabinosidase activity was measured with standard procedure at 60°C.²⁶ Protein concentration was assayed by Bradford method with bovine serum albumin as a standard, an arabinosidase was partially purified by (NH₄)₂SO₄ precipitation, DEAE-Sephacel, Butyl HIC, and HW-55F-toyopearl chromatography (Sigma). The protein of main band in native gradient gel exhibiting arabinosidase activity was transferred to polyvinylidene difluoride (PVDF) membrane by electro-blotting for NH₂-terminal amino acid sequencing (GeneCore Company, Shanghai, China).

2.3. Design of a RSD-PCR protocol and a group of RSD primers

To reduce non-specific amplifications, the RSD-PCR protocol was designed in two rounds. In the first round (1st round), the single strand target gene from genomic DNA was amplified by single specific primer under stringent conditions. PCR mixture of 1st round in 50 times dilution was used as the template in the second round (2nd round). Then, double strand target gene was amplified by a universal primer and the specific primer in this round (Fig. 1). The T_m in the experiment was estimated by using formula $T_m = 4 \times (G + C) + 2 \times (A + T)$.

The universal primers were a group of RSD-primers designed on the basis of the restriction sites commonly used at multiple cloning sites in vectors, which contained a 3'-end complementary to six bases of a selected restriction site, and a 5'-end of degenerated nucleotide sequences for pairing with variable bases to support the anneal of the primer to its template. To lower the melting temperature of primer-dimers, proper numbers of complementary bases at the 5'-end of RSD-primers were removed from the highly degenerated bases to make sure that at least three bases were not complementary at each end of a primer-dimer (Table 1). Presently, no formula has been found able to help this kind of primer design because the sequences for restriction enzymes are different. The

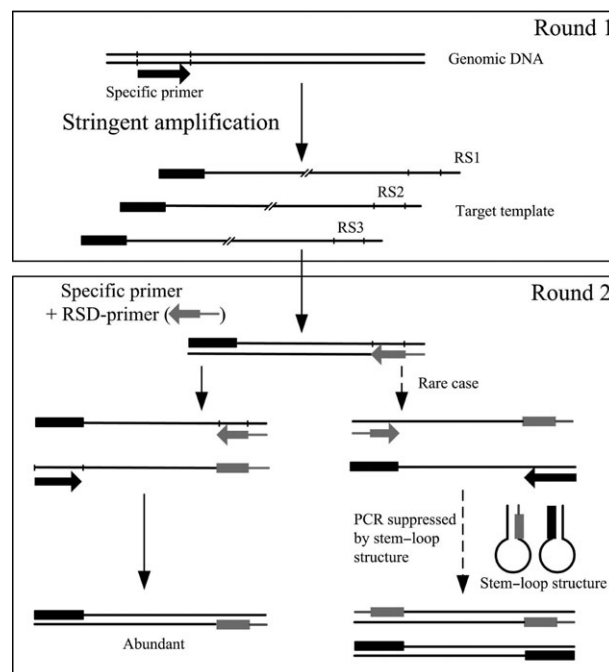


Figure 1. Schematic outline of RSD-PCR protocol for new gene cloning from chromosome DNA. In the round 1, the target template is amplified under stringent conditions from genomic DNA by using a specific primer which can be degenerated. In the round 2, the double strand target gene is amplified by using the same specific primer and one of the universal RSD-primers which can anneal with the corresponding restriction sites (RS) disperse in the target template.

Table 1. Primer sequences in gene cloning

Primers	Designation	Oligonucleotide primer	Basic sequence or reference
Universal primers	Hind-dep	5'-(t/c) (t/c) (t/c)(a/t/g)(a/c/g)(a/c/g) _{n5} AAGCTT-3'	Corresponding restriction sequence
	EcoR-dep	5'-(t/c)(t/c/g)(t/c)(a/c/g)(a/c/g)(a/t/g) _{n5} GAATTC-3'	
	Kpn-dep	5'-(a/c)(a/c)(a/c)(g/t/c)(g/a/t)(g/a/t) _{n5} GGTACC-3'	
	Nco-dep	5'-(t/g)(t/g)(t/g)(a/g/c)(a/t/c)(a/t/c) _{n5} CCATGG-3'	
	Pst-dep	5'-(a/g)(t/g)(a/c)(a/g)(t/c)(a/t/c) _{n5} CTGCAG-3'	
Degenerated specific primers	araN1	5'-AA(t/c)GGnACnGTnAAAGT-3'	Amino acid sequence NGTVKV
	araN2	5'-AA(t/c)GGnACnGTnAAGGT-3'	
	araN3	5'-AA(t/c)GGnACnGTnAA(a/g)GT-3'	
Specific primers	AK	5'-GCGTTGAATCTTTCC-3'	
	CUT	5'-TGTAGCAAGCTCTAC-3'	
	adh-N	5'-GGGGAATTCATGAAAGGTTTTGCAATG-3'	
	adh-C	5'-TTTCTCGAGTATTACAACAGGTTTGATTA -3'	
	ara-up	5'-CTCTGAATCTTTCC-3'	
	adh-up	5'-GCTAGAGGTCTTACG-3'	
	adh-down	5'-CGGTTTATGCCCTG-3'	
	M13-forward	5'-CCAGTCACGACGTTGT-3'	
	M13-reverse	5'-CAGGAAACAGCTATGAC-3'	

primers in Table 1 were designed manually by pairing the sequence of a primer in antiparallel orientation at all possible positions from two bases protruding at 5'-end to two bases protruding at 3'-end, and removing those base(s) from the first 3 Ns at 5'-end to make three uncomplimentary bases including protrusions at each end. In this way, the melting temperature of a primer-dimer should be at least 8°C lower than that of a template-primer hybrid when there is any G or C in the restriction site (G or C stands for 4°C, A or T stands for 2°C), so that most of the primer-dimers will be denatured at the stringent annealing temperature during PCR.

2.4. Performance and optimization of RSD-PCR

The acetate kinase gene *ak* of *T. maritima* was employed as a sample gene to optimize the RSD-PCR protocol. A gene-specific primer AK complementary to a region near the stop codon of acetate kinase gene (*ak*) was designed based on the NCBI online sequence (accession no. NC_000853) (Table 1). In the 1st round PCR of 40 cycles, the 50 µL *ex Taq*-DNA polymerase buffer contained 1.25 U *ex Taq*-DNA polymerase (TaKaRa), and 2.5 mM each of dATP, dTTP, dCTP, and dGTP. Different amount of genomic DNA templates were tested, and optimal annealing temperature in the first round (T_a^{1st}) was also determined for the best resolution. In the 2nd round, PCR conditions were optimized over the concentration of RSD-primer, the annealing temperature in the second round (T_a^{2nd}), and the amount of the single strand template amplified in the 1st round PCR. The optimized RSD-PCR cycling conditions on PTC-200 Peltier Thermal Cycler (MJ Research) was performed as follows: 40 cycles (94°C, 30 s; 54°C, 30 s; 72°C, 2 min) for the first round of PCR, and 35 cycles (94°C, 30 s; 48°C, 30 s; 72°C, 2 min) for the second round of PCR.

The PCR products were partitioned by electrophoresis on agarose gel with DNA markers (TaKaRa). DNA fragments in single bands were isolated from agarose gel, ligated into pMD19T (TaKaRa), and transformed into *E. coli* JM109 competent cells by electroporation using GenePulser Xcell (Bio-Rad). Positive clones were selected on LB-ampicillin (100 µg mL⁻¹) plates, and confirmed by colony PCR employing M13 forward and reverse primers (Table 1). To determine the sequences of inserted DNA, the recombinant plasmids were isolated from the clones, and sequenced by dideoxy-chain termination method at Invitrogen Ltd. (Shanghai, China).

2.5. Selective amplification from degenerated specific primers

To clone the new arabinosidase gene from *B. pumilus* ARA, degenerated primers were synthesized based on the NH₂-terminal amino acid sequence. In order to compare the specificity of amplification caused by the degeneration at the 3rd base from 3'-end, three oligonucleotides were synthesized as primers annealing to the complementary sequence of the gene deduced from amino acid sequence NGTVKV, which were designated as araN1, araN2, and araN3 (Table 1). In the 1st round PCR (50 µL), 0.2 µM degenerated araN1, araN2, or 0.4 µM araN3 was used in the corresponding reaction; and in the 2nd round PCR, 4 µM araN1, araN2, or 8 µM araN3 was used. The other parameters were the same as the optimized RSD-PCR protocol (Fig. 2A).

2.6. Chromosome walking by RSD-PCR

To find the promoter for the newly cloned arabinosidase gene, *abfB* of *B. pumilus* ARA, the specific upstream walking primer ara-up located ~108 bp downstream the start codon was designed (Table 1) based on partial

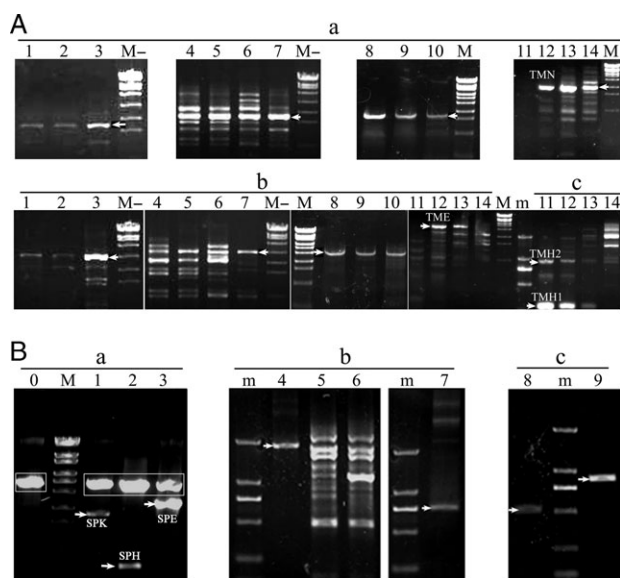


Figure 2. Amplification of a yeast gene and some bacterial genes by using the RSD-PCR protocol (5 μ L loading of the 2nd round products). (A) Following conditions were routinely used except otherwise indicated: for the 1st round PCR, 50 μ L reaction mixture contained 500 ng genomic DNA (*Thermotoga maritima*) and 0.02 μ M specific primer AK; for the 2nd round PCR, 50 μ L reaction mixture contained 1 μ L PCR products amplified in the 1st round, 0.2 μ M specific primer AK and 8 μ M RSD-primer which was (a) NcoI-dep, (b) EcoRI-dep, or (c) HindIII-dep. The corresponding specific PCR products were designated as (a) TMN, (b) TME, or (c) TMH1 and TMH2 as arrow indicated. Lanes 1–3: 50, 100, and 500 ng of genomic DNA per 50 μ L reaction mixture; lanes 4–7: the annealing temperature of the 1st round PCR was 44°C, 47°C, 51°C, and 54°C; lanes 8–10: the annealing temperature of the 2nd round PCR was 48°C, 44°C, and 40°C; and lanes 11–14: the RSD-primer in the 2nd round PCR was 16, 8, 4, and 2 μ M. (B) The RSD-PCR conditions were the same as basic procedures described in (i) except otherwise indicated. The DNA and primer amount: in the 1st round PCR (50 μ L), 3 μ g genomic DNA for *S. pombe*, 0.2 μ M degenerated araN1, araN2, or 0.4 μ M araN3 was used in the corresponding reaction; and in the 2nd round PCR, 4 μ M araN1, araN2, or 8 μ M araN3 was used. (a) *cut6* from *S. pombe* by (lane 0) CUT, and CUT and (lane 1–3) KpnI-dep, HindIII-dep, or EcoRI-dep. The corresponding specific bands were designated as (lane 1) SPK, (lane 2) SPH, or (lane 3) SPE as arrow indicated. The linearized stem-loop bands were indicated by the circle; (b) *abfB* and the upstream sequence from *B. pumilus* ARA by araN1 and PstI-dep (lane 4), araN2 (lane 5) and PstI-dep, araN3 and PstI-dep (lane 6), and ara-up and HindIII-dep (lane 7); and (c) genes flanking *adhB* from *T. ethanolicus* JW200 by *adh*-up and EcoRI-dep (lane 8), and *adh*-down and NcoI-dep (lane 9). DNA markers: M, λ -EcoT14 digest (19329, 7743, 6223, 4254, 3472, 2690, 1882, 1489, 925, 421, 74 bp); M-, the 3472 bp fragment missing of M; m, DNA marker of D2000 (2000, 1000, 750, 500, 250, 100 bp).

sequence. To locate the alcohol dehydrogenase (ADH) gene (*adhB*) in *T. ethanolicus* JW200, specific primers *adh*-N and *adh*-C (Table 1) were designed based on the sequence *adhB* of *T. pseudoethanolicus* 39E (accession no. U49975). *Eco*RI and *Xho*I restriction sites were introduced in each primer, respectively, for convenient cloning into expression vector pET20b for sequencing. According to the internal sequence obtained from cloned gene fragment, the walking primers *adh*-up and

adh-down were designed (Table 1). To test the RSD-PCR protocol over the eukaryotes, a gene-specific primer CUT (Table 1) near the start codon of acetyl-CoA carboxylase gene (*cut6*) in *S. pombe* was designed based on the NCBI online sequence (accession no. Z99261).

3. Results and discussion

3.1. Designing and optimizing the RSD-PCR protocol

The RSD-PCR protocol included the following steps: (i) design and synthesize a specific primer and a set of universal RSD-primers; (ii) amplify the target template from genomic DNA by using a specific primer which can be degenerated; and (iii) amplify the target gene from the diluted template by the specific primer and one of the universal primers (Fig. 1).

Before the RSD-primers in Table 1 were designed, another universal inosine-primer was tested, and the results indicated that 11 pairs between inosine bases and nucleotide bases were not sufficient to support the annealing of the primer for amplification. Amplification occurred between the same specific primers when using the inosine-primer to clone a gene encoding the most abundant extracellular protein from straw mushroom. This may be because that the T_m value of the inosine-primer was too low for annealing at the anneal temperature, or polymerizing at the elongation temperature. The same fragments were amplified even after the annealing and the elongation temperatures were lowered down to 38°C and 60°C using *Z* *Taq*-DNA polymerase (Takara) (data not shown). The RSD-primers were thus designed and proved useful in RSD-PCR when combined with a specific primer to amplify new DNA fragments. The parameters were further optimized with cloning of the published *ak* from *T. maritima* as an example. Fragments of ~1.7 and 2.7 kb were amplified by AK and the RSD-primers: NcoI-dep and EcoRI-dep (Fig. 2A-a and b), respectively, whereas ~0.3 and 1 kb were yielded from HindIII-dep (Fig. 2A-c). Five PCR parameters were tested for RSD-PCR protocol, and the best results were archived from: (i) 500 ng per 50 μ L reaction mixture of *T. maritima* genomic DNA in the 1st round PCR for target template amplification (Fig. 2A, lane 1–3); (ii) 10°C above T_m for T_a^{1st} (Fig. 2A, lane 4–7); (iii) 20 to 50 dilution times of target template (data not shown); (iv) 40°C to 48°C of T_a^{2nd} (Fig. 2A, lane 8–10); and (v) 4 μ M or higher RSD-primer concentration was required in the 2nd round PCR (Fig. 2A, lane 11–14).

The designated fragments in Fig. 2A were cloned and sequenced, proving the same sequence as the target genes published online (accession no. NC_000853). However, EcoRI-dep primer annealed to the *Eco*RI (GAATTC) similar site GAACTC in the target gene, whereas the other RSD-primers annealed to the corresponding restriction sites (Fig. 3). By using RSD-PCR

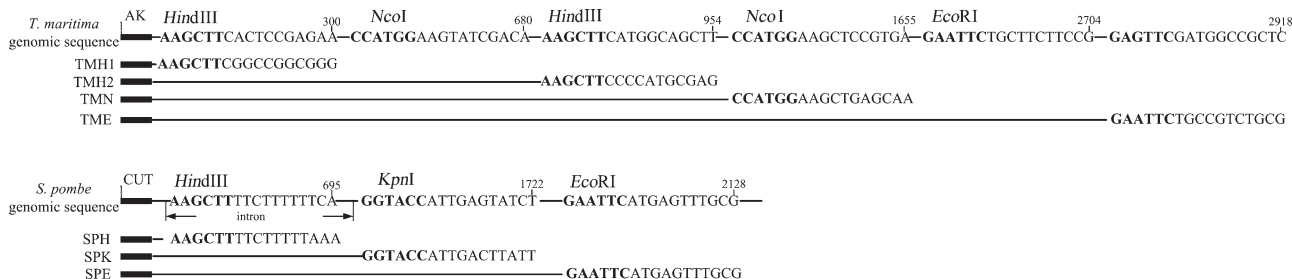


Figure 3. The sequenced fragments amplified by RSD-primers and the genomic templates. One clone from each primer set was picked out for sequencing.

protocol, *cut6* was amplified from a eukaryote genome of *S. pombe* by the specific primer CUT6 and a universal primer Hind-dep, EcoR-dep, or Kpn-dep (Fig. 2B-a). The sequencing results of the three bands showed that every PCR fragment sequence was consistent with online published sequence (accession no. Z99261), and RSD-primers annealed to the corresponding restriction sites (Fig. 3).

3.2. Cloning a new gene using degenerated primers

The 15 NH₂-terminal amino acids sequence was MNGTVKVNENIGRIS- for the purified arabinosidase of *B. pumilus* ARA. The degenerated primers were designed based on NGTVKV (Table 1, araN1, araN2, and araN3), and the occurrence of degeneration at 3rd base from 3' terminus was a key point for obtaining specific amplifications (Fig. 2B-b). When this position was not degenerated, the primer with a matched 3rd base (araN1) yielded the fragment of only one size (Fig. 2B-b, lane 4), and the primer with a mismatched 3rd base (araN2) yielded kinds of non-target fragments (Fig. 2B-b, lane 5); and when the primer containing both matched and mismatched 3rd bases (araN3), the PCR products showed all the contaminants (Fig. 2B-b, lane 6). The same phenomenon was observed when we clone a new gene from straw mushroom (data not shown). The fragments amplified by araN1 and Pst-dep in a single band (Fig. 2B-b, lane 4) was then sequenced, which approved to be a potential arabinosidase gene best homologous to the *abfB* from *B. subtilis* 168 of 67% identity by NCBI Blast analysis. The specific upwalking primer ara-up was then designed based on the sequence of cloned similar *abfB* above. The main fragment of 0.5 kb amplified by ara-up and Hind-dep (Fig. 2B-b, lane 7) from *B. pumilus* ARA was cloned for sequencing, and thus the complete similar *abfB* and the regutry nucleotide sequence of 2.2 kb fragment was obtained (accession no. DQ324528).

3.3. Studying the gene structure around *adhB* in *T. ethanolicus* JW200

A 1.2 kb *adhB* gene was amplified from genomic DNA of *T. ethanolicus* JW200 by using primer adh-N and

adh-C. On the basis of this sequence, the walking primers *adh-up* and *adh-down* were designed for cloning the flanking genes by RDS-PCR. The upstream walking product of 0.5 kb amplified by *adh-up* and EcoR-dep (Fig. 2B-c, lane 8) and downstream walking product of 0.9 kb amplified by *adh-down* and Nco-dep (Fig. 2B-c, lane 9) were cloned for sequencing, respectively. The potential promoter of *adhB* was identified 70–100 bp upstream the RBS. A truncated ORF upstream of but in the opposite orientation relative to *adhB* was translated into a polypeptide with amino acid sequence of 93% and 76% of identity with the *N*-acetylglutamate-gamma-semialdehyde dehydrogenase (*argC*) of *T. pseudoethanolicus* 39E and *T. tengcongensis* MB4. Further upstream walking was done by the RSD-PCR to obtain the complete ORF of *argC*. From downstream sequence, a truncated ORF identical to a similar polypeptide 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases located downstream of *T. pseudoethanolicus* 39E and *T. tengcongensis* MB4 *adhB* gene was found. Thus, the complete ORF of *adhB* and upstream *argC*, and a truncated downstream gene sequence of 3 kb from *T. ethanolicus* JW200 chromosomal DNA was obtained (accession no. DQ323135), and the gene structures were compared among these species in genus *Thermoanaerobacter* (Fig. 4).

3.4. Universal primers emerged from degenerated primers

Other than trying to reduce the degeneration according to the statistical reports about the codon usage in certain species, there is no direct evidence to indicate how to design a degenerated primer based on a few amino acid residues. The difference in the specificity of PCR amplifications by the generated primers for the arabinosidase gene showed that araN3, a mixture of complementary and non-complementary 3rd base from the 3'-end, would change PCR format rather than produce the mixed fragments from araN1 and araN2 (Fig. 2B-b, lanes 4–6). This result not only indicated the necessity to avoid degeneration in the last five bases at 3'-end of a specific primer, but also suggests that five or more complementary bases at 3'-end followed by a degenerated sequence facilitated selection for relatively specific

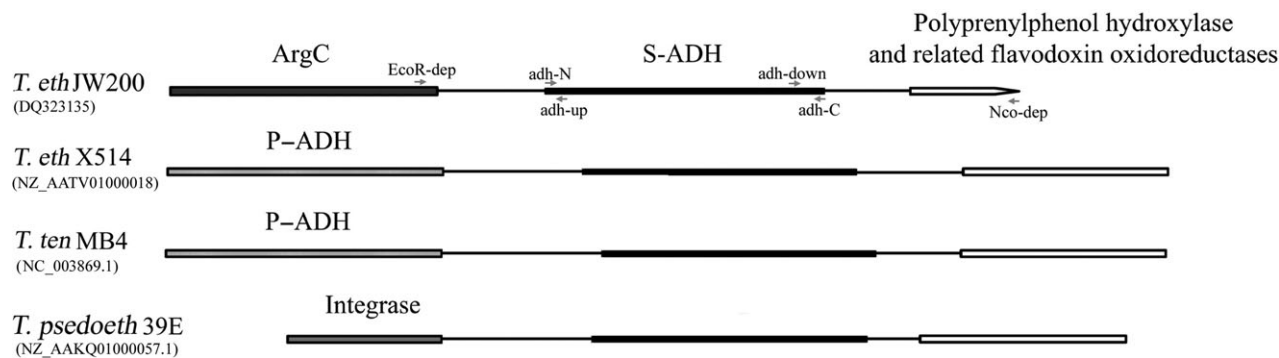


Figure 4. Comparison of the gene structures of *adhB* in *Thermoanaerobacter* spp. *T. eth* JW200, *T. ethanolicus* JW200; *T. eth* X514, *T. ethanolicus* X514; *T. pseudoeth* 39E, *T. pseudoethanolicus* 39E; *T. ten* MB4, *T. tengcongensis* MB4; ArgC, *N*-acetyl- γ -glutamyl-phosphate reductase (*argC*); S-ADH, secondary-alcohol dehydrogenase (*adhB*); P-ADH, primary-alcohol dehydrogenase (*adhA*). The numbers in the blankets indicate the corresponding GenBank accession numbers of the sequence. Arrows indicate the position and direction of the specific and universal primers used in the chromosome walking of *T. ethanolicus* JW200.

amplifications. Accordingly, the universal primers were designed as RSD-primers using six nucleotides (nt) recognized by a restriction enzyme and 11 degenerated bases, in which we thought that the addition of more degenerated bases could reduce the specificity. Theoretically, the G/C rich restriction sites are better choices for this purpose, although *Hind*III and *Eco*RI sites worked well in this work. The 3'-end can be any 5–8 bases without degeneration, but 6 nt restriction sites are easily recognized, whereas 4 nt restriction sites may occur too frequently in genomes to limit walking length. Thus, the specific primers were designed of the T_m ranged from 42°C to 48°C to pair with universal primers. Primers of higher T_m yielded contaminated PCR products when performed in the ideal annealing temperature based on universal primers which was 40 ~ 48°C as Fig. 2A, lane 8–10 indicated (data not shown). The uncertainty in the amplifications by using RSD-PCR protocol was mainly caused by the removal of some purine and pyrimidine bases from the degenerated 5'-end of RSD-primers because these removed bases might be complementary to the unknown end. It was necessary to lower the melting temperature of primer-dimers in this way because the binding of inosine to the regular bases gave insufficient strength to support annealing.

3.5. Overall strategy of RSD-PCR

RSD-PCR strategy is based on these principles: (i) restriction sites disperse throughout the genomes of double strand DNA either in prokaryotic or eukaryotic organisms are natural candidacy for universal pairing; (ii) a universal primer can be a combination of a 3'-end pairing with the bases of selected restriction sites in genomic DNA, and a 5'-end of degenerated sequence to support annealing; and (iii) to limit non-specific amplifications, genomic DNA can be greatly diluted after the target template is amplified by a specific primer. Thus,

two-round PCR amplifications were employed in the RSD-PCR protocol: amplify the target template from genomic DNA in the 1st round PCR by using a specific primer to accumulate the specific single strand template; and amplify the target gene in the 2nd round by using the specific primer and one of the universal primers. The protocol is especially useful in new gene cloning using a degenerate primer based on the amino acids sequence or conserved region without nested PCR, traditional library construction and screening process. Except for the genes listed in the article, RSD-PCR has been successfully applied in chromosome walking in *Candida* sp., *Bacillus* sp., and *Alcaligenes* sp. However, difficulties were found in cloning the new genes from the chromosome DNA of straw mushroom by RSD-PCR as well as cloning cDNA by RT-PCR, which might be caused by the complexity of the organism. Two-round PCR was previously reported, which enriched single-stranded DNA sequences either upstream or downstream from the known sequences before second round amplification.²⁰ This method increased final specificity, but the following random primer amplification still created non-target molecules. Thus, more PCR methods for gene walking were later reported using nested amplification to increase specificity.^{19,21,25}

3.6. The practical performance of the RSD-PCR protocol

The specificity of the final products was strongly controlled by the first single strand amplification. Although a standard PCR can yield specific products with a T_a range of 10 ~ 15°C,²⁷ the T_a ^{1st} in RSD-PCR was recommended for ~10°C above T_m . The amount of chromosome DNA depended on the genome size of the organism, and ~0.5 and 3 μ g was recommended for bacteria and yeasts. Meanwhile, the specific primer without degeneration should be reduced to one-tenth of the normal

concentration (0.02 μM) because it was a non-exponential amplification, but the concentration of degenerated specific primer should be varied with the degeneracy, e.g. 0.2 μM of degenerated specific primer yielded positive results in the cloning of *abfB* from *B. pumilus* ARA (Fig. 2B-b).

In the 2nd round PCR, there are three factors that need to be dealt with: (i) the existence of restriction sites in the single strand templates, (ii) the unknown sequence around the restriction sites, and (iii) the property of the specific primer. Because the target template from the 1st round is a new sequence, we do not know which restriction sites are at a suitable distance for second round amplification, a set of RSD-primers were synthesized, and each was performed with the specific primer for locating the target gene. Meanwhile, the number of matched bases in the region paired by the degenerated part of an RSD-primer is not predictable, and the binding between RSD-primer and chromosome DNA varies from time to time. This can explain why the effects of the primer concentration and annealing temperature varied with the type of RSD-primer used in amplification of *ak* from *T. maritima* (Fig. 2A). From the sequences of the cloned fragments, we can see that mismatches frequently occur in the 5'-end, and occasionally occur in the restriction site at 3'-end of RSD-primer (Fig. 3, TME). However, these mismatches did not affect amplification efficiency, indicating the importance of the specific primer in the selectivity of the amplification. The strength and specificity of the amplifications depend on the property of the specific primer. When the specific primer is not sufficiently unique, it can anneal to undesired site(s), and give several possibilities, among which the acceptable result is the single primer amplification as previously reported,²⁸ and a clear mixture of single and double primer amplifications as shown in Fig. 2B-a (*S. pombe*). In a few cases, a specific primer annealed to wrong site(s) instead of the target DNA resulting in non-target products (data not shown), and a new primer should be synthesized in the case.

3.7. Identification of the target gene from the PCR products

A stem-loop structure is occasionally formed following every cycle by single primer amplification products, because of the inverted terminal repeats (Fig. 1). Stem-loop structures are more stable than the primer-template hybrid, and therefore suppress exponential amplification.^{16,17} However, single primer amplification product still occurred in *cut6* amplification from *S. pombe* as shown in Fig. 2B-a, the largest fragments of the same length in lanes 1–3 as PCR products amplified by single primer CUT (lane 0). This non-target molecule amplification was greatly inhibited by decrease of annealing temperature to 42°C (data not shown). The reason for this is the

equilibrium between productive PCR primer annealing and non-productive self-annealing of the fragment's complementary ends, which arises within each PCR cycle at the primer annealing stage.²⁹ This critical temperature differs from the templates and specific primers. However, in most cases, the same size fragments appeared in different RSD-primer amplifications are treated as PCR backgrounds of non-target molecules. They can be easily distinguished from target fragments which usually present different sizes of different RSD-primer amplification (Fig. 2B-a). The main RSD-PCR product amplified by AK and Nco-dep only yielded one of the two calculated fragments of 1.6 kb (Figs 2A-a and 3), and a putative 0.6 kb fragment was missing, which may be because of *ex Taq* DNA polymerase amplification preference of length (Takara Bio Catalog) and RSD-primer anneal preference at the annealing temperature based on the template and primer sequences.

3.8. Advantages and applications of RSD-PCR

RSD-PCR has the advantages as follows: (i) chromosomal DNA is directly used as template in the 1st round PCR, and thus some inefficient steps such as digestion, and ligation are avoided; (ii) the amplification and the dilution of the single strand target template in the 1st round reduce the disturbance of non-target DNA, and makes it possible to employ a universal primer to the unknown end; (iii) the stringent conditions for both the 1st and the 2nd round PCR reduces the non-specific annealing of the primers to chromosomal DNA, and thus strongly controls the specificity of the amplifications; (iv) the RSD-PCR protocol does not need nested PCR selection, and becomes a useful technique for gene walking using degenerated primers based on NH₂-terminal or conserved amino acids sequences. RSD-PCR can be applied in the molecular cloning of (i) new genes from degenerated primers derived from NH₂-terminal amino acids sequences of proteins, or from conserved amino acids sequences in various enzymes; (ii) signal peptide or propeptide genes for secretory proteins or mature proteins; (iii) regulatory sequence of a gene or gene cluster; (iv) the full sequences of operons or their neighboring genes.

Funding

National Key Basic Science "973" program of China (2004CB719600); National Natural Science Foundation of China (30170511 and 30370034).

References

1. Triglia, T., Peter, M. G. and Kemp, D. J. 1988, A procedure for in vitro amplification of DNA segments that lie outside

- the boundaries of known sequences, *Nucleic Acids Res.*, **16**, 8186.
2. Ochman, H., Gerber, A. S. and Hartl, D. L. 1988, Genetic application of an inverse polymerase chain reaction, *Genetics*, **120**, 621–623.
 3. Silverman, G. A., Ye, R. D., Pollock, K. M., Sadler, J. E. and Korsmeyer, S. J. 1989, Use of the yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3, *Proc. Natl. Acad. Sci. USA*, **86**, 7485–7489.
 4. Keim, M., Williams, R. S. and Harwood, A. J. 2004, An inverse PCR technique to rapidly isolate the flanking DNA of dictyostelium insertion mutants, *Mol. Biotechnol.*, **26**, 221–224.
 5. Mueller, P. R. and Wold, B. 1989, In vivo footprinting of a muscle specific enhancer by ligation mediated PCR, *Science*, **246**, 780–786.
 6. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D. and Powell, S. 1990, A novel, rapid method for isolation of terminal sequences from yeast artificial chromosome (YAC) clones, *Nucleic Acid Res.*, **18**, 2887–2890.
 7. Rosenthal, A. and Jones, D. S. C. 1990, Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction, *Nucleic Acids Res.*, **18**, 3095–3096.
 8. Rosenthal, A., Mackinnon, R. N. and Jones, D. S. C. 1991, PCR walking from microdissection clone M54 identifies three exons from the human gene for the neural cell adhesion molecule L1 (CAM-L1), *Nucleic Acids Res.*, **19**, 5395–5401.
 9. Espelund, M. and Jackobsen, K. S. 1992, Cloning and direct sequencing of plant promoters using primer adapter mediated PCR on DNA coupled to a magnetic solid phase, *Biotechniques*, **13**, 74–81.
 10. Jones, D. H. and Winistorfer, S. C. 1992, Sequencing specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA, *Nucleic Acids Res.*, **20**, 595–600.
 11. Warshawsky, D. and Miller, L. A. 1994, A rapid genomic walking technique based on ligation-mediated PCR and magnetic separation technology, *Biotechniques*, **16**, 792–798.
 12. Sterky, F., Holmberg, A., Alexandersson, G., Lundberg, J. and Uhlen, M. 1998, Direct sequencing of bacterial artificial chromosome (BACs) and prokaryotic genomes by biotin-capture PCR, *J. Biotechnol.*, **246**, 810–813.
 13. Dai, S. M., Chen, H. H., Chang, C., Riggs, A. D. and Flanagan, S. D. 2000, Ligation-mediate PCR for quantitative in vivo footprinting, *Nat. Biotechnol.*, **18**, 1108–1111.
 14. Zhang, Z. and Gurr, S. J. 2002, A ‘step down’ PCR-based technique for walking into and the subsequent direct-sequence analysis of flanking genomic DNA, *Methods Mol. Biol.*, **192**, 343–350.
 15. Yan, Y., An, C., Li, L., Gu, J., Tan, G. and Chen, Z. 2003, T-lingker-specific ligation PCR (T-lingker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends, *Nucleic Acids Res.*, **31**, e68.
 16. Matz, M., Shagin, D., Bogdanova, E., et al. 1999, Amplification of cDNA ends based on template-switching effect and step-out PCR, *Nucleic Acids Res.*, **27**, 1558–1560.
 17. Chenchik, A., Diachenko, L., Moqadam, F., Tarabykin, V., Lukyanov, S. and Siebert, P. D. 1996, Full-length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA, *Biotechniques*, **21**, 526–534.
 18. Parker, J. D., Rabinovitch, P. S. and Burner, G. C. 1991, Targeted gene walking polymerase chain reaction, *Nucleic Acids Res.*, **19**, 3055–3060.
 19. Liu, Y., Mitsukawa, N., Oosumi, T. and Whittier, R. F. 1995, Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric inter-laced PCR, *Plant J.*, **8**, 457–463.
 20. Trueba, G. A. and Johnson, R. C. 1996, Random primed gene walking PCR: simple procedure to retrieve nucleotide fragments adjacent to known DNA sequences, *Biotechniques*, **21**, 20.
 21. Liu, Y. and Whitter, R. F. 1995, Thermal asymmetric inter-laced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking, *Genomics*, **25**, 674–681.
 22. Weigel, D., Ahn, J. H., Blazquez, M. A., et al. 2000, Activation tagging in *Arabidopsis*, *Plant Physiol.*, **122**, 1003–1013.
 23. Singer, T. and Burke, E. 2003, High-throughput TAIL-PCR as a tool to identify DNA flanking insertions, *Methods Mol. Biol.*, **236**, 241–272.
 24. Antal, Z., Rasclé, C., Fevre, M. and Bruel, C. 2004, Single oligonucleotide nested PCR: a rapid method for the isolation of genes and their flanking regions from expressed sequence tags, *Curr. Genet.*, **46**, 240–246.
 25. Tan, G., Gao, Y., Shi, M., et al. 2005, SiteFinding-PCR: a simple and efficient PCR method for chromosome walking, *Nucleic Acids Res.*, **33**, e122.
 26. Zeng, H., Xue, Y., Peng, T. and Shao, W. 2007, Properties of xylanolytic enzyme system in bifidobacteria and their effects on the utilization of xylooligosaccharides, *Food Chem.*, **101**, 1172–1177.
 27. Dominguez, O. and Lopez-Larrea, C. 1994, Gene walking by unpredictably primed PCR, *Nucleic Acids Res.*, **22**, 3247–3248.
 28. Parks, C. L., Chang, L. S. and Shenk, T. A. 1991, polymerase chain reaction mediated by a single primer: cloning of genomic sequences adjacent to a serotonin receptor protein coding region, *Nucleic Acids Res.*, **19**, 7155–7160.
 29. Lukyanov, K. A., Gurskaya, N. G., Bogdanova, E. A. and Lukyanov, S. A. 1999, Selective suppression of polymerase chain reaction, *Bioorganicheskaya Khimiya*, **25**, 163–170.