Gene 423 (2008) 92–95

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Methods paper

Reducing cloning artifacts for recovery of allelic sequences by T7 endonuclease I cleavage and single re-extension of PCR products — A benchmark

Kenji Saitoh [a,*], Wei-Jen Chen [b]

[a] Tohoku National Fisheries Research Institute, Shinhama, Shiogama 985-0001 Japan
[b] Saint Louis University, Department of Biology, 3507 Laclede Ave, St. Louis, MO 63103-2010, USA

ARTICLE INFO

Article history:
Received 7 January 2008
Received in revised form 5 June 2008
Accepted 23 June 2008
Available online 2 July 2008

Received by M. D'Urso

Keywords:
Heteroduplex
Template switching
Mismatch repair
Recombinant
Chimera
Nuclear gene
TA cloning

ABSTRACT

Occurrence of chimeric sequences and related artifacts in PCR cloning procedures gives us risks of over-estimation of haplotypes or alleles. Recombination among haplotypes occurs through template switching during PCR cycles or through random repair of mismatch sites on heteroduplex DNA by the host cell. To eliminate the chimeric cloning artifacts, we tested two alternative protocols using T7 endonuclease I cleavage of mismatch sites and re-extension of nascent strands. Though T7 endonuclease I effectively eliminated chimeric clones in some cases, it produced many short fragments. Protocol with single re-extension of PCR products successfully recovered non-recombinant clones with fewer short fragments. In spite of the improvement of allelic recovery through these two protocols, there were still a few recombinants that remained in both reaction mixtures, and thus interpretation of the results for haplotype diversity in a PCR-amplified DNA population should be cautionary. Because re-extension in a diluted reaction mixture is quick, inexpensive and effective, it is advisable to use this procedure for recovery of chromosomal alleles with PCR cloning.

1. Introduction

With advances in molecular techniques, DNA sequences from nuclear or cytoplasmic gene loci have been the tools of choice for various studies in evolutionary biology and ecology. For such studies, cloning PCR products is an efficient method to search for alleles, paralogous loci, sequence heterogeneity in DNA populations, or hidden organismal communities (Pendleton et al., 1993; Morrel et al., 2003; Acinas et al., 2004; Shigenobu et al., 2005). A number of clones are sequenced to sort out individual allelic sequences in a PCR-amplified DNA population. In addition to the real template multiplicity, variants can be produced through nucleotide mis-incorporation and template switching during PCR cycles (Odelberg et al., 1995; Kanagawa, 2003). Host cell DNA repair system can also produce variants. When the host cell assimilates a plasmid containing heteroduplexes, a mismatch repair system works randomly utilizing either strand as template and thus makes chimeric sequences (Speksnijder et al., 2001; Longeri et al., 2002; Kanagawa, 2003; Beser et al., 2007). These PCR or cloning errors give rise to risks on over-estimation of number of haplotypes or alleles. To eliminate such PCR and cloning artifacts some authors have proposed a number of protocols (Smith and Modrich, 1997; Lowell and Klein, 2000; Thompson et al., 2002; Kanagawa, 2003). We present here a benchmark test in order to assess the efficiency for eliminating the sequence artifacts produced during PCR or cloning from two alternative methods: T7 endonuclease I cleavage (T7 cleavage) and re-extension of nascent strands (re-extension). We also compared the results from two different PCR conditions (stringent and relaxed conditions) which differ in their number of thermocycles and the annealing temperatures.

2. Materials and methods

2.1. Specimens

We used two tetraploid loaches of the genus *Cobitis* (*Cobitis striata* large race [specimen #1] and an unnamed form of the 'yamato' complex [#2]) (Saitoh et al., 2000; Saitoh, 2003). About 1500 bp region of exon-3 of the recombination activating gene 1 (RAG1) was the target of this study. Specimen #1 has four alleles including relatively divergent (3.8% p-distance) ones, while #2 has two close alleles (0.9% p-distance). Tetraploids offer a good model system of template multiplicity and complexity of natural occurrence with at most four alleles showing various extent of divergence.

Abbreviation: nt, nucleotide.

* Corresponding author. Tel.: +81 22 365 9932; fax: +81 22 367 1250.
E-mail address: ksaitoh@affrc.go.jp (K. Saitoh).

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.06.025

2.2. PCR, enzyme treatments and cloning

PCR primers (R1_2533F; CTGAGCTGCAGTCAGTACCATAAGATGT and R1_4078R; TGAGCCTCCATGAACTTCTGAAGRTAYTT) worked for a wide spectrum of teleostean species (López et al., 2004; Chen et al., in press), and we expect all the alleles could be amplified. PCR reaction mixtures of 15 μl in total volume contained about 5–50 ng of genomic DNAs, 5 pmol each of primers, 0.2 mM each of dNTPs, 0.375 U of Ex*Taq* DNA polymerase (Takara, Shiga, Japan) in the 1× Ex*Taq* reaction buffer. PCR reaction started with an initial denaturation step at 95°C for 3 min, followed by 30 cycles (a stringent condition) or 35 cycles (a relaxed condition to maximize yields) of denaturation at 94°C for 30 s, annealing at 55°C (the stringent condition) or 52°C (the relaxed condition) for 30 s, and 72°C for 2 min. The reaction stopped after a final extension step at 72°C for 30 min. The yields of the PCR reaction were approximately determined on an agarose gel comparing fluorescent intensities. We first sequenced the PCR products directly to assess heterogametic nucleotide sites, and then cloned them into *Escherichia coli* DH5α prepared according to Inoue et al. (1990).

For direct cloning of the PCR products without enzyme treatment, 0.2–0.3 μl (diluted with milli-Q water) of the PCR products received ligation reaction with 12.5–25 ng of pGEM-t vector (Promega, WI, USA) in a 5 μl volume for overnight at 15°C.

For the T7 cleavage, ca. 10 ng of the PCR products (0.025–0.1 μl depending on yields) received digestion by 5 U of the enzyme (New England Biolabs, MA, USA) in 10 μl reaction mixtures at 55°C for 1 h. The reaction stopped with addition of 1 μl of 2 mg/ml proteinase-K (Nacalai, Kyoto, Japan) at 37°C for 30 min and subsequently at 80°C for 15 min.

For the re-extension, we set up 10 μl of new PCR reaction mixtures containing about 10 ng of the PCR products (0.025–0.1 μl), the primer set, dNTPs, buffer and the DNA polymerase as described above. The new reaction mixtures underwent the following thermo-steps: a denaturation at 95°C for 5 min, annealing at 55°C for 30 s and an extension at 72°C for 30 min.

We added 1 μl of these enzyme-treated DNA solutions to ligation reaction mixture of 4 μl containing 5 ng of pGEM-t (5 μl in total) and incubated for 6 h or overnight at 15°C. Ligation mixtures accepted equal volume of TE buffer (10 mM tris, 1 mM EDTA, pH8) to stop reaction, and 4 μl of the solution was transformed to 50 μl of competent DH5α as usual (Inoue et al., 1990).

2.3. Sequencing and assembling

We picked 16 to 32 white colonies up from each transformant to PCR-amplify inserts for sequencing. The same PCR primers, M13 primers and two internal primers (R1_3236F; GAAGATGGTACGTGAAATGGA and R1_3427R; CTCCATGAAAGGCTTGGCAGA) worked for sequencing with BigDye v.3.1 (Applied Biosystems, CA, USA) or ET (GE Healthcare, Buckinghamshire, England) terminator sequencing kits on an ABI 3100 sequencer.

We ignored nt sites where variant nt appeared in a single clone (private sites) to sort out consensus sequences in assemblies. We also excluded private clones with unique nt combinations of variable sites from assemblies. Errors such as mutations or recombinations during PCR and cloning procedures might occur randomly, and this randomness makes error sites or nt combinations singletons that we excluded. When two or more clones showed an identical nt combinations of variable sites except for the private sites, we recognized them as a putative allele.

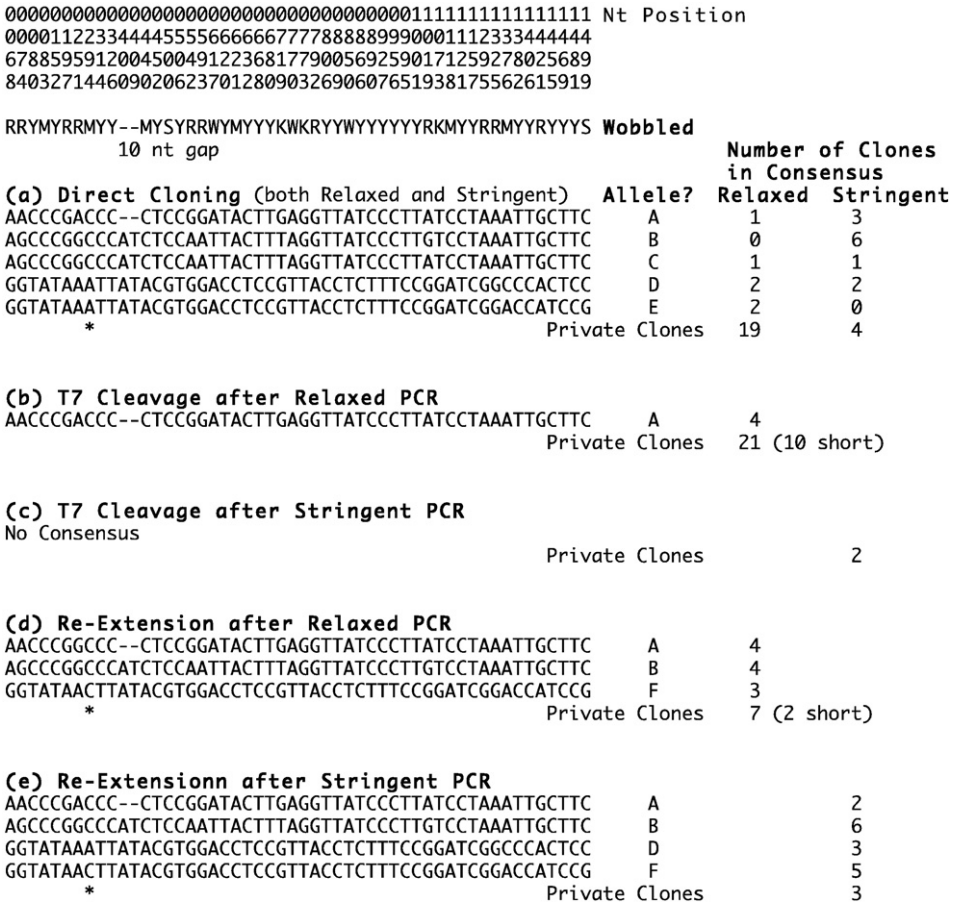

Fig. 1. Clones and alleles recognized among colonies with various methods of cloning of the specimen #1 (*C. striata* large race). Hyphens show gap sites from 400–409. Wobbled sequence pattern appeared in the PCR-direct sequencing charts. There was one mismatch between putative alleles (E and F) recognized in the products of direct cloning and enzyme-treated cloning (asterisk). Private sites are omitted.

3. Results and discussion

3.1. Direct cloning

Among 32 white colonies picked up under the relaxed PCR condition, 25 and 28 colonies contained PCR insert for the specimen #1 and #2 respectively. The sequencing charts of 25 clones (Fig. 1a) as well as of PCR-direct sequencing from the specimen #1 showed existence of two allele groups distant from each other. In this case, however, we were able to recognize only two putative alleles (D and E) close to each other (0.3% p-distance). No distant allele was recognizable. There were many (19 out of 25 clones) private clones. For the case with specimen #2, we were able to recognize four putative alleles (Fig. 2a) from 28 clones. A few (4 out of 28 clones) private clones were still recovered.

Regarding to the experiment under the stringent PCR condition for the specimen #1, all of 16 white colonies picked contained insert. It also gave a better result with smaller number of private clones (4 out of 16 clones) (Fig. 1a). Three more putative alleles can henceforth be assigned (A, B and C). However, five alleles totally recognized from a tetraploid animal, which must has a maximum of four alleles at a locus, was unrealistic.

3.2. Enzyme treatments after PCR

Enzyme treatment in two ways yielded better results than the direct cloning in many cases (Figs. 1e, 2b–e). More than 100 (up to about 1000) white colonies appeared from the small amount of ligated DNAs. For specimen #1, clones from the product of stringent PCR condition with re-extension determined unambiguously four alleles with fewer private clones (Fig. 1e). An extra putative allele (F) was found. The sequence of allele-F had one nt mismatch (C to A) with allele-E that was obtained from the direct cloning (Fig. 1a). The allele-F sequence repeatedly

```
00000000000001 Nt Position
00111345688890
78177063200854
40669869069460

RMRYYRSYYYYAYY Wobbled

                                Number of Clones
(a) Direct Cloning   Allele?  in Consensus
ACGCCGCCTTTCCC          A          4
GAATTAGTCCCGTT          B         16
ACGCCGCCTCCCCC          C          2
GAATTAGTCTTGTT          D          2
       Private Clones              4


(b) T7 Cleavage after Relaxed PCR
ACGCCGCCTTTCCC          A          4
GAATTAGTCCCGTT          B          8
       Private Clones              8 (6 short)


(c) T7 Cleavage after Stringent PCR
ACGCCGCCTTTCCC          A          2
GAATTAGTCCCGTT          B          7
       Private Clones              2 (2 short)


(d) Re-Extension after Relaxed PCR
ACGCCGCCTTTCCC          A         10
GAATTAGTCCCGTT          B          7
       Private Clones              2 (1 short)


(e) Re-Extension after Stringent PCR
ACGCCGCCTTTCCC          A         11
GAATTAGTCCCGTT          B         12
       Private Clones              2 (1 short)
```

Fig. 2. Clones and alleles recognized among colonies with various methods of cloning of the specimen #2 (*Cobitis* 'yamato' complex). Wobbled sequence pattern appeared in the direct sequencing charts. Private sites are omitted.

appeared in other rounds of cloning from independent PCR tubes as private clones (in direct cloning of the stringent PCR condition and relaxed PCR+T7 cleavage) giving 10 clones in total among 115 clones throughout for the specimen #1, while allele-E appeared only twice (Fig. 1a). Therefore, the allele-E is most likely an artifactual variant of allele-F for the specimen #1. Also, allele-C from the specimen #1 is most likely of artifact, because it was absent from all the 74 clones of enzyme-treated products. Similarly, the allele-C and D for the specimen #2 are most likely of artifact (Fig. 2a) since both of them were absent from all of 75 clones from the enzyme-treated products (Fig. 2b–e). From these results, we can identify four alleles (A, B, D, and F) including relatively divergent (3.8% p-distance) ones from the specimen #1 and two close alleles (0.9% p-distance) (A and B) from #2.

Enzyme treatments, however, did not always give good resolution. PCR products with the relaxed condition brought about many private clones even after the enzyme treatments (Figs. 1b,d, 2b). This tendency is noticeable in the specimen #1 in which the genetic distances between alleles are large. Template switching in the later PCR cycles (Odelberg et al., 1995; Kanagawa, 2003) might be responsible for this tendency rather than heteroduplex formation. T7 cleavage was prone to produce shorter inserts or clones without detectable insert. In an extreme case, there were only two clones with insert among 32 white colonies from product of stringent PCR with T7 cleavage (Fig. 1c). Gel slice to size-fractionate inserts before ligation might be helpful in this case, but it is time consuming.

3.3. Concluding remarks

In our benchmark test, re-extension of PCR products with a higher annealing temperature and fewer PCR cycles yielded the best results for both specimens. The annealing temperature and number of thermocycles tested here are routinely applied in many laboratories. Further optimization to reduce thermocycles with quantitative PCR to avoid heteroduplex formation (Kanagawa, 2003) is unnecessary. Re-extension procedure need neither gel slice nor buffer exchange between experimental steps, and it is less time consuming. It does not need any additional enzymes (e.g. T7 endonuclease I) and special equipments. It also turned out that only one re-extension is enough to obtain colonies necessary for determining alleles. Single re-extension is better than a few cycles of extensions (Thompson et al., 2002) which may introduce a chance of heteroduplex formation again.

Both re-extension and T7 cleavage, however, did not completely eliminate private clones or possible recombinants, and thus interpretation of the results for haplotype diversity in a PCR-amplified DNA population should be still cautionary. It is especially the case in search for organismal community from environmental samples or for paralogous loci of gene families, because the maximum number of haplotypes is unknown.

We detected two and four alleles for RAG1 from allotetraploid fish derived from hybridization between distantly related species (Saitoh et al., 2000; Saitoh, 2003). A gap in a coding region from specimen #1 (allele-A) and absence of distant allele from specimen #2 indicate genome re-shaping such as gene silencing after polyploidization. PCR cloning turned out to be an efficient method to detect such modes of diploidization in tetraploid animals. Because of quickness, economy and efficiency, it is advisable to use the re-extension procedure for recovery of chromosomal alleles with PCR cloning.

Acknowledgments

This study was partly supported by Grants-in-Aid from the Japan Society for Promotion of Science (JSPS 17207007) and the National Science Foundation, US (EF-0431326). We thank Richard L. Mayden in Saint Louis University for giving the opportunity to perform this collaborative work. DDBJ/Genbank entries of the sequence data obtained in this study are as follows; AB373204-AB373209. This

paper is a scientific contribution from the Fisheries Research Agency, Japan under contribution number FRA-TNFRI-B105.

References

Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., Polz, M.F., 2004. Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430, 551–554.

Beser, J., Hagblom, P., Fernandez, V., 2007. Frequent in vitro recombination in internal transcribed spacers 1 and 2 during genotyping of *Pneumocystis jirovecii*. J. Clin. Microbiol. 45, 881–886.

Chen, W.-J., Miya, M., Saitoh, K., Mayden, R.L., in press. Phylogenetic utility of two existing and four novel nuclear loci in reconstructing tree of life of ray-finned fishes: the order Cypriniformes (Ostariophysi) as a case study. Gene. doi:10.1016/j.gene.2008.07.016.

Inoue, H., Nojima, H., Okayama, H., 1990. High efficiency transformation of *Escherichia coli* with plasmids. Gene 96, 23–28.

Kanagawa, T., 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J. Biosci. Bioeng. 96, 317–323.

Longeri, M., Zanotti, M., Damianti, G., 2002. Recombinant DRB sequences produced by mismatch repair of heteroduplexes during cloning in *Escherichia coli*. Eur. J. Immunogenet. 29, 517–523.

López, J.A., Chen, W.-J., Ortí, G., 2004. Esociform phylogeny. Copeia 2004, 449–464.

Lowell, J.L., Klein, D.A., 2000. Heteroduplex resolution using T7 endonuclease I in microbial community analyses. BioTechniques 28, 676–681.

Morrel, P.L., Lundy, K.E., Clegg, M.T., 2003. Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. Proc. Natl. Acad. Sci. U. S. A. 100, 10812–10817.

Odelberg, S.J., Weiss, R.B., Hata, A., White, R., 1995. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. Nucleic Acids Res. 23, 2049–2057.

Pendleton, J.W., Nagai, B.K., Murtha, M.T., Ruddle, F.H., 1993. Expansion of the *Hox* gene family and the evolution of chordates. Proc. Natl. Acad. Sci. U. S. A. 90, 6300–6304.

Saitoh, K., 2003. Mitotic and meiotic analyses of the 'large race' of *Cobitis striata*, a polyploid spined loach of hybrid origin. Folia Biol. (Krakow) 51 (Suppl.), 101–105.

Saitoh, K., Kobayashi, T., Ueshima, R., Numachi, K., 2000. Analyses of mitochondrial and satellite DNAs on spined loaches of the genus *Cobitis* from Japan have revealed relationships among populations of three diploid–tetraploid complexes. Folia Zool. 49 (S1), 9–16.

Shigenobu, Y., Saitoh, K., Hayashizaki, K., Ida, H., 2005. Nonsynonymous site heteroplasmy in fish mitochondrial DNA. Genes Genet. Syst. 80, 297–301.

Smith, J., Modrich, P., 1997. Removal of polymerase-produced mutant sequences from PCR products. Proc. Natl. Acad. Sci. U. S. A. 94, 6847–6850.

Speksnijder, A.G.C.L., Kowalchuk, G.A., De Jong, S., Kline, E., Stephen, J.R., Laanbroek, H.J., 2001. Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequence. Appl. Environ. Microbiol. 67, 469–472.

Thompson, J.R., Marcelino, L.A., Polz, M.F., 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucleic Acids Res. 30, 2083–2088.