Journal of Biochemistry and Molecular Biology, Vol. 40, No. 2, March 2007, pp. 172-179

JOURNAL OF Biochemistry and Molecular Biology

PCR-mediated Recombination of the Amplification Products of the *Hibiscus tiliaceus* Cytosolic Glyceraldehyde-3-phosphate Dehydrogenase Gene

Linghui Wu, Tian Tang, Renchao Zhou and Suhua Shi*

State Key Laboratory of Biocontrol and Key Laboratory of Gene Engineering of the Ministry of Education, School of Life Sciences, Sun Yat-Sen University, 510275 Guangzhou, China

Received 25 February 2006, Accepted 25 September 2006

PCR-mediated recombination describes the process of *in vitro* chimera formation from related template sequences present in a single PCR amplification. The high levels of genetic redundancy in eukaryotic genomes should make recombination artifacts occur readily. However, few evolutionary biologists adequately consider this phenomenon when studying gene lineages. The cytosolic glyceraldehyde-3-phosphate dehydrogenase gene (*GapC*), which encodes a NADP-dependent nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase in the cytosol, is a classical low-copy nuclear gene marker and is commonly used in molecular evolutionary studies. Here, we report on the occurrence of PCR-mediated recombination in the *GapC* gene family of *Hibiscus tiliaceus*. The study suggests that recombinant areas appear to be correlated with DNA template secondary structures. Our observations highlight that recombination artifacts should be considered when studying specific and allelic phylogenies. The authors suggest that nested PCR be used to suppress PCR-mediated recombination.

Keywords: Cytosolic glyceraldehyde-3-phosphate dehydrogenase gene (*GapC*), Evolutionary study, *Hibiscus tiliaceus*, Nuclear gene, PCR-mediated recombination, Secondary structure

Introduction

PCR-mediated recombination describes the process of *in vitro* chimera formation from related DNA template sequences present in a single PCR amplification. It arises primarily due to the presence of incompletely extended PCR products, or via polymerase template switching in the absence of temperature cycling (Shammas *et al*., 2001). This phenomenon was initially identified in multi-gene family (Scharf *et al*., 1988), and has since been demonstrated for nature templates from heterozygous *Adh* loci (Bradley *et al*., 1997) and polyploid cotton (Cronn *et al*., 2002). PCR-mediated recombination has greatly promoted the development of *in vitro* evolution (Crameri *et al*., 1998) and provides a useful means of constructing recombinant molecules (Fang *et al*., 1999), but unrecognized artifactually produced recombinant products may bias results or alter interpretations in phylogeny and evolutionary studies (Cronn *et al*., 2002).

There is an increasing need to use nuclear genes in evolutionary studies to test sophisticated historical models. However, as nuclear genes are often present as gene families, the potential exist to amplify paralogs related by gene duplications and to generate recombinant amplification products mediated by PCR. While the challenge presented by paralogs has been widely recognized, many evolutionary biologists are inadequately aware of the problem of PCR-mediated recombination.

There are three distinct isoforms of glyceraldehydes 3-phosphate dehydrogenases (GAPDH) in higher plants, and one of these, cytosolic glyceraldehyde-3-phosphate dehydrogenase, is a NADP-dependent nonphosphorylating GAPDH which catalyzes the oxidation of glyceraldehyde 3-phosphate to 3-phosphoglycerate in the cytosol (Cerff, 1982). The cytosolic glyceraldehyde-3-phosphate dehydrogenase gene (*GapC*) has been successfully used as a classical low-copy nuclear gene maker in molecular evolutionary studies (Olsen and Schaal, 1999; Pérusse and Schoen, 2004). Moreover, it often appears as gene family in many plant species, such as, *Zea mays* (Russell and Sache, 1989; Manjunath and Sache, 1997), *Amsinckia spectabilis* (Pérusse and Schoen, 2004), and *Arabidopsis thaliana* (Marri *et al*., 2005). However, no study has addressed PCR-mediated recombination of the *GapC* gene or the potential risks of using such artifactual recombinants in evolutionary studies.

Here, the authors demonstrate the occurrence of PCR-mediated recombination in the *GapC* gene family of *Hibiscus*

*To whom correspondence should be addressed.
Tel: 86-20-84113677; Fax: 86-20-34022356
E-mail: lssssh@mail.sysu.edu.cn

tiliaceus, and report that recombinant areas appear to be correlated with DNA template secondary structures. Nested PCR is advocated as a means of suppressing PCR-mediated recombination.

Materials and Methods

Plant materials. Leaves of one cultivated *Hibiscus tiliaceus* individual from Guangzhou, China (denoted hereafter as CG) and of one wild *H. tiliaceus* individual from Ryukyu, Japan (denoted hereafter as JR) were sampled, and stored with silica gel in zip-lock plastic bags until required.

DNA isolation, PCR amplification, cloning and sequencing. Total DNA was extracted using the Doyle and Doyle method (Doyle and Doyle, 1987). Exon-primed intron-crossing (EPIC) amplifications of *GapC* were performed using the universal primer set (GPDX7F and GPDX9R) designed by Strand *et al.* (1997) (Table 1), using the PCR system of Olsen and Schaal (Olsen and Schaal, 1999) in the presence of a negative control. Amplification products were visually examined after 1% agarose gel electrophoresis. Three bands of amplification products were observed for each individual and each of these bands were spliced from agarose gel and purified using a QIAquick PCR purification kit (QIAGEN). The purified PCR products so obtained were ligated to pGEM-T Easy Vectors (Promega). Subcloning, microbial cultures and plasmid isolation were conducted according to *Molecular Cloning* (Sambrook *et al*., 2001). Sixteen clones from CG and fifteen clones from JR were randomly selected and the corresponding bands were purified, i.e., in total 48 clones for CG and 45 clones for JR were analyzed, respectively. All purified products were sequenced on an ABI PRISM™ 3730 sequencer using BigDye Terminator V3.0 Cycle Sequencing kits. Sequences have been deposited in GenBank under accession no. DQ299458-299487.

Sequence analyses. Sequences were assembled and edited using SeqMan5.05 (DNASTAR Inc.), and proofread by eye. The identities of the DNA sequences were obtained using blastn in GenBank (Altschul *et al*., 1997). The boundaries of exons/introns in the obtained sequences were determined using FSPLICE (Softberry Inc.) combined with BLAST (blastn, tblastx and bl2seq) (Altschul *et al*., 1997; Tatusova and Madden, 1999). Coding sequences were translated to amino acid sequences using the ExPASy Translate tool (ExPASy Proteomics Server). *GapC* identities of obtained DNA sequences were confirmed using blastp (Altschul *et al*., 1997) and their reduced amino acids sequences. DNA sequences were aligned using Clustal_X (Thompson *et al*., 1997). Mfold version 3.1 (mfold web server) (Zuker, 2003) was used to predict nucleic acid folding and the hybridization of involved sequences.

Amplifications of recombinant products. To verify recombination artifacts, we also designed four sets of inner primers specific for each suspicious recombinant product using Primer Premier 5.00 (PREMIER Biosoft International) and Oligo 6.31 (Molecular Biology Insights. Inc.) (Table 1). Nested PCR was performed to validate PCR-mediated recombination using the universal primer set combined with each specific inner primer set. Nested PCR was conducted using Olsen and Schaal's system (Olsen and Schaal, 1999) with modified cycling conditions, i.e., 95°C (2 min); then 20 cycles of 95°C (1 min), 46-62°C (1 min), 72°C (2 min); and a final extension at 72°C (10 min). Primary PCR reactions were performed using genomic DNA as templates, and GPDX7F and GPDX9R as primers; and secondary PCR was carried out using diluted primary PCR products (1/20-1/200) as templates, and inner primer sets as primers. Real (biological) recombinant products were expected to be amplified by corresponding inner primer pairs, whereas potential parental sequences can only be amplified by single primers (sense primer or anti-sense primer).

Results and Discussion

Exon-primed intron-crossing (EPIC) amplifications of *GapC* in *H. tiliaceus* using the universal primer set produced three distinct bands of length *ca.* 790-bp, 870-bp and 950-bp (Fig. 1). Extensive BLAST analyses then revealed that the sequences of clones from all amplification products were homologs of *GapC* with DNA similarities of from 58% to 86%. When examining the aligned sequences of *GapC* in *H. tiliaceus* in detail, we found four sequences, 3 of the 48 clones from Chinese CG and 1 of the 45 clones from Japanese JR, which appeared to be recombinant products (Fig. 2). The probabilities of recombinant artifacts were 6.25% (3/48) in CG and 2.22% (1/45) in JR, respectively. Suspicious recombination occurred

Table 1. Primer sets used in this study

Primer set*	Sense primer	Sense primer sequence	Anti-sense primer	Anti-sense primer sequence
1	GPDX7F	5' GATAGATTTGGAATTGTTGAGG 3'	GPDX9R	5' AAGCAATTCCAGCCTTGG 3'
2	SW137	5' GGACTGGAGAGGTGGTAGAG 3'	AW796	5' CCTGTATTGACATCCGATTT 3'
3	C2_02S	5' TATTTCGCAATGATTGTGGTT 3'	C2_02A	5' GCAATGGTGTCCCAAATTACTAC 3'
4	C3_07S	5' TATTTCGCAATGATTGTGGTT 3'	C3_07A	5' TATTGATAAGAGGAAGACTG 3'
5	C3_10S	5' ATTTCCCCCCTTTTCACTCT 3'	C3_10A	5' GTATTGACATCCGATTTCGGT 3'
6	J2_04S	5' TTTTGCAACCGTATGTGTATCG 3'	J2_04A	5' GGCAAATAAGATATCATAGGACG 3'

*: 1, universal primer; 2, primer set used to amplify a parental sequence *CG_2_01* (positive control); 3-6, primer sets used to amplify suspicious recombinant artifacts *CG_2_02*, *CG_3_07*, *CG_3_10*, and *JR_2_04*, respectively.

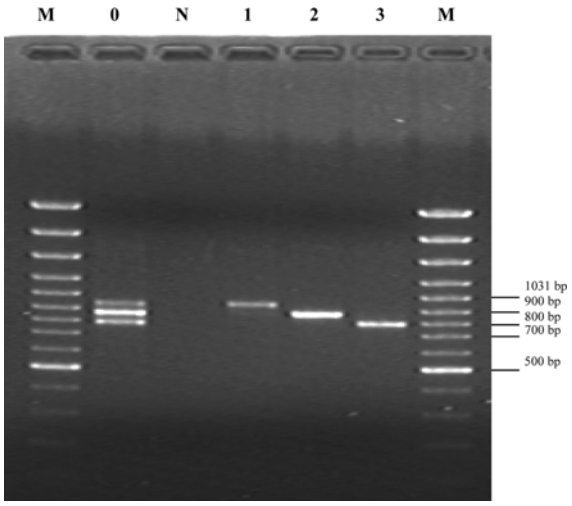


Fig. 1. Exon-primed intron-crossing (EPIC) amplifications of *GapC* in *H. tiliaceus* using the universal primer set GPDX7F and GPDX9R. Lane 0: PCR product showing three distinct bands; Lane N: negative control; Lanes 1-3: purified PCR products of each band of lengths of about 950-bp, 870-bp and 790-bp, respectively; Lane M: 100 bp DNA ladder.

between the 790- and 870-bp bands in CG (Fig. 2A-C), and between the 870- and 950-bp bands in JR (Fig. 2D). Two of the four suspicious recombinants were chimera formations from *CG_3_01* and *CG_2_01* (Fig. 2A, B), while the other two were chimera formations from *CG_3_03* and *CG_2_01*, and *JR_2_11* and *JR_1_01*, respectively (Fig. 2C, D). The aligned sequences of the potential parents and recombinants involved in Fig. 2 are provided in Fig. 3.

To test whether the observed recombinants arise via biological recombination or PCR-mediated recombination, we performed nested PCR using the universal primer set combined with a corresponding specific inner primer set for each recombinant (See Fig. 3 for the positions of specific inner primers). None of the nested PCRs succeeded in amplifying suspicious recombinants while the positive control amplified expected band for a potential parent sequence (Fig. S1 in Supplementary materials). These results indicated that recombinant products were derived via PCR-mediated recombination rather than biological recombination.

We further predicted the nucleic acid folding and hybridization of potential parental sequences at the PCR extension temperature (72°C) to examine the correlation between structural features of templates and PCR-induced *in vitro* chimera formation. Significant stem-loop structures were detected in all potential parent sequences and we compared positions of secondary structures in templates with recombination points for respective recombination events (Table 2). The secondary structure of template sequences *CG_2_01*, *CG_3_03* and their reverse-complement counterparts appeared to be involved in the chimera formations of recombinant *CG_2_02* (Fig. 4). That is, they showed overlapping positions of hairpins and recombination points (Table 2). For example, the *CG_3_03* hairpin at 579-596 bp overlapped the recombination point at 591-636 bp (Fig. 2C and Fig. 4B). As for the recombinants *CG_3_07* from templates *CG_3_01* and *CG_2_01* (Fig. 2A), the hairpins of templates were sometimes found to be close located to recombination points (see Table 2). For example, loop2 of *CG_3_01* at 383-397 bp was found to be close to the recombination point at 411-430 bp, whereas loop2 of *CG_2_01* at 458-472 bp was close to the recombination point at 486-505 bp. However, the association between the secondary structural features of templates and PCR-induced recombination was not obvious for the recombination of *JR_2_04* with *JR_2_11* or *JR_1_01* as templates (Fig. 2D and Table 2). We also examined the secondary structures of templates for the PCR-induced recombination of the homologous genes *Myb3* and *Myb5* from *Gossypium* allotetraploids (Cronn *et al.*, 2002); *Myb3* had one type of recombinant (R1), whereas

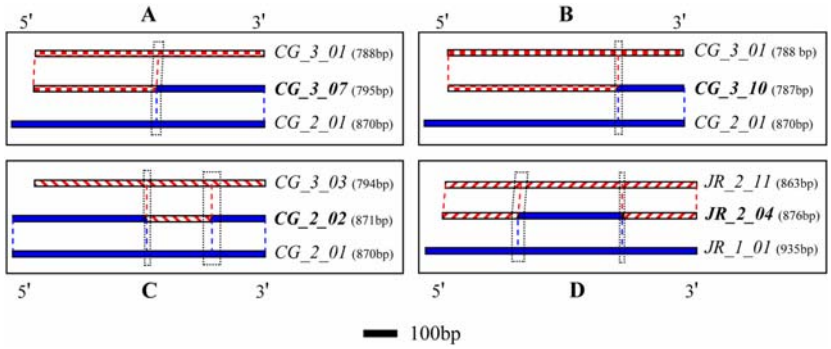


Fig. 2. Schematic representation of PCR-mediated recombination of the *GapC* gene family in *H. tiliaceus*. (A) recombinant *CG_3_07* from *CG_3_01* and *CG_2_01*; (B) recombinant *CG_3_10* from *CG_3_01* and *CG_2_01*; (C) recombinant *CG_2_02* from *CG_3_03* and *CG_2_01*; and (D) recombinant *JR_2_04* from *CG_2_11* and *CG_1_01*. For each case of PCR-mediated recombination, recombinant products are indicated in bold, and both parental sequences are indicated by the different types of bold lines. The code of each sequence is given following the criteria "individual_# band_# clone" and length (in bp) is indicated in parenthesis. Frames drawn with dashed lines indicate recombination points where sequence fragments were identical in all three clones. Aligned sequences of potential parents and recombinants for each PCR-mediated recombination are given in Fig. 3.

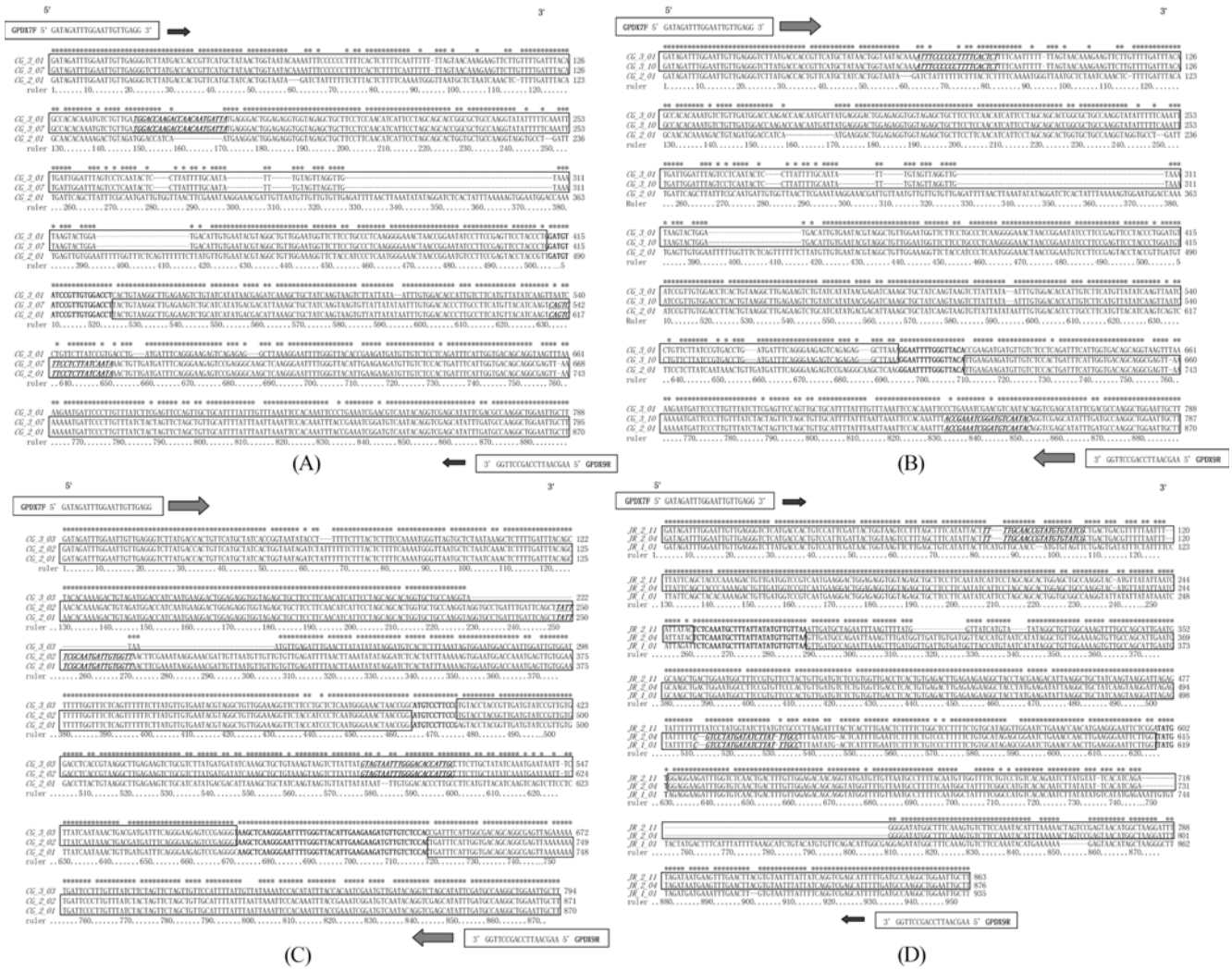


Fig. 3. Alignment of sequences involved in PCR-mediated recombinations as shown in Fig. 2. (A) recombinant *CG_3_07* from *CG_3_01* and *CG_2_01*; (B) recombinant *CG_3_10* from *CG_3_01* and *CG_2_01*; (C) recombinant *CG_2_02* from *CG_3_03* and *CG_2_01*; and (D) recombinant *JR_2_04* from *CG_2_11* and *CG_1_01*. The rectangular frames indicate recombinant sequences that are the same as template sequences. Sequences in bold letters are recombination points where sequence fragments were identical in all three clones. Positions for the primer pair GPDX7F and GPDX9R are also shown. The positions of specific internal primers are indicated by underlined bold italic.

Myb5 had seven types (R1-7). However, only a hairpin at 466-475 bp appeared to be associated with PCR-mediated recombination type R2 in *Myb5* (see Supplementary materials). The secondary structure features of templates therefore may contribute, but do not wholly determine PCR-induced *in vitro* chimera formation.

A variety of factors conceivably influence the degree of *in vitro* chimera formation during PCR amplification, including target sequence length, the number of partially homologous templates resident in the genome, cycling parameters (such as extension time and number of cycles), possibly nucleotide composition, the degree of divergence among potentially interacting loci (Cronn *et al.*, 2002), and the structure-specific exo/endonuclease activities of DNA polymerases (Tombline *et al*., 1996). To suppress PCR-mediated recombination, we used long elongation times and reduced cycling to the mid-point of amplification. Although decreased numbers of PCR cycles reduce the output of target sequences, this defect can be resolved by nested PCR. We successful amplified target fragments at desirable intensities, free from artifactually produced recombinant products by nested PCR with a 2 min extension time over 20 cycles of primary and of secondary PCR (Fig. S1 in Supplementary materials). In comparison with previously suggested approaches, nested PCR is more target specific than simple adjustments of PCR parameters and is more practicable than amplifying target loci from genomic libraries for most species. Moreover, predictions of the structural features of templates may be helpful in terms of primer design in future studies and would allow the risk of PCR-mediated recombination to be minimized.

Table 2. Relationship between secondary structure and PCR-mediated recombination. The positions of predicted loop structures for all template sequences are given in bp from 5' to 3'. We also show the locations of recombination points for each recombination event in respective parental sequences (see also Fig. 3 for sequences of recombination points). Overlapping positions for secondary structures and recombination points are indicated in red

	Secondary structure (5' to 3')			Recombination point (5' to 3')[a]					
	Loop1	Loop2	Loop3	A1	B1	C1	C2	D1	D2
CG_3_01	218-233	383-397		411-430	593-608				
CG_2_01	203-218	458-472		486-505	676-691	462-472	667-712		
CG_3_03	579-596					385-395	591-636		
JR_2_11	207-222	324-345	631-644					252-279	599-603
JR_1_01	566-583							256-283	616-620
CG_3_01*	391-407	555-572	646-665	358-377	180-195				
CG_2_01*	398-414	422-432	652-669	365-384	179-194	398-408	158-203		
CG_3_03*	389-396	579-596				399-409	158-203		
JR_2_11*	51-58	523-536						584-611	260-264
JR_1_01*	75-85	231-247						656-683	332-336

[a]Recombination points are numbered according to their order of appearance in Fig. 2. For example "C1" indicates the first recombination points from 5' to 3' in Fig. 2C.
*Reverse-complement counterpart

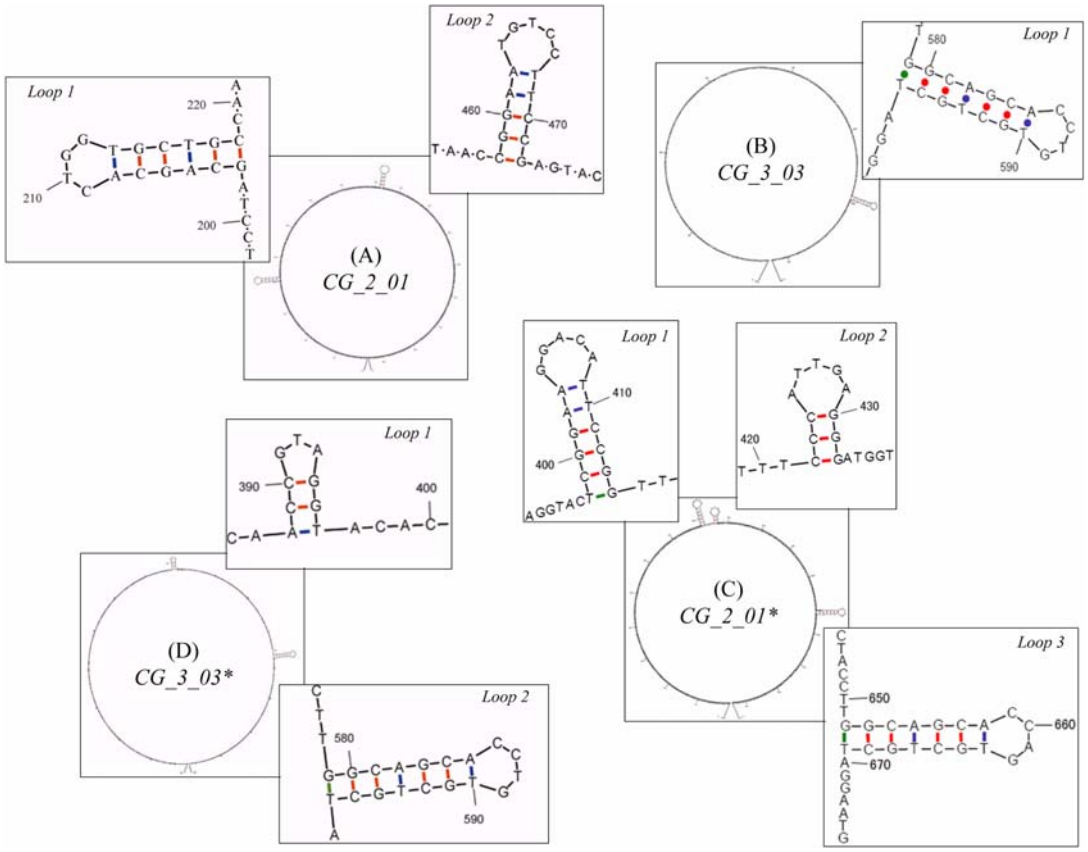


Fig. 4. Predicted secondary structures of *CG_2_01*, *CG_3_03* and their reverse-complement counterparts (A) *CG_2_01*, (B) *CG_3_03*, (C) *CG_2_01**, and (D) *CG_3_03**. The enlarged regions of loop are given for respective sequence. Secondary structures of the remaining template sequences are given in Supplementary Materials.

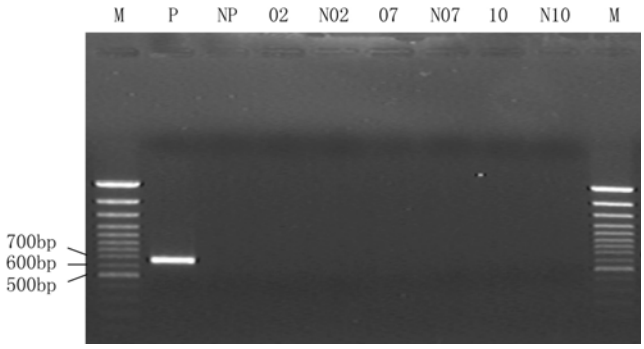
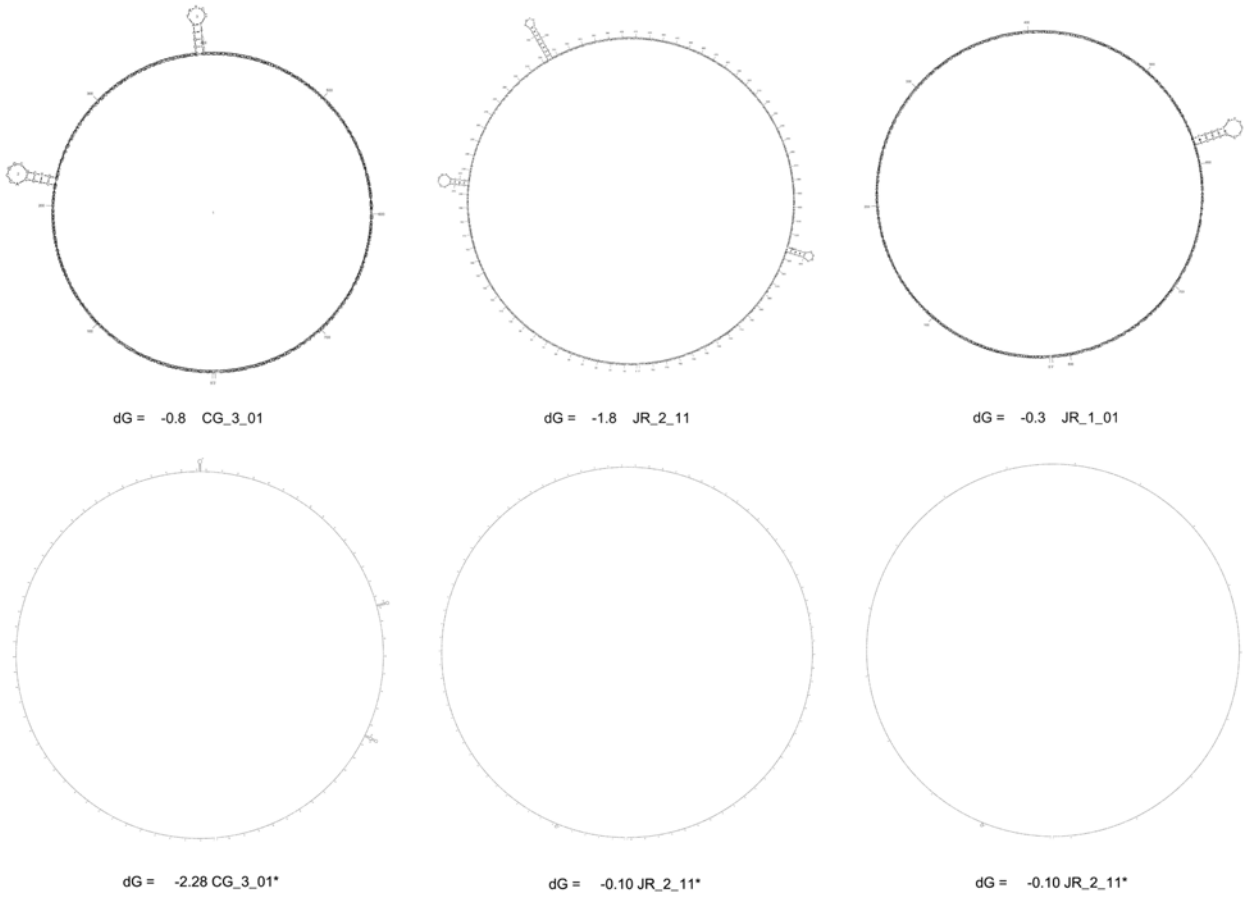


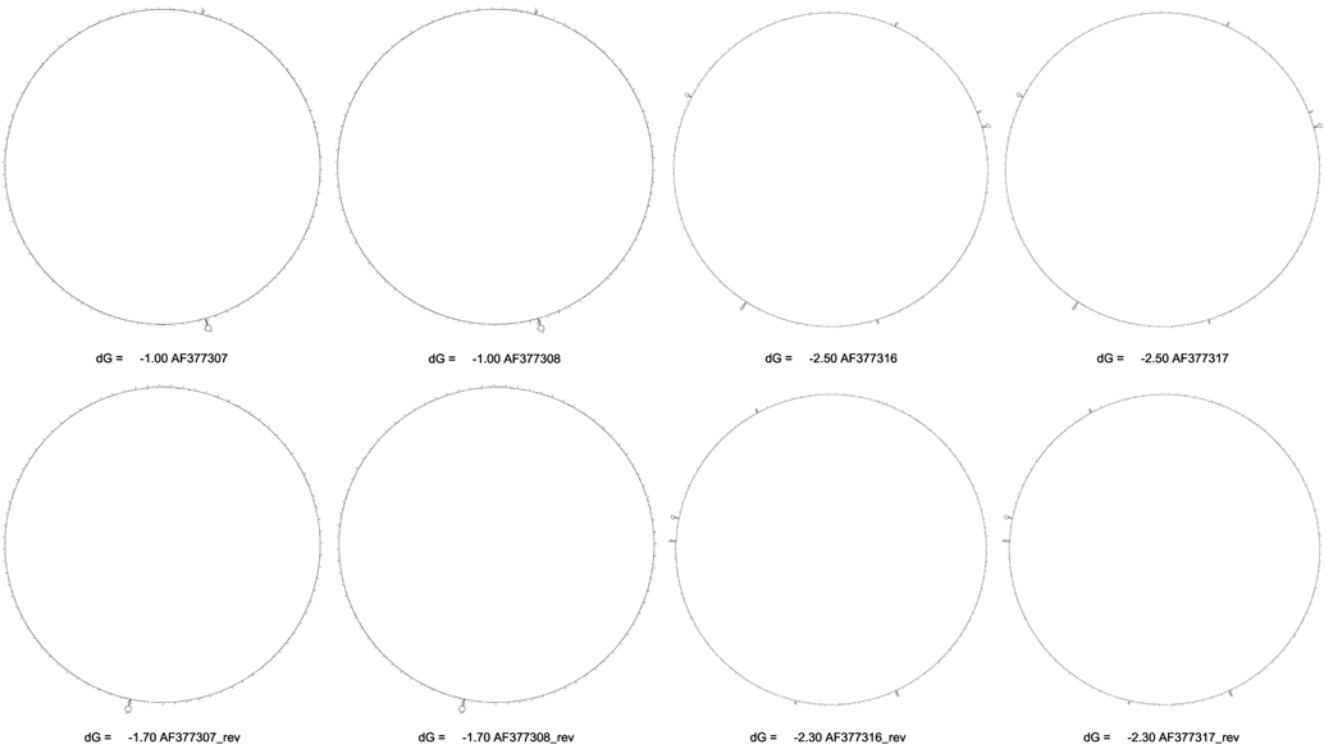
Fig. S1. Amplification of PCR-mediated recombination products using nested PCR. Lane (M): 100 bp DNA ladder; Lane (P): positive control, amplification of template *CG_2_01* using SW137 and AW796 as inner primer set; Lanes (02), (07), and (10): amplification of recombination products *CG_2_02*, *CG_3_07*, and *CG_3_10* using the respective inner primer sets C2_02S/C2_02A, C3_07S/C3_07A, and C3_10S/C3_10A; Lanes (NP), (N02), (N07), and (N10) are negative controls of lanes (P), (02), (07), and (10), respectively. The amplification of recombinant *JR_2_04* using J2_04S/J2_04A as an inner primer set showed the same pattern. The sequences and positions of each inner primer set are given in Table 1 and Fig. 3.

In summary, we show that PCR-mediated recombination can occur in a classical low-copy nuclear gene marker, which emphasizes that great caution should be taken when using gene lineages in population and evolutionary studies. We also demonstrate that PCR extension pauses are sometimes associated with template secondary structures, although these may not be the decisive cause of PCR-mediated recombination. Information on secondary structure would be useful to minimize the risk of PCR-recombination in advance. To suppress PCR-mediated recombination, nested PCR is a useful approach as it offers the advantages of target specificity and practicability.

Acknowledgments We thank Dr. Hasegawa Masami and Mr. Cairong Zhong for sample collection, Kenian Chen and Yang Shen for help with Mfold, and Dr. Yuguo Wang and Dr. Yelin Huang for technical assistance and helpful advice. This study was supported by the Chang Hungta Science Foundation and the Start-up Research Funds for Young Teachers from Sun Yat-Sen University, National Natural Science Foundation of China (30500049, 30470119, 30300033, and 30230030), the Open Foundation of State key Laboratory of Biocontrol, and the International Foundation for Science (IFS - D/3836).



File S1. Secondary structures of template sequences for PCR-induced recombination in GapC from *Hibiscus tiliaceus*. Page1 CG_3_01, Page2 JR_2_11, Page3 JR_1_01, Page4 CG_3_01*, Page5 CG_3_03*, Page6 JR_2_11*, Page8 JR_1_01*.



File S2. Secondary structures of template sequences for PCR-induced recombination in homologous *Myb3* and *Myb5* genes from *Gossypium hirsutum* A-genome and D-genome. Page1 AF377307, *Gossypium hirsutum* D-genome myb-like transcription factor *Myb3* gene; Page2 AF377308, *Gossypium hirsutum* A-genome myb-like transcription factor *Myb3* gene; Page4 AF3773016, *Gossypium hirsutum* D-genome myb-like transcription factor *Myb5* gene; Page5 AF377317, *Gossypium hirsutum* A-genome myb-like transcription factor *Myb5* gene; Page6-8 reverse complement reverse-complement counterparts to AF377307, AF377308, AF377316, and AF377317, respectively.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang J., Zhang Z., Miller W. and Lipman D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*. **25**, 3389-3402.

Bradley, R. D. and Hillis, D. M. (1997) Recombinant DNA sequences generated by PCR amplification. *Mol. Biol. Evol*. **14**, 592-593.

Cerff, R. (1982) Evolutionary divergence of chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from angiosperms. *Eur. J. Biochem*. **126**, 513-515.

Crameri, A., Raillard, S. A., Bermudez, E. and Stemmer, W. P. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288-291.

Cronn, R., Cedroni, M., Haselkorn, T. and Grover C. (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor. Appl. Genet*. **104**, 482-489.

Doyle, J. J. and Doyle, J. L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*. **19**, 11-15.

Fang, G., Weiser, B., Visosky, A., Moran T. and Burger H. (1999) PCR-mediated recombination: A general method applied to construct chimeric infectious molecular clones of plasma-derived HIV-1 RNA. *Nat. Med*. **5**, 239-242.

Manjunath, S. and Sache, M. M. (1997) Molecular characterization and promoter analysis of the maize cytosolic glyceraldehyde-3-phosphate dehydrogenase gene family and its expression during anoxia. *Plant Mol. Biol.* **33**, 97-112.

Marri, L., Sparla, F., Pupillo, P. and Trost, P. (2005) Co-ordinated gene expression of photosynthetic glyceraldehyde-3-phosphate dehydrogenase, phosphoribulokinase, and CP12 in *Arabidopsis thaliana*. *J. Exp. Bot.* **56**, 73-80.

Olsen, K. M. and Schaal, B. A. (1999) Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. *Proc. Natl. Acad. Sci. USA* **96**, 5586-5591.

Pérusse, J. R. and Schoen, D. J. (2004) Molecular evolution of the *GapC* gene family in *Amsinckia spectabilis* populations that differ in outcrossing rate. *J. Mol. Evol.* **59**, 427-436.

Russell, D. A. and Sachs, M. M. (1989) Differential expression and sequence analysis of the maize glyceraldehyde-3-phosphate dehydrogenase gene family. *Plant Cell.* **1**, 793-803.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (2001) *Molecular cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press, New York, USA.

Scharf, S. J., Friedmann, A., Brautbar, C. Szafer E., Steinman L., Horn G., Gyllensten U. and Erlich H. A. (1988) HLA class II allelic variation and susceptibility to *Pemphigus vulgaris*. *Proc. Natl. Acad. Sci. USA* **85**, 3504-3508.

Shammas, F. V., Heikkila, R. and Osland, A. (2001) Fluorescence-based method for measuring and determining the mechanisms of recombination in quantitative PCR. *Clinica Chimica Acta*

304, 19-28.

Strand, A. E., Leebens-Mack, J. and Milligan, B. G. (1997) Nuclear DNA-based markers for plant evolutionary biology. *Mol. Ecol.* **6**, 113-118.

Thompson, J. D., Gibson, T. J., Plewniak, F. Jeanmougin F. and Higgins D. G. (1997) The Clustal_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

Tombline, G., Bellizzi, D. and Sgaramella, V (1996) Heterogeneity of primer extension products in asymmetric PCR is due both to cleavage by a structure-specific exo/endonuclease activity of DNA polymerases and to premature stops. *Proc. Natl. Acad. Sci. USA* **93**, 2724-2728.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415.