ISSN 1672-9145 **Acta Biochimica et Biophysica Sinica** 2007, **39**(9): 675–683 **CN** 31-1940/Q

Isolation of *Ty1-copia-like* Retrotransposon Sequences from the Apple Genome by Chromosome Walking Based on Modified SiteFinding-polymerase Chain Reaction

Guiling ZHAO[1,2], Zhihong ZHANG[1], Haiyue SUN[1], He LI[1], and Hongyan DAI[1*]

[1] *College of Horticulture, Shenyang Agricultural University, Shenyang 110161, China;*
[2] *College of Forestry, Shenyang Agricultural University, Shenyang 110161, China*

Abstract Long terminal repeat (LTR) retrotransposons are powerful tools for studying genetic biodiversity, genome evolution, gene mutation, gene cloning and gene expression. The scarcity of retrotransposon sequence information restricts the development of these studies in higher plants. In the present study, 31 reverse transcriptase (RT) genes of *Ty1-copia-like* retrotransposons were identified from the apple genome by amplifying the RT coding region using degenerate primers. Nineteen RT genes showed extreme heterogeneity in terms of fragment size, base pair composition and open reading frame integrality. Originating from one 266 bp cloned RT gene, a 1966 bp *Ty1-copia-like* retrotransposon (named *Tcrm1*), including RT-ribonuclease H-LTR domain sequences, was achieved by chromosome walking based on modified SiteFinding-polymerase chain reaction. The comparison between *Tcrm1* and other LTR retrotransposons in gene structure and sequence homology shows that *Tcrm1* is the first *Ty1-copia-like* retrotransposon including an LTR domain in the apple genome. Dot blot analysis revealed that *Tcrm1* copy number in the apple was approximately 1×10^3 copies per haploid genome.

Keywords *Malus domestica*; *Ty1-copia-like* retrotransposon; long terminal repeat sequence; modified SiteFinding-polymerase chain reaction

Retrotransposons are mobile genetic elements that transpose through reverse transcription of an RNA intermediate. Due to their ubiquitous distribution, high copy numbers, widespread chromosomal dispersion, considerable sequence heterogeneity and insertional polymorphism in eukaryotes [1–5], retrotransposons are potentially powerful tools for studying genetic biodiversity, genome evolution, gene mutation, gene cloning and gene expression [6–13]. Retrotransposons are classed as LTR or non-LTR retrotransposons on the basis of presence or absence of long terminal repeat (LTR) sequences on both ends. All LTR retrotransposons have the same general structure, which is 5′-LTR-primer binding site (PBS)-coding region-polypurine tract (PPT)-LTR-3'. According to their similarity of sequence and the order of encoded gene products, LTR retrotransposons can be further subdivided into *Ty1-copia* and *Ty3-gypsy* [14]. Because of sequence variability in the non-coding regions, particularly in the LTR region, only a very small number of complete sequences of LTR retrotransposons have been isolated from plants [15–20].

In the past few years, several retrotransposon-based molecular marker systems have been explored [21–27]. The most popular of these is the sequence-specific amplification polymorphism (S-SAP) technique developed by Waugh *et al*. [21]. S-SAP, a technique of revealing the genetic distribution of retrotransposable elements, is a combination of the general principle of amplified fragment length polymorphism (AFLP) and sequence-specific polymerase chain reaction (PCR). Compared with AFLP, this method produces a higher number of polymorphisms within and between species, a more even distribution across the genome and a better quality of amplification products [21,26]. In addition, S-SAP-based taxonomic data are more consistent with geographical and morphological criteria than AFLP-based markers [25,28]. It should be noted that

Received: February 26, 2007 Accepted: April 15, 2007

This work was supported by a grant from the National Natural Science Foundation of China (No. 30500348)

*Corresponding author: Tel, 86-24-88487636; Fax, 86-24-88487144; E-mail, daihy@syau.edu.cn

DOI: 10.1111/j.1745-7270.2007.00328.x

©Institute of Biochemistry and Cell Biology, SIBS, CAS

S-SAP is effective in identifying variation among somatoplasm clones [29]. However, the development of S-SAP requires sequence information of the terminal regions of the retrotransposon [21]. Currently, the scarcity of retrotransposon LTR sequences limits the application of S-SAP to higher plants.

To date, several strategies have been developed to isolate retrotransposon LTR sequences, including coding sequence trapping [15,18], genomic library screening [30, 31], reverse transcription-PCR (RT-PCR) [32,33], ligation-mediated PCR (biotin-capture PCR and cassette PCR) [16,17,19,20] and sequence similarity searching [33,34]. However, none of these strategies has been widely applied, because of their inefficiency or difficulty, or both. For example, while sequence similarity searching is simple, it can only be applied where sequences encompassing LTR domains are known. The other four methods can directly identify LTR sequences, but require complicated manipulations, such as genomic library screening, restriction cleavage, size optimization, ligation, ribonucleaseH (RNaseH) primer biotinylation and cDNA synthesis.

SiteFinding-PCR [35] aims to isolate unknown genes adjacent to a known DNA sequence. In principle, this PCR is primed by a SiteFinder at low annealing temperature, after which the target molecules are amplified exponentially using gene-specific and SiteFinder primers, and then they are screened out by another gene-specific primer and a vector primer. However, non-target molecules cannot be amplified exponentially owing to the suppression effect of their stem-loop structure, and they can not be screened out. Tan *et al.* [35] succeeded in using SiteFinding-PCR to explore chromosome walking. Compared with other PCR and chromosome walking combinations, this method does not require size optimization, biotinylation and cDNA synthesis.

Apples are important temperate fruits and are widely distributed throughout the world, with an annual production of 63 million tons (http://faostat.fao.org/DesktopDefault.aspx?PageID=291&lang=en 2005). Bud sport is prevalent in apples, and most of the apple clones grown worldwide, such as "Fuji", "Gala" and "Delicious", are bud sport cultivars. The mechanism(s) of bud sports is poorly understood. Previous studies have shown that retrotransposons contribute to understanding the mechanism(s) of bud sports of fruit trees; for example, parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor [18]. Kobayashi *et al.* [36] found that mutations in grape skin color were induced by retrotransposon insertion. Up to August 2006, 97 retrotransposon sequences from the apple genome had been submitted to the GenBank database. Of these, 92 were partial size retroelements that were obtained by degenerative PCR or sequence similarity searching [37]. Three full-size terminal repeat retrotransposons in miniature (TRIM) were isolated using a universal primer method [38]. One full-size LTR retrotransposon was deposited in the database under GenBank accession No. AM167520, but the method of isolation was not reported. Another full-size LTR retrotransposon (*dem1*) was characterized using coding sequence trapping [18], and then based on LTR sequence information of *dem1*, the S-SAP approach was developed to characterize and distinguish bud sport clones of apple [27].

In the present study, based on one of the reverse transcriptase (RT) sequences isolated by degenerate primers aimed at RT conserved motifs, modified SiteFinding-PCR was explored to identify RT-RNaseH-LTR domain sequences of *Ty1-copia-like* retrotransposons from the apple genome. This work lays the foundation for establishing retrotransposon-based molecular marker systems of apple and will be useful for understanding the mechanism(s) of apple bud sports.

Materials and Methods

DNA extraction

Five apple (*Malus domestica*) cultivars, "Judafuji", "Hongwangjiang", "Yanfu", "Linyiduan" and "Liaozaofu", were grown at the Changli Institute of Pomology, Academy of Agricultural and Forestry Sciences, Hebei Province (Changli, China). Unexpanded fresh leaves were collected, placed into liquid nitrogen, and stored at –70 ºC before use. Total DNA was extracted according to the protocol of Lodhi *et al.* [39] with the following modifications: cetyl trimethyl ammonium bromide (CTAB) solution [2% (*W/V*) CATB, 100 mM Tris-HCl, pH 8.0, 20 mM EDTA, pH 8.0, 1.4 M NaCl, 0.2% (*V/V*) 2-mercaptoethanol] stored at room temperature was heated to 65 ºC for use. Chloroform isoamyl alcohol (24:1, *V/V*) was used instead of chloroform octanol (24:1, *V/V*), and was centrifuged at 12,875 *g* for 10 min when necessary. Treatment with 1 µl of RNaseA (10 mg/ml) per 100 µl of DNA solution at 37 ºC for 1 h was performed before DNA precipitation.

Degenerative PCR amplification of RT gene sequences

Two degenerative primers (DP1: 5′-ACNGCNTTPyPyTNCAPyGG and DP2: 5′-APuCATPuTCPuT-

©Institute of Biochemistry and Cell Biology, SIBS, CAS

CNACPuTA, where N=A+T+C+G, Pu=A+G and Py=T+C) designed by Kumar *et al*. [4] were used. The sequences of the two primers correspond to the highly conserved peptide sequences TAFLHG (motif 1) and YVDDML (motif 3), which flank the internal domain of RT [**Fig. 1(A)**]. The PCR mixtures contained 10–50 ng apple genomic DNA, 2 μl 10×PCR buffer (100 mM Tris-HCl, pH 8.8, 500 mM KCl, 0.8% Nonidet P-40), 25 mM $MgCl_2$, 0.4 μl dNTPs (2.5 mM each; TaKaRa, Dalian, China), 6 pmol each degenerative primer (DP1 and DP2) and 0.5 U *Taq* DNA polymerase (TIANGEN, Beijing, China). The final volume was made up to 20 μl with sterile distilled water. The amplifications were performed in 0.2 ml Eppendorf tubes on a PTC-150 PCR System (MJ Research, Waltham, USA) with the following cycling conditions: 94 ºC for 2 min, 35 cycles at 94 ºC for 30 s, 53 ºC for 60 s, 72 ºC for 90 s, and a final extension at 72 ºC for 5 min.

SiteFinding-PCR amplification of RT-RNaseH-LTR gene sequences

Primers and protocol for modified SiteFinding-PCR The protocol of SiteFinding-PCR was adopted with the following modifications (**Fig. 2**): the desirable products were directly cloned into TA vectors, the concentration of primers was reduced to 50%, and the annealing temperatures in steps B and C were decreased to 60.2 ºC and 64 ºC, respectively (**Fig. 3**). The SiteFinder and the corresponding primers were identical to those of Tan *et al*. [35] [**Fig. 1(B)**]. Six gene-specific primers (GSP) [**Fig. 1(A)**] were designed using Oligo 6.0 software (Molecular Biology Insights, Cascade, USA). The distance between GSP2 and GSP3 was 46 bp while the distance between GSP6 and GSP7 was 49 bp.

SiteFinding reaction PCR mixtures of 20 μl included 10–100 ng genomic DNA, 2 μl dNTP (2.5 mM each), 0.7 μl 10×LA PCR buffer (TaKaRa), 5 pmol SiteFinder and 0.5 U *Taq* DNA polymerase (TIANGEN). The PCR reactions were performed in 0.2 ml Eppendorf tubes on a PTC-150 PCR System with a single cycle. The cycling conditions were 92 ºC for 2 min, 95 ºC for 1 min, 25 ºC for 1 min, slope 43 ºC with 0.2 ºC per second and 68 ºC for 10 min.

Nested PCR For the first round of PCR, 5 μl of primer mixtures [25 pmol SiteFinder primer (SFP) 1, 5 pmol GSP1 and 3×LA PCR buffer] were added to the tubes containing SiteFinding reaction mixtures (20 μl) and then the PCR was run with the following cycling conditions: an initial denaturation at 94 ºC for 1 min, followed by 30 cycles at

Fig. 1 SiteFinder and the primers

(A) Two degenerative primers (DP) and six gene-specific primers (GSP) are indicated by black lines and black arrows, respectively. Two motifs corresponding to degenerative primers are shown. (B) SiteFinder and its primers (SFP1 and SFP2). The GCCT nucleotides at the 3′ end of the SiteFinder, with the help of NNNNNN, were used to anneal with the complimentary site on genomic DNA at low temperature and initiate SiteFinding-polymerase chain reaction (PCR). SFP1 and SFP2 were used in the first and second rounds of PCR, respectively.

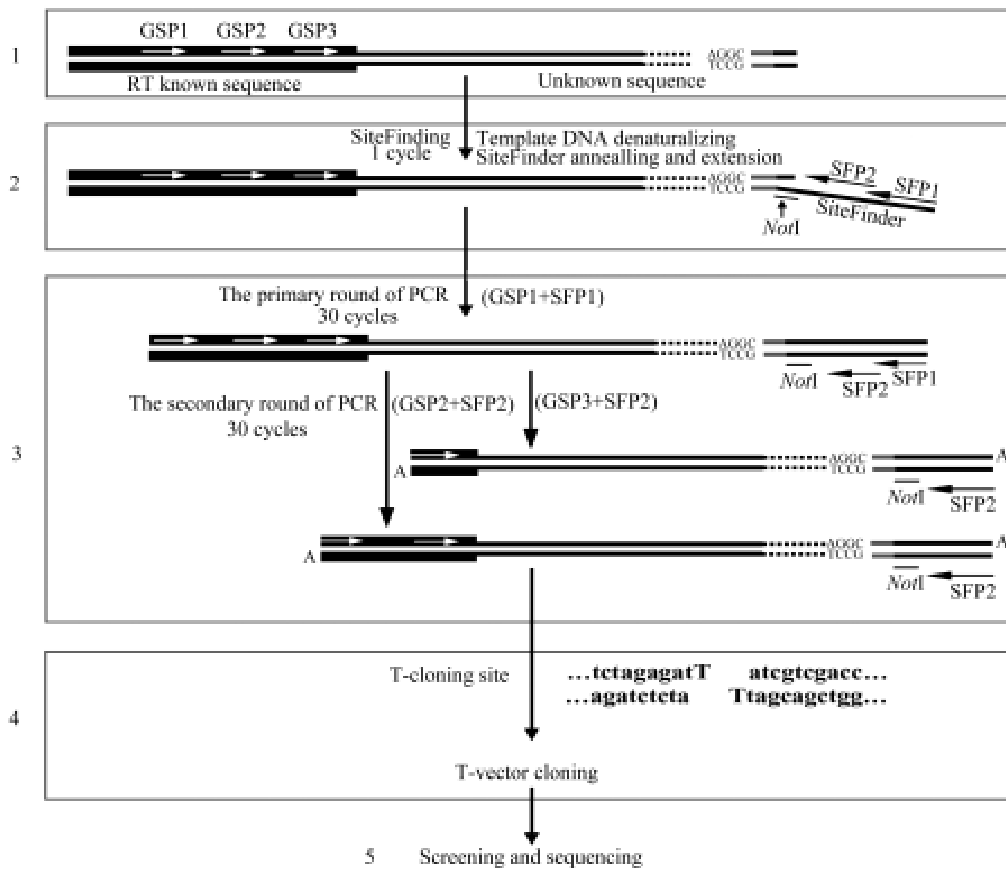

Fig. 2 Outline of modified SiteFinding-polymerase chain reaction (PCR)

1, original target molecule genomic double-strand templates with known and unknown sequences. Known and unknown sequences are depicted with thick and thin lines, respectively. Gene-specific primers (GSP)1–3 could anneal with known sequences (white arrows). 2, SiteFinding reaction. After denaturation, low temperature annealing and extension, the strand with AGGC was generated as double-stranded target molecules with SiteFinder. 3, nested PCR. The products of step 2 were exponentially amplified by nested PCR. The first round of PCR was with primers GSP1 and SiteFinder primer (SFP)1, and the second round of PCR was with primers GSP2/GSP3 and SFP2. 4, cloning target molecules. The largest fragments of the second round of PCR were purified and then cloned into a T-vector. 5, screening and sequencing. White colonies were selected to extract plasmid DNA from and the inserts were digested with *Eco*RI and *Hin*dIII restriction endonucleases to confirm identity of band size. The target molecules were subsequently sequenced. This outline was adopted in step B and the same outline was used for step C except that GSP4, GSP6 and GSP7 primers rather than GSP1, GSP2 and GSP3 primers, respectively, were used.

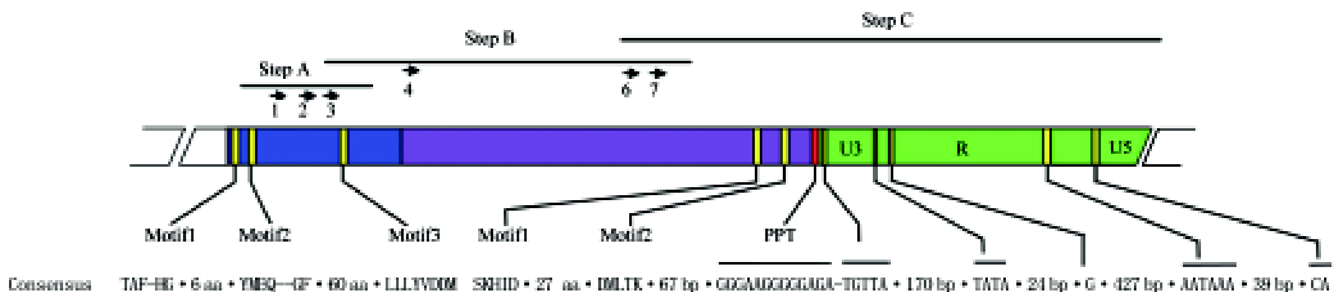

Fig. 3 Chromosome walking procedure, gene structure and character sequences of *Tcrm1*

The three black lines indicate the steps in chromosome walking. The six single arrows represent the six gene-specific primers (GSP) (1, 2, 3, 4, 6 and 7 stand for GSP1, GSP2, GSP3, GSP4, GSP6 and GSP7). The blue, pink and green boxes correspond to reverse transcriptase (RT) gene, ribonucleaseH (RNaseH) and long terminal repeat (LTR) regions, respectively. The yellow boxes are the character sequences and conserved motifs (comparison between *Tcrm1* and *Tto1-1*). The red box represents the polypurine tract (PPT). The numbers of amino acids or nucleotides separating the motifs are indicated between dots. Every short line is an amino acid or a nucleotide.

©Institute of Biochemistry and Cell Biology, SIBS, CAS

95 ºC for 10 s, 60.2 ºC for 6 min, and an elongation step at 72 ºC for 5 min.

For the second round of PCR, 5 µl PCR products of the first round were diluted with 495 µl sterile distilled water, and then 0.4 µl diluted solution was added to each of two Eppendorf tubes (tube A and tube B) containing 19.6 µl PCR mixtures of the second round. Tube A contained 0.5 U *Taq* DNA polymerase (TIANGEN), 0.2 µl dNTP (2.5 mM each), 3 µl 10× LA PCR buffer, 2 pmol each of GSP2 and SFP2, and 15.4 µl sterile distilled water. In tube B, GSP2 was replaced with GSP3, but all other components and their concentrations were identical to those in tube A. The PCR cycling conditions were the same as those for the first round of PCR.

The above description of nested PCR refers to step B. The cycling conditions, components of reaction mixtures and their concentrations in step C (**Fig. 3**) were identical to those of step B, except that the primers GSP1, GSP2 and GSP3 used in step B were replaced by GSP4, GSP6 and GSP7, respectively, in step C.

Aliquots of PCR products of the second round were electrophoresed on 1.5% agarose gels containing 0.5 µg/ml ethidium bromide. The desirable bands were judged by the difference in product sizes in the second round PCR, which were 46 bp in step B and 49 bp in step C. Each experiment was repeated three times.

Cloning, colony selection and DNA sequencing

The desirable bands were purified with the Agarose gel DNA purification kit (TaKaRa), cloned into the TA-cloning site of the pMD18-T vector (TaKaRa), and then transformed into the JM109 strain of *Escherichia coli*. White colonies were selected for extracting plasmid DNA. The inserts were digested with *Eco*RI and *Hin*dIII restriction endonucleases to confirm band size and identity.

DNA sequencing was performed on an ABI PRISM 3730 sequencer (ABI, Foster city, USA) with M13 primers and BigDye terminator v3.1 (ABI). This work was done by Shanghai Sangon Biological Engineering Technology & Services (Shanghai, China).

Sequence analysis

Basic sequence analysis was done using DNAMAN software (Lynnon Biosoft, Vaudreuil, Canada). Homology searches were carried out using the programs BLASTn and BLASTx against the sequences in the GenBank database (http://www.ncbi.nlm.nih.gov). Multiple alignments of the deduced amino acids to other retrotransposons were done using DNAMAN. The corresponding sequences used for comparison, including *Copia* (*Drosophila melanogaster* retrotransposon P04146), *Tto1-1* (*Nicotania tabacum* retrotransposon BAA11674) and *Tnt1-94* (*Nicotania tabacum* retrotransposon P10978), were obtained from the GenBank database. The apple retrotransposons mentioned here were obtained by searching the GenBank database with the key words "*Malus domestica* retrotransposon".

Dot blotting

The 289 bp heterogeneous PCR product amplified with a primer pair SP1/SP2 located in the LTR region of *Tcrm1* (SP1: 5′-GTTAGAATTCCCTCCTCGTGA-3′ and SP2: 5′-TAGAGGCTGACTAACTCAAGA-3′) was used as a probe labeled with the DIG high prime DNA labeling and Detection starter kit II (Roche, Basel, Switzerland) according to the manufacturer's instructions. Genomic DNA and the 289 bp heterogeneous PCR product were serially diluted, denatured in 1 M NaOH with 0.2 M EDTA (pH 8.2) for 20 min, and transferred onto a nylon membrane (Hybond-N+; Amersham, Piscataway, USA) using a vacuum dot blotter. Hybridization was performed at 65 ºC for 18 h. *Tcrm1* copy numbers were calculated by using the following equation [40]:

Copy number=(size of the haploid genome×average proportion of nuclear genomic DNA hybridizing to the probe)/size of probe element. Signal intensity was acquired with VisionWorks LS gel software (Upland, USA).

Results

Ty1-copia-like retrotransposon RT genes from the apple genome

PCR amplification with degenerate primers DP1 and DP2 yielded a single band of the expected size (approximately 270 bp) for five apple samples. The five fragments were cloned, and 31 individual subclones were sequenced. Based on a search against the GenBank database, these 31 sequences showed clear similarity to RT sequences, and their deduced amino acids showed significant homology to RT (data not shown).

Unlike previous work that showed high sequence heterogeneity was detected within and between species, 12 out of the 31 cloned RT genes were repeated sequences and only 19 RT sequences (deposited in the GenBank nucleotide database under the accession numbers DQ410748–DQ410766) showed heterogeneity. The ratios of AT and CG of these 19 RT sequences varied from 1.25 to 1.82.

The sizes of all 19 RT sequences, ranging from 254 to 266 bp, were shorter than 273 bp [3], which indicated that deletion mutations were ubiquitous in *Ty1-copia-like* retrotransposons from the apple genome. The deduced amino acid homologies between each pair of the 19 fragments varied from 17.2% (DQ410751 and DQ410757) to 87.4% (DQ410765 and DQ410748). Eleven of the 19 RT gene sequences were found to encode a single open reading frame (ORF) (deposited in the GenBank protein database under the accession numbers ABD76543–ABD76553), while the ORF of the other eight sequences were interrupted by either stop codons or translational frame-shifts.

Chromosome walking based on modified SiteFinding-PCR

Originating from one 266 bp cloned RT gene (DQ410750), RT-RNaseH-LTR domain sequences of a *Ty1-copia-like* retrotransposon were elucidated from the apple genome by chromosome walking (step B and step C) based on modified SiteFinding-PCR (**Fig. 3**). The largest fragments in step B and step C were recovered, cloned and sequenced (**Fig. 4**).

In step B, three sense retrotransposon specific primers (GSP1, GSP2 and GSP3) were designed from the cloned RT gene sequence (DQ410750) to generate overlapping sequences, so that a contig sequence was established (**Fig. 3**). Four specific DNA bands, approximately 450, 600, 650 and 830 bp in size [**Fig. 4(A)**, lane 2], were obtained in the second round of PCR with the primer pair GSP3 and SFP2. The sequencing results showed that the largest fragment was 836 bp. Along the sequence, the distances

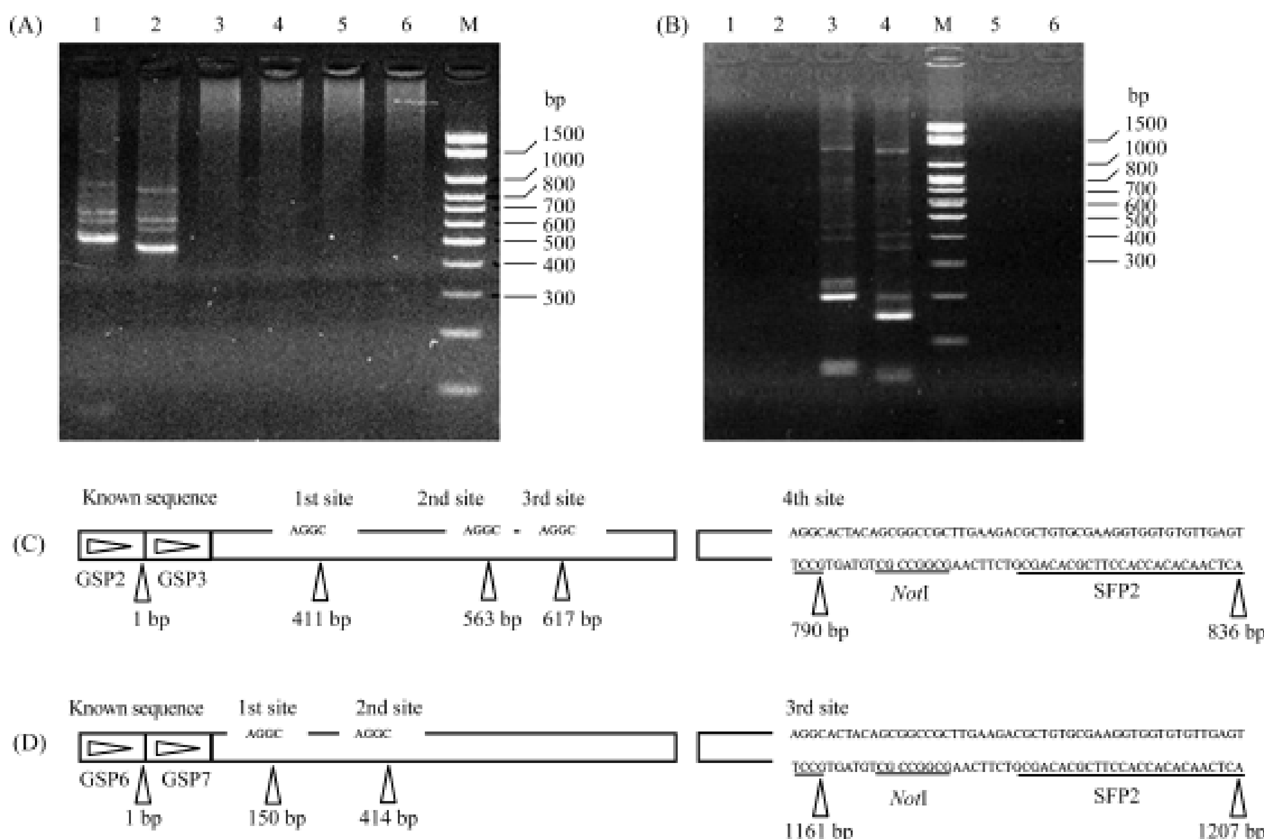

Fig. 4 Chromosome walking based on modified SiteFinding-polymerase chain reaction (PCR)

(A) Products of the second round of PCR in step B. Lane M, 100 bp DNA ladder; lanes 1 and 2, apple; lanes 3 and 4, strawberry; lanes 5 and 6, distilled water. Lanes 1–6 contained PCR products obtained with primer couples SiteFinder primer (SFP)2/gene-specific primer (GSP)2, SFP2/GSP3, SFP2/GSP2, SFP2/GSP3, SFP2/GSP2 and SFP2/GSP3, respectively. The largest fragment in lane 2 was cloned and sequenced as indicated in (C). (B) Products of the second round of PCR in step C. Lane M, 100 bp DNA ladder [same as (A)]; lanes 1 and 2, strawberry; lanes 3 and 4, apple; lanes 5 and 6, distilled water. The PCR products in lanes 1–6 were obtained with primer couples SFP2/GSP6, SFP2/GSP7, SFP2/GSP6, SFP2/GSP7, SFP2/GSP6 and SFP2/GSP7, respectively. The largest fragment in lane 3 was cloned and sequenced as indicated in (D). (C) Four GCCT sites (white arrowheads) on the 836 bp sequences and the distance between them and GSP3. These distances, if 46 bp (the distance between the C of AGGC and the last A of the primer SFP2) was added, were consistent with the gel electrophoresis results (A). (D) Four GCCT sites (white arrowheads) on the 1207 bp sequences and the distance between them and GSP6. These distances, if 46 bp (the distance between the C of AGGC and the last A of the primer SFP2) was added, were consistent with the gel electrophoresis results in (B).

©Institute of Biochemistry and Cell Biology, SIBS, CAS

between GSP3 and the other four AGGC sites were 411, 563, 617 and 790 bp [**Fig. 4(C)**]. If 46 bp (the distance between the C of AGGC and the last A of the primer SFP2) was added, the distances between GSP3 and the other four AGGC sites would then be consistent with the gel electrophoresis [**Fig. 4(A)**, lane 2, and **Fig. 4(C)**]. By conducting a BLASTn and a BLASTx search, 10 retrotransposon sequences and 37 retrotransposon RTs showed high homology with the sequence of 836 bp. The identities were 80% and 45%, respectively. The partial sequences of three plant LTR retrotransposons (*Copia*, 1136–1270 aa; *Tto1-1*, 1060–1219 aa; *Tnt1-94*, 1059–1218 aa) were selected to compare with the deduced amino acids of the 836 bp sequence, and the results showed that the average identity was 57.65%. However, the 836 bp sequence did not encompass the conserved motif 1 and motif 2 that characterize the RNaseH region of *Ty1-copia-like* retrotransposons (**Fig. 3**).

In step C, three sense retrotransposon specific primers (GSP4, GSP6 and GSP7) were designed from the 836 bp sequence obtained in step B. DNA bands of approximately 200, 450 and 1200 bp were obtained in the second round of PCR using primers GSP6 and SFP2 [**Fig. 4(B)**, lane 3]. The sequencing results showed that the largest fragment was 1207 bp, which had a 147 bp overlapping sequence with the 836 bp sequence obtained in step B. The distances between the primer GSP6 and the other three AGGC sites along this 1207 bp sequence [**Fig. 4(D)**], were consistent with the gel electrophoresis if 46 bp (the distance between the C of AGGC and the last A of the primer SFP2) was added [**Fig. 4(B)**, lane 3, and **Fig. 4(D)**]. The deduced amino acids of the 1161 bp sequence contained not only the conserved motif 1 and motif 2 that characterize the RNaseH region of retrotransposons but also the potential PPT and LTR region (**Fig. 3**).

Therefore, one upstream degenerative amplification (step A, generated a fragment of 266 bp) and two downstream steps (step B and step C, generated largest fragments of 790 and 1161 bp, with overlapping sequences of 109 and 147 bp, respectively) of chromosome walking on the basis of modified SiteFinding-PCR established a contig sequence of 1966 bp (**Fig. 3**), which has been deposited in the GenBank nucleotide database under the accession number DQ898280 and named *Tcrm1*.

Primary structure of *Tcrm1* and homology analysis

The 1289 bp central domains of *Tcrm1* contained an ORF encoding typical RT and RNaseH. Several short highly conserved amino acids that characterize RT and the RNaseH region of *Ty1-copia-like* retrotransposons were found in the *Tcrm1* ORF (**Fig. 3**).

Tcrm1 had the same gene structure as *Ty1-copia-like* retrotransposons. From 5′ to 3′, it encoded the RT and the RNaseH, sequentially followed by the PPT (5′-GGGAAGGGGGAGA-3′) and the 666 bp LTR region. Three typical regions of retrotransposon LTR, U3, R and U5 [41] (**Fig. 3**), were also discovered in *Tcrm1*. The U3 region was 197 bp, starting with 5′ TG 3′ (position 1299–1300 bp) and ending with a G residue (position 1475) of the U3-R boundary. U3 had a potential TATA box (position 1468–1471 bp), which was similar to the description of Joshi [42]. The characteristic sequence AATAAA (position 1923–1928 bp) in the R region was identical to the consensus polyadenylation signal [43]. The CA motif, which marks the R-U5 boundary, was found at the 39th nucleotide after the identified polyadenylation signal.

The differences between *Tcrm1* and other reported apple retrotransposons are as follows: (1) compared with 92 other retroelements with only coding region sequences, *Tcrm1* encompasses LTR gene sequences of non-coding region; (2) compared with three full-size TRIM with no ORF encoding typical proteins of retrotransposons [38], *Tcrm1* contains an ORF encoding typical RT and RNaseH; and (3) as for the only reported full-size LTR retrotransposon (*dem1*) [18], Yao *et al.* did not classify *dem1* into a subgroup but proved that *dem1* showed high homology with *dea1* and *del1*, which are *Ty3-gypsy* retrotransposons. Another full-size LTR retrotransposon in the GenBank database (AM167520) showed high homology with *dem1* by blast searching. This indicates that the only two full-size LTR retrotransposons (*dem1* and AM167520) belong to the *Ty3-gypsy* retrotransposons. However, unlike these full-size LTR retrotransposons, *Tcrm1* possesses the typical structure of the *Ty1-copia-like* retrotransposons and three distinct regions of LTR. In addition, compared with *dem1*, the identity of *Tcrm1* decreased to 41.38% on 145 nucleic acid sequences of RT regions and only 2.08% identity was obtained on deduced amino acid sequences of corresponding regions. Therefore, the differences in gene structure and sequence homology proved that *Tcrm1* is the first *Ty1-copia-like* retrotransposon including LTR domain sequences from the apple genome.

Copy number of *Tcrm1* in the apple

To estimate the copy number of *Tcrm1* in the apple genome, a 289 bp heterogeneous gene in the LTR region of *Tcrm1* was used as a probe for dot blot analysis. The haploid genome size for apple was 743–796 Mbp/C [44], and the average proportion of nuclear genomic DNA

hybridizing to the probe was 3.76×10^{-4} (determined by Launch VisionWorksLS gel software). Based on these data, and using the equation given in the "Materials and Methods" section, the result was as follows. *Tcrm1* copy number= $[(743\times10^{6})\times(3.76\times10^{-4})]/289=967$. If the haploid genome size for apple was 796 Mbp/C, *Tcrm1* copy number would be 1036. Therefore, the total copy number of *Tcrm1* in the apple genome was approximately 1×10^{3} copies per haploid genome.

Discussion

Compared to other PCR methods for isolating an unknown region adjacent to a known DNA sequence, SiteFinding-PCR is simple, fast, inexpensive, sensitive, efficient and easy to obtain long specific fragments from [35]. However, cloning target molecules still requires restriction cleavage, which is relatively complicated, expensive and time-consuming. In the present study, the process of cloning target molecules was simplified. After the nested PCR, the desired products were directly cloned into TA vectors rather than cleaved with *Not*I and then cloned into pBluescriptSK(+) vectors linearized by *Not*I and *Eco*RV. The other modifications of SiteFinding-PCR were to reduce concentrations of primers to 50% and decrease annealing temperatures to 60.2 ºC and 64 ºC in steps B and C, respectively. This modified SiteFinding-PCR, which was simpler, less expensive and faster than SiteFinding-PCR, succeeded in amplifying the RT-RNaseH-LTR domain sequences of *Ty1-copia-like* retrotransposons from the apple genome. The advantages of SiteFinding-PCR in terms of specificity, sensitivity and efficiency were maintained. The specific fragments obtained were shorter than those described by Tan *et al.* [35], possibly due to differences in species, AGGC numbers, PCR mixtures or cycling conditions.

Our study described a simple method that allowed isolation of LTR regions of retrotransposons from apple genomic DNA. It was developed on the basis of known RT genes. However, the method could be applied to genomes of any higher plant species, even if there is no genomic information adjacent to the LTR available. In that case, the sequence adjacent to the LTR can be easily isolated using the method of degenerate amplification, because the RT or RNaseH regions contain conserved motifs, thereby allowing degenerate primers to be designed. Apart from obtaining LTR sequences, the technique can be useful to isolate more genome sequences of higher plant species, such as rapid isolation of full gene sequences, cloning of promoters or regulatory elements from cDNA fragments, and amplification of DNA sequences of exon-intron boundaries.

A key point pertaining to this method is whether GCCT oligonucleotides at the 3′-end of the SiteFinder can initiate the SiteFinding reaction at a low temperature or not, which depends on GCCT numbers in the genome and on an appropriate annealing temperature. In fact, any 4, 5 or 6 nt oligonucleotides can be used to initialize the reaction at adjusted annealing temperatures in the SiteFinding reaction [35]. Therefore, there should be no limitations based on the length and quality of the genome sequences to using this technique in higher plants.

The future goals related to this work will be to further walk along both ends of *Tcrm* by applying modified SiteFinding-PCR, to detect retroelement insertional sites and insertional genes. At the same time, specific primers for *Tcrm1* LTR sequences will be designed to develop molecular marker systems including S-SAP and inter-retrotransposon amplified polymorphism (IRAP), which will reveal polymorphisms associated with this retroelement within the apple bud sports. These studies will be very useful to understanding the mechanism(s) of apple bud sports.

References

1 Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A. *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 1992, 20: 3639–3644

2 Flavell AJ, Smith DB, Kumar A. Extreme heterogeneity of *Ty1-copia* group retrotransposons in plants. Mol Gen Genet 1992, 231: 233–242

3 Voytas DF, Cummings MP, Koniczny A, Ausubel FM, Rodermel SR. copia-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 1992, 89: 7124–7128

4 Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS, Waugh R, Flavell AJ. The *Ty1-copia* group of retrotransposons in plants: Genomic organisation, evolution, and use as molecular markers. Genetica 1997, 100: 205–217

5 Feschotte C, Jiang N, Wessler SR. Plant transposable elements: Where genetics meets genomics. Nat Rev Genet 2002, 3: 329–341

6 Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 1990, 9: 3353–3362

7 VanderWiel PL, Voytas DF, Wendel JF. *Copia*-like retrotransposable element evolution in diploid and polyploid cotton (*Gossypium* L.). J Mol Evol 1993, 36: 429–447

8 Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M. Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 1996, 93: 7783–7788

9 Kumekawa N, Ohtsubo E, Ohtsubo H. Identification and phylogenetic analysis of gypsy-type retrotransposons in the plant kingdom. Genes Genet Syst 1999, 74: 299–307

10 Hirochika H, Okamoto H, Kakutani T. Silencing of retrotransposons in

©Institute of Biochemistry and Cell Biology, SIBS, CAS

arabidopsis and reactivation by the *ddm1* mutation. Plant Cell 2000, 12: 357–369

11 Bennetzen JL. Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 2000, 42: 251–269

12 Kumar A, Hirochika H. Applications of retrotransposons as genetic tools in plant biology. Trends Plant Sci 2001, 6: 127–134

13 Casacuberta JM, Santiago N. Plant LTR-retrotransposons and MITEs: Control of transposition and impact on the evolution of plant genes and genomes. Gene 2003, 311: 1–11

14 Kumar A, Bennetzen JL. Plant retrotransposons. Annu Rev Genet 1999, 33: 479–532

15 Grandbastien MA, Spielmann A, Caboche M. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 1989, 337: 376–380

16 Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis TH, Flavell AJ. Rapid isolation of plant *Ty1-copia* group retrotransposon LTR sequences for molecular marker studies. Plant J 1999, 19: 711–717

17 Kilstrup M, Kristiansen KN. Rapid genome walking: A simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. Nucleic Acids Res 2000, 28: E55

18 Yao J, Dong Y, Morris BA. Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. Proc Natl Acad Sci USA 2001, 98: 1306–1311

19 Pelsy F, Merdinoglu D. Complete sequence of *Tvv1*, a family of Ty 1 copia-like retrotransposons of *Vitis vinifera* L., reconstituted by chromosome walking. Theor Appl Genet 2002, 105: 614–621

20 Holligan D, Zhang X, Jiang N, Pritham EJ, Wessler SR. The transposable element landscape of the model legume Lotus japonicus. Genetics 2006, 174: 2215–2228

21 Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 1997, 253: 687–694

22 Kalendar R, Grob T, Regina M, Suoniemi A, Schulman AH. RAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 1999, 98: 704–711

23 Breto MP, Ruiz C, Pina JA, Asins MJ. The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species. Mol Phylogenet Evol 2001, 21: 285–293

24 Garcia-Martinez J, Martinez-Izquierdo JA. Study on the evolution of the grande retrotransposon in the zea genus. Mol Biol Evol 2003, 20: 831–841

25 Queen RA, Gribbon BM, James C, Jack P, Flavell AJ. Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Mol Genet Genomics 2004, 271: 91–97

26 Labra M, Imazio S, Grassi F, Rossoni M, Sala F. *Vine-1* retrotransposon-based sequence-specific amplified polymorphism for *Vitis vinifera* L. genotyping. Plant Breeding 2004, 123: 180–185

27 Venturi S, Dondini L, Donini P, Sansavini S. Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. Theor Appl Genet 2006, 112: 440–444

28 Ellis TH, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ. Polymorphism of insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 1998, 260: 9–19

29 Tahara M, Aoki T, Suzuka S, Yamashita H, Tanaka M, Matsunaga S, Kokumai S. Isolation of an active element from a high-copy-number family of retrotransposons in the sweetpotato genome. Mol Genet Genomics 2004, 272: 116–127

30 Lee D, THN Ellis, L Turner, RP Hellens, WG Cleary. A copialike element in Pisum demonstrates the use of dispersed repeated sequences in genetic analysis. Plant Mol Biol 1990, 15: 707–722

31 Manninen I, Schulman AH. BARE-1, a *copia*-like retroelement in barley (Hordeum *vulgare* L.). Plant Mol Biol 1993, 22: 829–846

32 Hirochika H, Hirochika R. *Ty1-copia* group retrotransposons as ubiquitous components of plant genomes. Jpn J Genet 1993, 68: 35–46

33 Le QH, Wright S, Yu Z, Bureau T. Transposon diversity in Arabidopsis thaliana. Proc Natl Acad Sci USA 2000, 97: 7376–7381

34 McCarthy EM, McDonald JF. Long terminal repeat retrotransposons of Mus musculus. Genome Biol 2004, 5: R14

35 Tan G, Gao Y, Shi M, Zhang X, He S, Chen Z, An C. SiteFinding-PCR: A simple and efficient PCR method for chromosome walking. Nucleic Acids Res 2005, 33: e122

36 Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-induced mutations in grape skin color. Science 2004, 304: 982

37 Sun J, Fang JG, Zhang Z. Partial sequence of reverse transcriptase of *Ty1-copia-like* retrotransposon in apple. Journal of Fruit Science 2005, 22: 193–197

38 Antonius-Klemola K, Kalendar R, Schulman AH. TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports. Theor Appl Genet 2006, 112: 999–1008

39 Lodhi MA, Ye GN, Weeden NF, Reisch BI. A simple and efficient method for DNA extraction from grapevine cultivars *Vitis* species and Ampelopsis. Plant Molecular Biology Reporter 1994, 12: 6–13

40 MacRae AF. A pentamer-repeat-containing DNA sequence in *Texas bluebonnet* (*Lupinus texensis* Hook.) Genome 1998, 41: 553–559

41 Temin HM. Structure, variation and synthesis of retrovirus long terminal repeat. Cell 1981, 27: 1–3

42 Joshi CP. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 1987, 15: 6643–6653

43 Mogen BD, MacDonald MH, Graybosch R, Hunt AG. Upstream sequences other than AAUAAA are required for efficient messenger RNA 3′-end formation in plants. Plant Cell 1990, 2: 1261–1272

44 Arumuganathan K, Earle ED. Estimation of nuclear DNA content of plants by flow cytometry. Plant Mol Biol Rep 1991, 9: 229–233

Edited by
Minghua XU