African Journal of Biotechnology Vol. 9(50), pp. 8583-8596, 13 December, 2010
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB10.1259
ISSN 1684–5315 © 2010 Academic Journals

Full Length Research Paper

Genome-wide detection of *Ty1-copia* and *Ty3-gypsy* group retrotransposons in Japanese apricot (*Prunus mume* Sieb. et Zucc.)

Fei Wang[1#], Zhaoguo Tong[1#], Jun Sun[2], Yuying Shen[3], Jun Zhou[4], Zhihong Gao[1*] and Zhen Zhang[1*]

[1]College of Horticulture, Nanjing Agricultural University, 1 Tongwei Road, Nanjing 210095, People's Republic of China.
[2]College of Horticulture, Anhui Agricultural University, 130 West Changjiang Road, Hefei 230036, People's Republic of China.
[3]Zhejiang College of Construction, Hangzhou 311231, People's Republic of China.
[4]Southwest Forestry University, Bailongsi Road, Kunming 650224, People's Republic of China.

Accepted 19 October, 2010

The conserved domains of reverse transcriptase (RT) genes of approximately 260 bp for *Ty1-copia* and 430 bp for *Ty3-gypsy* groups of long terminal repeat (LTR) retrotransposons were amplified from Japanese apricot (*Prunus mume* Sieb. et Zucc.) using degenerate oligonucleotide primers. Sequence analysis showed that 32.3% of *Ty1-copia* and 27.5% of *Ty3-gypsy* RT sequences possessed stop codons and/or frameshifts, and all sequences were AT-rich. *Ty1-copia* retrotransposon has higher heterogeneity than *Ty3-gypsy* retrotransposon, but the latter has a higher copy number revealed by southern dot blot hybridization. Phylogenetic analysis illustrated that some of the clones were more closely related to the representative elements present in other plant species than to other clones of Japanese apricot. Transcription was not detected by reverse transcription-polymerase chain reaction amplification for either *Ty1-copia* or *Ty3-gypsy* retrotransposons in young leaves of plants treated with UV light and 2,4-dichlorophenoxyacetic acid either individually or in both combinations, even though the ratios of dN/dS of the open reading frames among members of each subgroup of both group retrotransposons were less than 1. This is the first report on the presence of RT sequences of *Ty1-copia* and *Ty3-gypsy* group retrotransposons in Japanese apricot genome.

Key words: *Prunus mume*, retrotransposon, heterogeneity, copy number, transcriptional activity.

INTRODUCTION

Transposable elements (TEs) are mobile elements of DNA that can move and replicate within their host genomes and have been found to be widely prevalent in most eukaryotes. According to the trait of transposition mediated by either RNA or DNA, TEs could be separated into two principal classes (Wicker et al., 2007).

Class I elements or retrotransposons, are the most common type of mobile genetic element (Flavell et al., 1992) and tran-spose through an RNA intermediate by a "copy and paste" mechanism (Grandbastien, 1992). Class II elements, or DNA transposons, move by a "cut and paste" mecha-nism involving excision and reintegration via a DNA intermediate (Bowen and Jordan, 2002). Additionally, class I can also be classified into two major groups: Long terminal repeat (LTR) and non-LTR retrotransposons depending on the presence or absence of LTRs (Kapitonov and Jurka, 2008). In plants, LTR retrotransposons are further subdivided into the *Ty1-*

*Corresponding author. E-mail: gaozhihong@njau.edu.cn. Tel: +86 25 84395724. Fax: +86 25 84395724.

Abbreviations: TEs, Transposable elements; **LTR,** long terminal repeat; **POL,** polyprotein; **PR-IN-RT,** protease-integrase-reverse transcriptase; **2,4-D,** 2,4-dichlorophenoxyacetic acid; **PCR,** polymerase chain reaction; **RT,** reverse transcriptase; **UV,** ultraviolet; **dN,** average pairwise values of the number of nonsynonymous substitutions per nonsynonymous site; **dS,** number of synonymous substitutions per synonymous; **NCBI,** National Center for Biotechnology Information.

#These authors contributed equally to this work.

copia and *Ty3-gypsy* groups based on sequence similarity and the order of genes coding for the polymerase function within poly-protein (POL) region (Wilhelm and Wilhelm, 2001). *Ty1-copia* elements have a *pol* gene organized as protease-integrase-reverse transcriptase (PR-IN-RT), whereas the *pol* gene of *Ty3-gypsy* elements is organized as PR-RT-IN (Havecker et al., 2004).

It is documented that retrotransposons contribute to increasing genome size in the plant kingdom, and play important roles in the evolution of genome size. The great differences in genome sizes of plant species are putatively attributed to the presence of different amounts of retrotransposons. Generally, the larger genome of plants may have a capacity to maintain the large copies of retrotransposons. For instance, the 5000-Mb barley genome includes at least 70% of these elements (Vicient et al., 2001) while in smaller rice genome (430 Mb) they represent only 17% of its genome (McCarthy et al., 2002). More recently, the different size of two orthologous regions of barley and rice had been proved to be mainly caused by the presence of LTR retrotransposons, solo-LTRs and their derivatives in these regions (Park et al., 2004).

In general, most retrotransposons are inactive or transcriptionally silent during normal plant growth and development. However, biotic and abiotic stresses could increase transcript levels of some characterized retrotransposons. For example, *Tnt1* and *Reme1* elements, the first found in tobacco and the last in melon species, can be induced at specific stages by abiotic stress (e.g. jasmonic acid, salicylic acid, 2,4-dichlorophenoxyacetic acid (2,4-D) and UV irradiation) or biotic stress (e.g. cucumber mosaic virus (CMV)) (Beguiristain et al., 2001; Grandbastien et al., 1997; Ramallo et al., 2008). In addition, the element of *Tos17* within the genus *Oryza* is expressed under tissue culture condition. Retrotransposons insertions near or within genes are known to be involved in generating mutations as a result of alterations in gene expression and/or structure of the resulting gene product (Kobayashi et al., 2004). Although some transposable element insertions can have a deleterious impact on the host genome, other insertions could benefit the host, providing genetic variabilities or mediating favorable structural changes in the genome that increase host fitness (Kidwell and Lisch, 2000). Therefore, transposition and amplification of retrotransposons are considered to be powerful contributions to plant gene, genome structure and evolution (Bennetzen, 2000).

Japanese apricot, a deciduous tree of the genus *Rosaceae*, also known as mei and ume, originated in the southwest China, distributed with more than 400 varieties worldwide, and has been widely cultivated in China and Japan. The fruits of Japanese apricot have a long history of use as traditional foods and beverages, such as ume-jam, pickled Japanese apricot, ume-juice, ume-candies and so on (Chu, 1999). Moreover, the fruits have also been proved to be useful to cure or prevent many kinds of diseases. For instance, it has been widely used in folk remedies and as an herbal medicine to alleviate fever, laxness, cough and intestinal disorder (Shi et al., 2009), improve human blood fluidity (Chuda et al., 1999) and inhibit the proliferation of cancer cells (Jenog et al., 2006). The need to improve Japanese apricot fruit as a nutraceutical with respect to a range of characteristics is now recognized, however, there are no reports on the existence of TEs in the Japanese apricot genome. The genome characterization with reference to the content, heterogeneity, activity and overall distribution of retrotransposons might contribute to our understanding about *Prunus mume* genome organization and its evolution. Moreover, identification of TEs can help to create new molecular markers which could be used in fruit quality selection.

The advent of polymerase chain reaction (PCR) techniques using primers designed from conserved enzyme domains have facilitated quick and extensive surveys of TEs in many plants. Reverse transcriptase (RT) is the most conserved domain in retrotransposon sequences and it is commonly used for phylogenic analysis and identification of new retrotransposons. In the current study, our aim include to document the presence of RT sequences of *Ty1-copia* and *Ty3-gypsy* group retrotransposons in Japanese apricot genome and investigate their sequence heterogeneity, copy numbers, phylogenetic relationships and transcriptional activity.

MATERIALS AND METHODS

Plant material and isolation of nucleic acids

The young leaves of Japanese apricot (*P. mume* 'koume') were collected from 8-year-old trees growing in national germplasm orchard of College of Horticulture, Nanjing Agricultural University, Nanjing, Jiangsu, China. Total DNA, used for PCR amplification and Southern dot blotting, was extracted according to the method of Lodhi et al. (1994).

Potted Japanese apricot young trees were either irradiated with a 36-W germicidal lamp that emitted predominantly ultraviolet (UV) light of wavelength 254 nm at a distance of 40 cm or sprayed with 2.0 mg/l 2,4-D at room temperature. The combined experiment, in which both treatments were applied simultaneously, was also conducted at the same time. The durations for the three treatments were 0, 12, 24, 48 and 96 h. About two leaves, closely around the terminal buds were sampled for the isolation of total RNA according to the method described by Gambino et al. (2008). Genomic DNA was eliminated by treating each RNA sample with RNase-free DNase□ (TaKaRa, Japan) according to the instruction manual.

Amplification of RT fragments of retrotransposons

The degenerate primers, Ty1-F: 5'-ACNGCNTTYYTNCAYGG-3', Ty1-R: 5'-ARCA TRTCRTCNACRTA-3' (Kumar et al., 1997) and Ty3-F: 5'-AGMGRATGTGYGTSGAYTAT-3', Ty3-R: 5'-CAMCCMR AAMWCACAMTT-3' (Kumekawa et al., 1999) (N=A+T+C+G, Y=T+C, R=A+G, W=A+T, M=A+C) were used to amplify RT domains of *Ty1-copia* and *Ty3-gypsy* group LTR retrotransposons,

respectively. The PCR amplification reaction contained 1 × PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTP, 0.8 µM of each primer of *Ty1-copia* or *Ty3-gypsy* primer pairs, 1 U of rTag DNA polymerase (TaKaRa, Japan) and 100 ng genomic DNA. The final volume was brought to 50 µl with sterile distilled water. PCR cycling conditions comprised an initial denaturation step at 94°C for 5 min, followed by 30 cycles at 94°C for 60 s, 52°C (*Ty1-copia*) or 50°C (*Ty3-gypsy*) for 90 s and 72°C for 60 s, and then a final elongation step at 72°C for 10 min.

PCR products for *Ty1-copia* and *Ty3-gypsy* were separated by electrophoresis on 2.0% agarose gels in 1 × TAE (4 mM Tris–acetate, 1 mM EDTA) buffer and visualized under UV light after staining with ethidium bromide.

Cloning and sequencing

The PCR products were purified with agarose gel DNA purification kit (TaKaRa, Japan), cloned in the pGEM-T Easy vector (Promega, USA) according to the manufacturer's instructions, and then transformed into DH5α strain of *Escherichia coli*. Plasmids DNA containing the expected insert were extracted and sequenced. DNA sequencing reactions were carried out by Shanghai Invitrogen Biological Co. Ltd., Shanghai, China.

Sequence analysis

The RT sequences were named by the letters PM and PMG for *Ty1-copia* and *Ty3-gypsy* group retrotransposons, respectively, and given a corresponding number. The RT nucleotide sequences were translated into putative amino acids with the help of the Transeq tool at http://bips.u-strasbg.fr/EMBOSS/. The frameshift mutations were detected using ERRWISE (http://coot.embl.de/ERR_WISE/). The deduced amino acid sequences of RT regions with frameshifts and/or stop codons were discarded from the subsequent analyses. The putative amino acid sequences of the RT domains were aligned using the program ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html) and displayed with GeneDoc (http://www.psc.edu/biomed/genedoc). The *Ty1-copia* and *Ty3-gypsy* group RT sequences of other plant species are given in Table 1, along with their database accession numbers, organisms and sequence definitions. A comparative phylogenetic tree for each group LTR retrotransposons was created based on amino acid sequences of the RT domains, using MEGA 4.0 software by the *p*-distance and Neighbor-Joining methods with 1000 replication bootstrap tests (Tamura et al., 2007). Alignments of the deduced amino acid sequences for each group of *Ty1-copia* and *Ty3-gypsy* retroelements, used for analysis of synonymous and nonsynonymous substitution rates, were made with T-Coffee web server at http://www.tcoffee.org/ (Poirot et al., 2003). These alignments were then used to generate the alignment of the corresponding nucleotide sequences using PAL2NAL (http://www.bork.embl.de/ pal2nal) (Suyama et al., 2006). After being converted into Phylip format, the average pairwise values of the number of nonsynonymous substitutions per nonsynonymous site (dN), the number of synonymous substitutions per synonymous (dS) and the ratios of the dN/dS were calculated by the yn00 program of the PAML 4.2b package (Yang and Nielsen, 2000) for each subgroup of the two types of LTR retrotransposon elements.

Southern dot blot analysis

Heterogeneous PCR products, *Ty1-copia* and *Ty3-gypsy* LTR retrotransposon RT domain sequences, were used as probes, which were labeled using digoxigenin-dUTP following the manufacturer's instruction manual of DIG High Primer DNA Labeling and Detection Starter Kit I (Roche, Germany). Genomic DNA and heterogeneous PCR products were serially diluted, subsequently denatured by 1 M NaOH + 0.2 M EDTA (pH = 8.0) for 15 min and in water bath at 100°C for 10 min, and then quickly chilled in an ice-water bath for 3 to 5 min. The denatured genomic DNA and PCR products were transferred onto Immobilon-Ny^+ membranes (Millipore, USA). On one Immobilon-Ny^+ membrane, the genomic DNA was diluted to 80, 100, 200 and 400 ng, and PCR product of the *Ty1-copia* RT domain was diluted to 400, 500, 800 and 1000 pg (Figure 1a). On the other Immobilon-Ny^+ membrane, the genomic DNA was diluted to 50, 80, 100 and 200 ng, and PCR product of the *Ty3-gypsy* was diluted to 250, 400, 500 and 800 pg (Figure 1b). Hybridization was carried out at 42°C for 20 h in DIG Easy hybridization liquid with the hybridization furnace (OV4, Biometra, Germany). The filters were washed and visualized by immunological detection according to the methods described in the instruction manual of DIG High Primer DNA Labeling and Detection Starter Kit I. The hybridization signals were subsequently detected and analyzed with Life Science Research Products and System Engineering (Shanghai FuRi Science and Technology Co. Ltd.). The linear regression equation relating the natural logarithm of the copy number of probe sequences in the dilution spots and the corresponding densitometric readings was used to estimate the copy number of the sequences probed in the samples of genomic DNA.

Reverse transcription polymerase chain reaction (RT-PCR)

One microgram RNA was reverse-transcribed using Reverse Transcriptase M-MLV (RNase H^-) (TaKaRa, Japan) for first-strand cDNA synthesis with 2.5 µM oligonucleotide dT primer and 5 µM random hexamer priming method according to the manufacturer's recommendations. Each RNA sample was controlled for genomic DNA contamination without reverse transcriptase addition into cDNA synthesis mixture. PCR was performed using the same degenerate primers and cycling conditions of amplifying RT genes from genomic DNA. In addition, the control gene of *Actin* was used in PCR reactions on all RNA templates as a positive control with the forward primer 5'-CAATGTGCCTGCCATGTATG-3' and reverse primer 5'-CCAGCAGCTTCCATTCCAAT-3' (Zhang et al., 2007). At least three different RNA isolations and cDNA syntheses were used as replicates for PCR reactions.

Nucleotide sequence accession numbers

The nucleotide sequences of Japanese apricot RT clones determined in this study have been deposited in the National Center for Biotechnology Information (NCBI) GenBank database under accession numbers DQ494215-DQ494253 and GU143560-GU143582 for *Ty1-copia* group LTR retrotransposons, GQ372892-GQ372916 and GU143583-GU143597 for *Ty3-gypsy* group LTR retrotransposons.

RESULTS

Detection of *Ty1-copia* RT fragments from Japanese apricot genome

Ty1-copia group RT sequences from the cultivar of 'koume' genomic DNA were isolated by PCR with degenerate primers, Ty1-F and Ty1-R, of which the sequences correspond to the well conserved amino acid motifs of TAFLHG and YVDDML, respectively. The length of the fragment approximated to 260 bp that was similar to the expected size (data not shown). A total of 70 clones

Table 1. List of RT sequences of LTR retrotransposons from other plant species used in current study for comparative analysis.

Sequence name	Organisms	Sequence definitions	Accession No	Type
Apple	*Malus × domestica*	Reverse transcriptase	ABS11055	Copia
Brown mustard	*Brassica nigra*	Reverse transcriptase	CAD11834	Copia
Chickpea	*Cicer arietinum*	*Ty1-copia* retrotransposon partial pol pseudogene, clone cart384	AM283486	Copia
Grape	*Vitis vinifera*	Hypothetical retrotransposon protein	CAN67792	Copia
Himalayan poplar	*Populus ciliata*	Reverse transcriptase	AAT73704	Copia
Maidenhair tree	*Ginkgo biloba*	Reverse transcriptase	AAA33351	Copia
Norway spruce	*Picea abies*	Reverse transcriptase	CAA11921	Copia
Perennial kale	*Brassica oleracea*	Reverse transcriptase	CAD11838	Copia
Potato	*Solanum tuberosum*	Probable RNA-directed DNA polymerase, *Ty1-copia*-like retrotransposon (fragment)	S20016	Copia
Pumello	*Citrus maxima*	Reverse transcriptase	CAJ41402	Copia
Reme1	*Cucumis melo*	Putative reverse transcriptase sequence of melon	CAJ65852	Copia
Strawberry	*Fragaria × ananassa*	*Ty1-copia*-like retrotransposon, partial sequence of clone F8	EF429149	Copia
Sweet potato	*Ipomoea batatas*	Gag-Pol protein	BAD34493	Copia
Tnt1	*Nicotiana tabacum*	Tobacco DNA sequence for retroviral-like transposon *Tnt* 1-94	X13777	Copia
Tomato	*Lycopersicon esculentum*	*Ty1-copia* class retrotransposon fragment (clone RT6)	T06321	Copia
Trifoliate orange	*Poncirus trifoliata*	RNA-directed RNA polymerase	CAJ85621	Copia
African oil palm	*Elaeis guineensis*	Reverse transcriptase	CAD45567	Gypsy
Apple	*Malus× domestica*	Clone GyRTFJ-43 retrotransposon *Ty3-gypsy* reverse transcriptase gene, partial cds	EF623978	Gypsy
Barrel medic	*Medicago truncatula*	Integrase, catalytic region; Ribonuclease H	ABD28291	Gypsy
Broomrape	*Orobanche crenata*	Reverse transcriptase	ABD43107	Gypsy
Grape	*Vitis vinifera*	Gag-Pol precursor	BAD18986	Gypsy
Maidenhair tree	*Ginkgo biloba*	Reverse transcriptase	CAA12930	Gypsy
Marine alga	*Porphyra yezoensis*	*Ty3/gypsy*-like retrotransposon *PyRE3* gene for reverse transcriptase, partial cds	AB108540	Gypsy
Mungbean	*Vigna radiata*	Clone VRG-7 *gypsy*-like retrotransposon, partial sequence	AY683009	Gypsy
Muskmelon	*Cucumis melo*	Pol protein	AAO45752	Gypsy
Pea	*Pisum sativum*	Gag/pol polyprotein	AAQ82033	Gypsy
Pineapple	*Ananas comosus*	Probable retrotransposon polyprotein *dea1* (fragment)	T07863	Gypsy
Potato	*Solanum demissum*	Putative gag-pol polyprotein	AAT38744	Gypsy
Rice	*Oryza sativa* (japonica cultivar-group)	Similar to sorghum bicolor 22 kDa kafirin cluster; polyprotein	BAA95869	Gypsy
Strawberry	*Fragaria × ananassa*	*Ty3-gypsy*-like retrotransposon reverse transcriptase gene (clone 6), partial sequence	EF443063	Gypsy
Thale cress	*Arabidopsis thaliana*	Reverse transcriptase	BAB40828	Gypsy

of the 260 bp PCR products were randomly selected for sequencing, and the results showed that 62 sequences were unique, and the actual sizes ranged from 259 to 269 bp. The nucleotide sequences of these clones were very similar to many other plant *Ty1-copia* group RT sequences,

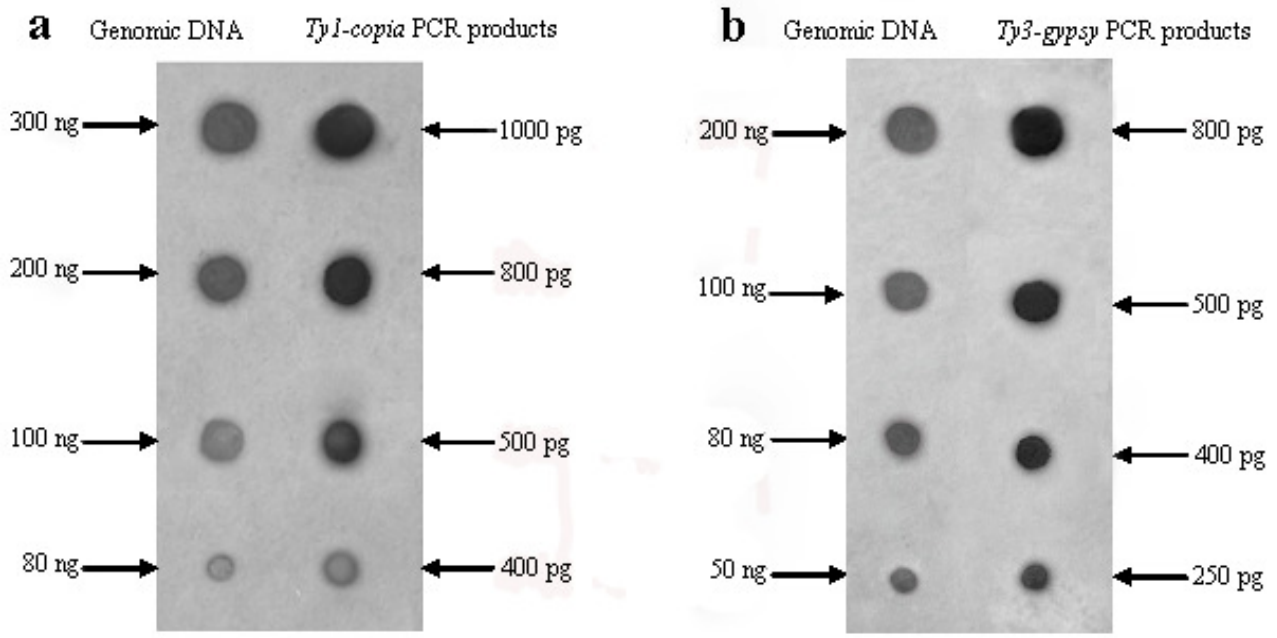

Figure 1. Determination of the total copy number of LTR retrotransposons in the *P. mume* genome. **a.** Determination of the total copy number of *Ty1-copia* group retrotransposons in the Japanese apricot genome. Genomic DNA corresponding to 80, 100, 200 and 400 ng, and PCR product of the *Ty1-copia* RT domain, corresponding to 400, 500, 800 and 1000 pg, were fixed to the filter. The filter was probed with a heterogeneous no-radioactive labeled PCR probe. **b.** Determination of the total copy number of *Ty3-gypsy* group retrotransposons in the Japanese apricot genome. Genomic DNA corresponding to 50, 80, 100 and 200 ng, and PCR product of the *Ty3-gypsy* RT domain, corresponding to 250, 400, 500 and 800 pg, were fixed to the filter. The filter was probed with a heterogeneous no-radioactive labeled PCR probe.

which confirmed that they were authentic. In addition, these fragments were found to be rich in the bases of A and T, with the ratio of AT/GC ranging between 1.11 and 1.86 for PM 28 and PM 24, respectively. All the RT sequences were conceptually translated into amino acids and analyzed for the presence of stop codons and/ or frameshifts in their coding regions. Of the 62 unique sequences, 20 (32.3%) contained premature stop codons and/or frameshifted translations, while the remaining 42 (67.7%) sequences had potentially functional RT fragments. Alignment of the 42 putative amino acid sequences showed that most of them contained characteristic peptide motifs, TAFLHG and LYGLKQ, and they all contained YVDDM, located in 5'-end, central region and 3'-end of the *Ty1-copia* group RT genes, respectively (Nakatsuka et al., 2002; Sun et al., 2008) (Figure 2). Pairwise comparisons showed that the average amino acid sequence heterogeneity was 45.5%, with a range between 0 and 60.2%. Despite the fact that two pair sequences, PM 29, PM 32 and PM 21, PM 57, shared 100% sequence identity between each pair at amino acid sequence level, the difference in their nucleotide sequences accounted for 2.7 and 1.9%, respectively. A NJ-tree based on *P*-distance and supported with 1000 replicates of bootstrapping was reconstructed for the *Ty1-copia* group RT sequences obtained from this study and other plant species identified from the GenBank database (Figure 3). All the RT sequences could be classified into six major subgroups, designated as SG1-SG6, respectively, and the amino acid identities between individual sequences belonging to the same subgroup ranged from 60.9 to 95.4% (SG1), 56.3 to 98.9% (SG2), 94.2 to 100% (SG3), 67.4 to 98.9% (SG4), 68.6 to 100% (SG5) and 60.2 to 98.9% (SG6). The RT sequences of other plant species could be clustered into the same subgroup with some sequences of Japanese apricot. For example, the sequences of Norway spruce (*Picea abies*), tomato (*Lycopersicon esculentum*), Himalayan poplar (*Populus ciciata*), tobacco (*Nicotiana tabacum*), melon (*Cucumis melo*), potato (*Solanum tuberosum*) and sweet potato (*Ipomoea batatas*) comprised the subgroup (SG6) with PM 2, PM 17, PM 24, PM 39, PM 44, PM 49, PM 51, PM 54, PM 55 and PM 56. However, some of these *P. mume* sequences were assigned to different subgroups with low identical residues, such as the case of two sequences PM 12 and PM 44, which had a divergence value of 60.2%, and were placed in two separate subgroups, SG2 and SG6, respectively.

Detection of *Ty3-gypsy* RT sequences from Japanese apricot genome

Using degenerate PCR primers, Ty3-F and Ty3-R, *Ty3-*

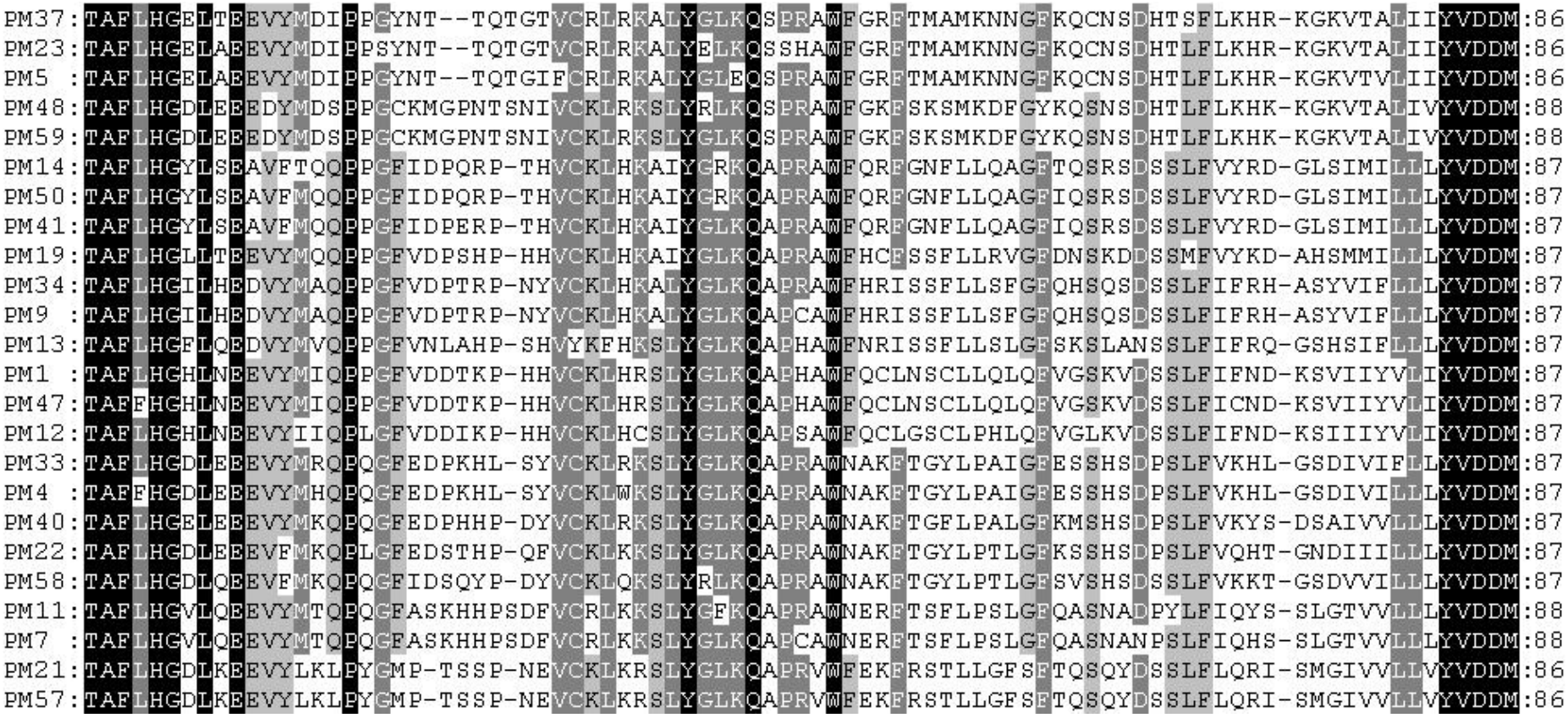

Figure 2. Sequence alignments of deduced amino acids corresponding to the RT domains from *Ty1-copia* group RT sequences of *P. mume*. The numbers of amino acid residues are shown on the right hand of each sequence. Gaps are indicated as (-) and consensus sequence are shown below. The four shading levels indicate degree of residue conservation within a column: black (100% conserved), dark gray (80% conserved or greater), light gray (60% conserved or greater) and no shading (< 60% conserved).

gypsy RT fragment with the expected size of about 430 bp was amplified (data not shown), recovered and 50 clones were randomly selected for sequencing. After discarding the repetitive sequences, the remaining 40 unique nucleotide sequences were used to query the GenBank database, and the results showed that these sequences were very similar to the *Ty3-gypsy* group retrotransposons of other species. Each sequence was also abundant with the bases of A and T, with the ratio of AT/GC varying from 1.01 (PMG 8) to 1.95 (PMG 3). Among the 40 sequences, 11 (27.5%) sequences presented premature stop codons and/or frame shift mutations within their translation frames, and thus the remaining 29 sequences were unaffected by either stop codons or frameshifts when translated. Alignment of the 29 putative amino acid sequences showed that most of them contained the peptide motif VMPFGL in the central region of the *Ty3-gypsy* RT gene, but at both ends they presented some variations in their translated primer sequences, especially in 3'-end (Figure 4). The overall average amino acid divergence between individual *Ty3-gypsy* RT sequences was 39.9%, lower than that of *Ty1-copia* group retrotransposons. A NJ-tree based on *p*-distance and

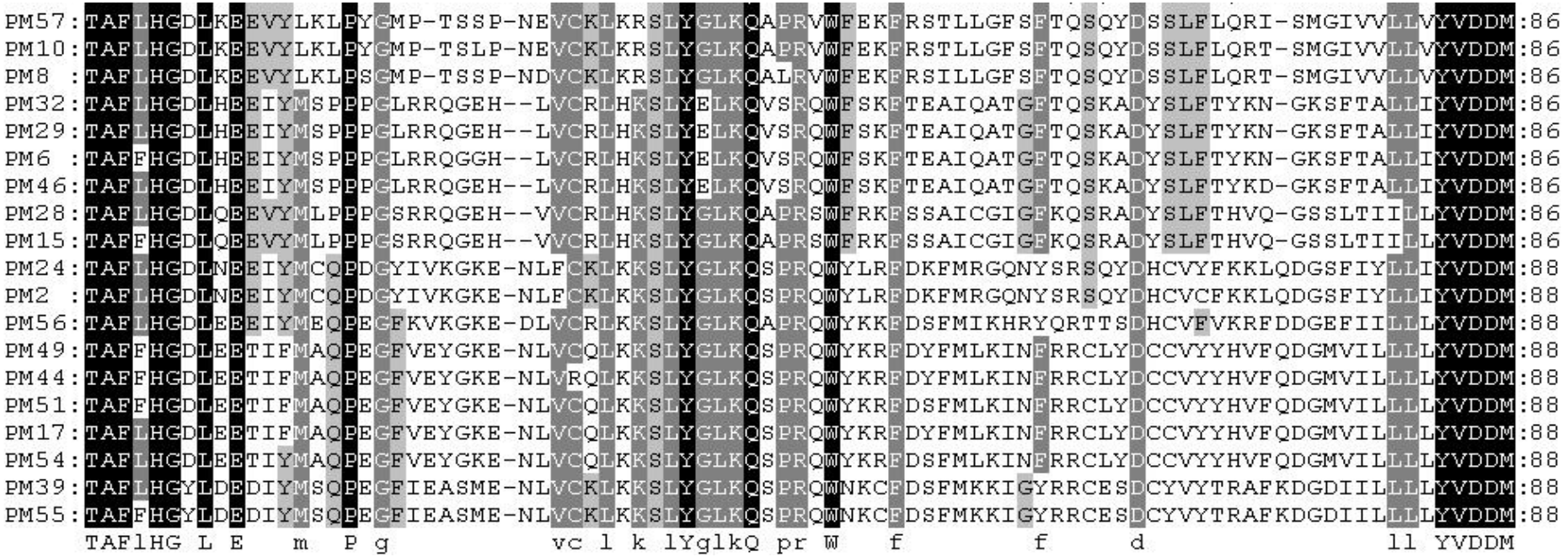

Figure 2. Contd.

supported with 1000 replicates of bootstrapping was generated for the *Ty3-gypsy* group amino acid RT sequences obtained from this study and other plant species. The phylogenetic relationships among these sequences are shown in Figure 5. All the RT sequences could be separated into three distinct subgroups, named as SF1-SF3, respectively, and the amino acid identities between individual sequences belonging to the same family ranged from 82.5 to 100% (SF1), 51 to 100% (SF2) and 50.3 to 100% (SF3).

Synonymous and nonsynonymous substitution analysis

The numbers of synonymous and nonsynonymous substitutions per site were estimated for the RT sequences of each subgroup of both types of retroelements. Table 2 showed that all the dN/dS ratios were less than 1, denoting that purifying selection had acted on the LTR retrotransposon elements in the *P. mume* genome.

Copy number

To estimate the number of *Ty1-copia* and *Ty3-gypsy* group retrotransposon copies in Japanese apricot, two heterogeneous populations of 260 bp fragments and 430 bp fragments from the RT genes were used as probes in dot blot hybridization analysis according to the method described by MacRae (1998) (Figure 1). The copy numbers for the two types of LTR retrotransposons were calculated with the following equation:

Copy number = (Size of genome (bp) × Average proportion of nuclear genomic DNA hybridizing to probe) / Size of probe (bp)

The genome size of *Prunus* species is about 300 Mb (1C= 0.3 pg), and thus the copy numbers of *Ty1-copy* and *Ty3-gypsy* group retrotransposons is about 7.9×10^3 and 1.0×10^4 molecules per genome, respectively. Assuming average sizes of *Ty1-copia* and *Ty3-gypsy* group LTR retrotransposons are 7 and 10 kb, respectively (Hill et al., 2005), both two groups of LTR retrotransposons account for approximately 52% of the *P. mume* genome (*Ty1- copia* = 18.4%, *Ty3-gypsy*

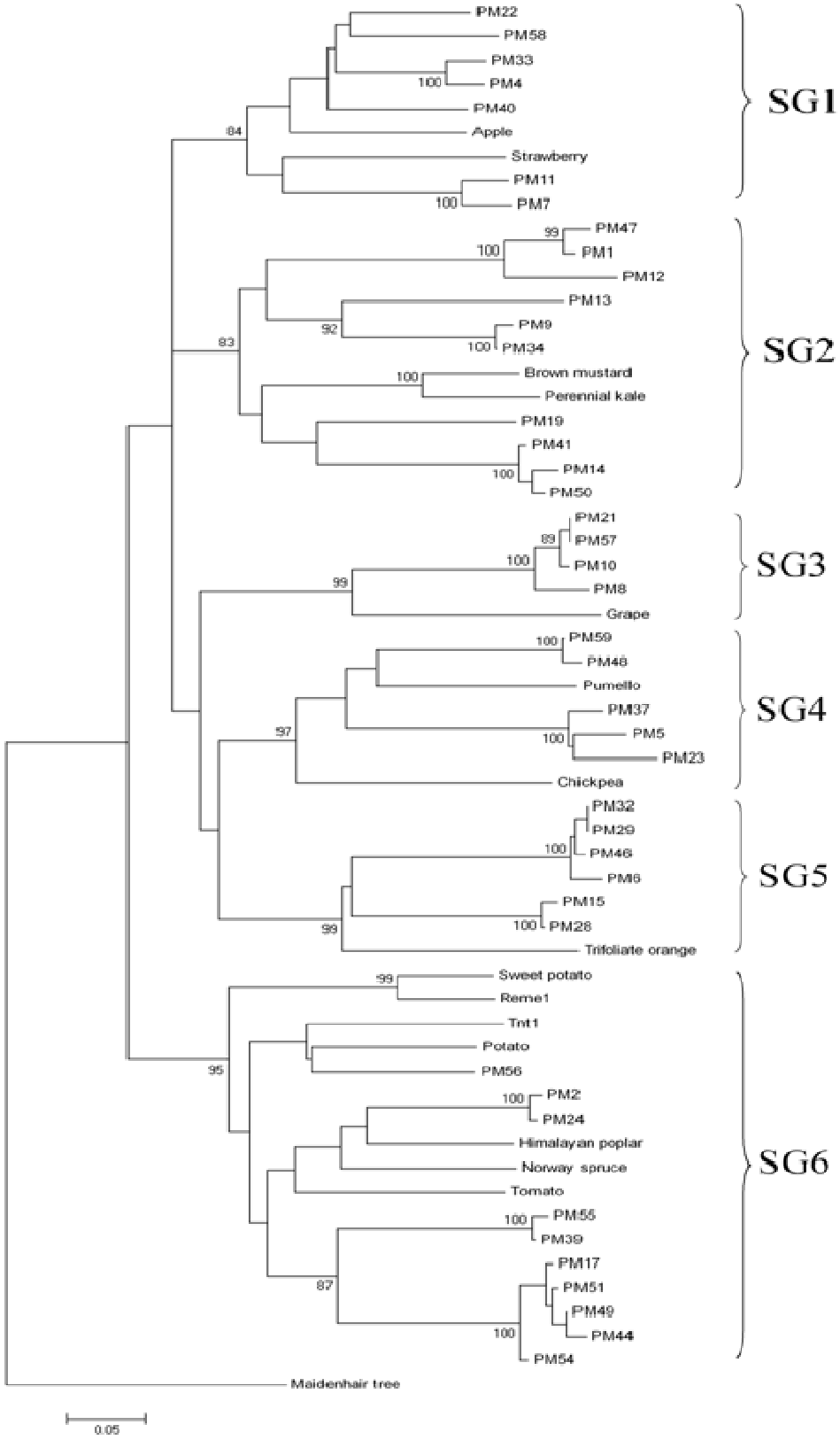

Figure 3. Phylogenetic relationships among the *Ty1-copia* RT sequences of *P. mume* and other plant species. Bootstrap values of 80% or more are indicated at the branch nodes. The *Ty1-copia* RT sequence of Maidenhair tree is used as an outgroup to root the tree. Bar represents 5% sequence divergence. See Table 1 for details of other plant species sequences.

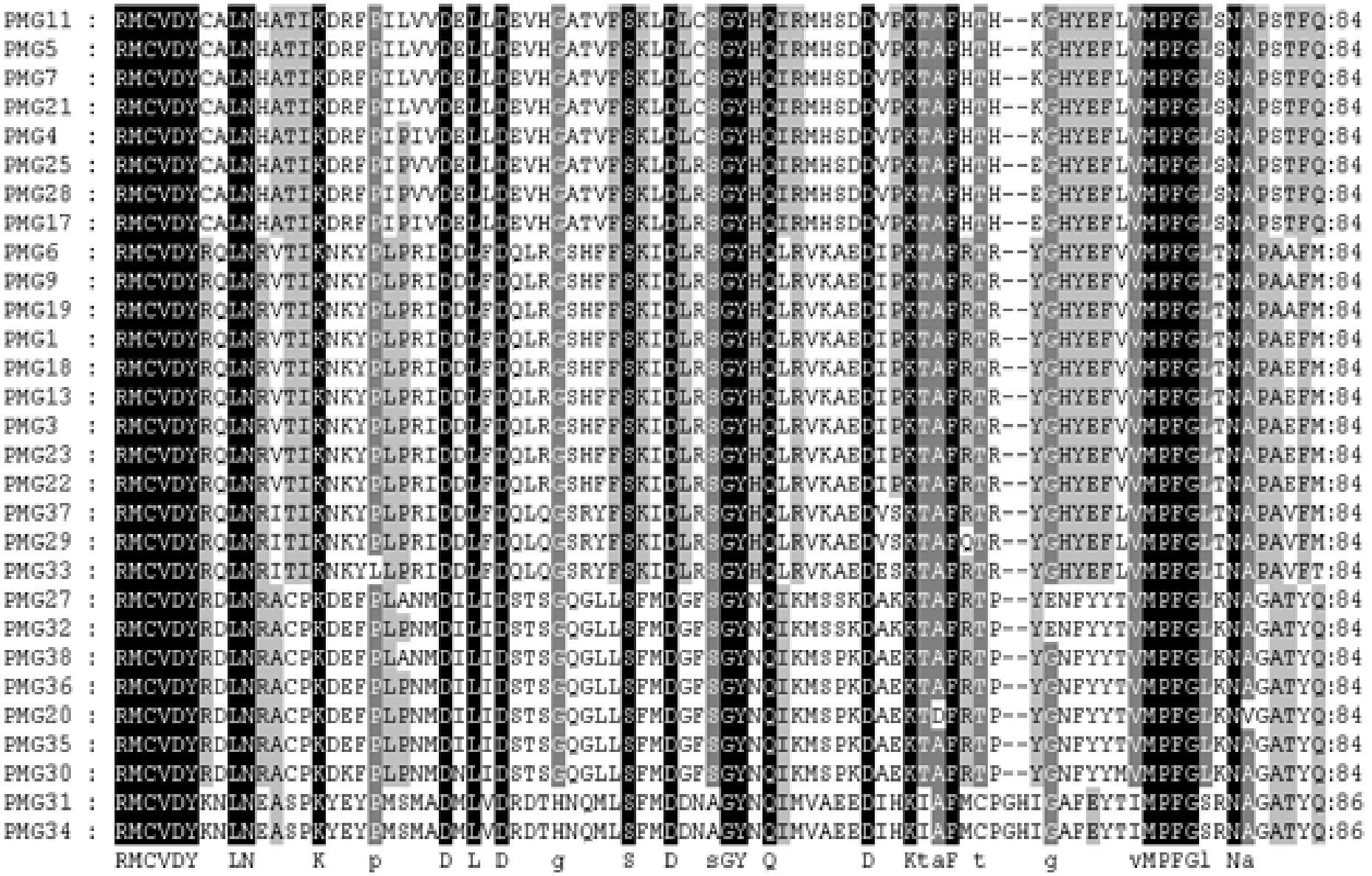

Figure 4. Sequence alignments of inferred amino acids corresponding to the RT domains from *Ty3-gypsy* group RT sequences of *P. mume*. Sequence positions for each sequence are indicated on the right of each sequence. Gaps are indicated as (-) and consensus sequence are shown below. The four shading levels indicate degree of residue conservation within a column: black (100% conserved), dark gray (80% conserved or greater), light gray (60% conserved or greater), and no shading (< 60% conserved).

Figure 4. Contd.

= 33.3%).

RT-PCR analysis

The transcriptional activity of LTR retrotransposons in Japanese apricot was investigated by RT-PCR technology using two pairs of degenerate primers Ty1-F, Ty1-R and Ty3-F, Ty3-R for *Ty1-copia* and *Ty3-gypsy* RT domains, respectively. Nevertheless, the amplification products corresponding to the expected size of the RT gene fragments of both types of LTR retrotransposons were not observed except the control gene of *Actin* (data not shown).

DISCUSSION

Japanese apricot, belonging to the *Rosaceae* family, has been a traditional medicine, a popular alcoholic beverage and a familiar and commonly consumed food. To the best of our knowledge, although retrotransposons have been widely found in the genomes of most eukaryotes, there has been no research regarding characterization of TEs obtained from the fruit tree of Japanese apricot. The present study is the first to survey the diversities of both types of *Ty1-copia* and *Ty3-gypsy* retrotransposons in *P. mume* genome.

RT domains of *Ty1-copia* as well as *Ty3-gypsy* group retrotransposons were successfully amplified by PCR using degenerate primers from *P. mume* genome. Our data showed that *Ty1-copia* group retrotransposons in Japanese apricot genome were highly heterogeneous, the divergences of predicted amino sequences of RT fragments were from 0 to 60.2%, like those observed in other plant species (Dixit et al., 2006; Nakatsuka et al., 2002; Price et al., 2002). The *Ty3-gypsy* RT clones of Japanese apricot also exhibited significant levels of sequence heterogeneity in their amino acids. However, they were less heterogeneous when compared to *Ty1-*

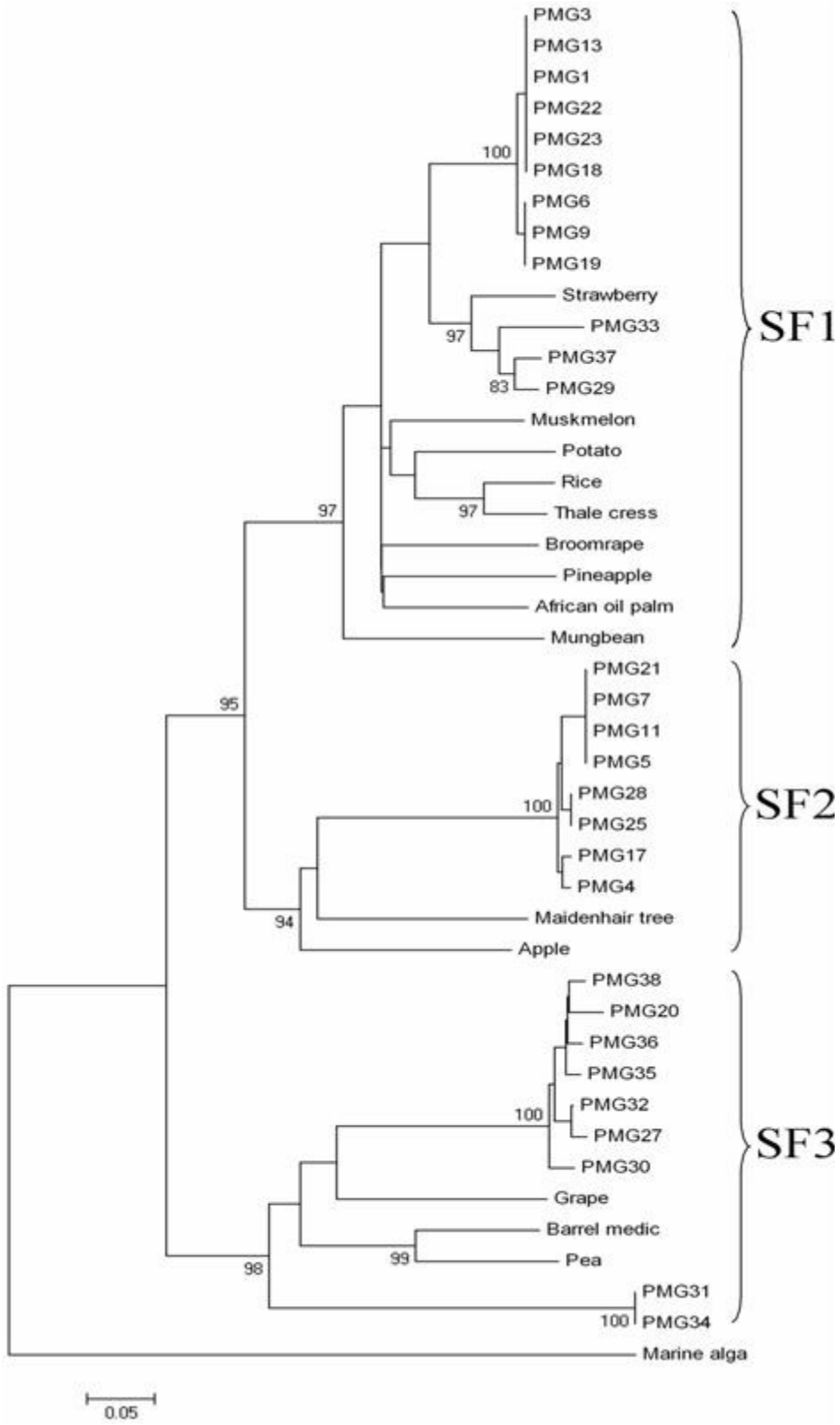

Figure 5. Phylogenetic relationships among the *Ty3-gyspy* RT sequences of *P. mume* and their homologues. The numbers indicate the bootstrap values from 1000 replicates, and only the values of 80% or more are given. The *Ty3-gypsy* RT sequences of Marine alga is used for outgroup rooting. Bar represents 5% sequence divergence. The details of RT sequences of other plants are given in Table 1.

Table 2. Average nonsynonymous (dN) and synonymous (dS) substitution rates for SG and SF subgroups belonging to *Ty1-copia* and *Ty3-gypsy*, respectively.

Subgroup	dN [a]	dS [a]	dN/dS
SG1	0.657 ± 0.183	2.556 ± 2.559	0.257
SG2	0.302 ± 0.045	2.400 ± 2.601	0.126
SG3	0.657 ± 0.183	2.556 ± 2.559	0.257
SG4	0.031 ± 0.013	0.164 ± 0.057	0.190
SG5	0.149 ± 0.027	1.028 ± 1.142	0.145
SG6	0.207 ± 0.033	1.577 ± 0.988	0.131
SF1	0.043 ± 0.010	0.656 ± 1.166	0.065
SF2	0.016 ± 0.007	0.023 ± 0.013	0.685
SF3	0.463 ± 0.060	0.694 ± 0.302	0.667

[a] The values were averages of dN's and dS's for comparisons of each RT coding sequence with other member sequences of the same subgroup. The sequences used were obtained from Japanese apricot only.

copia RT sequences as reflected by relatively low degree of sequence divergence between individual sequences with an overall average of 39.9% in contrast to 45.5% in the case of *Ty1-copia* RT sequences. Similar observations of lesser sequence divergence among *Ty3-gypsy* RT elements than that of *Ty1-copia* RT elements had been reported in strawberry (Ma et al., 2008). This might be because the *Ty3-gypsy* group retrotransposons are less conserved in the regions chosen for priming, especially in 3' -end. The high levels of sequence heterogeneities among the RT sequences of LTR retrotransposons of the same species might play a role in the evolution of control of retrotransposon activity (Grandbastien, 1998).

The total copy number of *Ty1-copia* group retrotransposons is nearly 7.9×10^3 molecules per genome, comprising 18.4% of the Japanese apricot genome, while the number of *Ty3-gypsy* is approximately 1.0×10^4, comprising nearly 33.3% of the *P. mume* genome. This suggests that *Ty3-gypsy* group retrotransposons might play more important roles in the diversity and evolution of the Japanese apricot genome. The same phenomenon was also observed by Sun et al. (2008), who reported that the *Ty3-gypsy* group sequences are abundant and constitute at least 33.5% in the genome of apple (*Malus* × *domestica*), while *Ty1-copia* content is only 4.6%. However, the contribution of *Ty1-copia* group retrotransposons (13%) is higher than that of the *Ty3-gypsy* group (10%) in the *Citrus sinensis* genome (Rico-Cabanas and Martinez-Izquierdo, 2007). These results suggested that different types of LTR retrotransposons had proliferated at different rates in diverse plant species and retrotransposons comprise a significant proportion of nuclear genomes, making them one of the most important components affecting the evolution of genome size in the plant kingdom.

Although plant genomes contain many copies of retrotransposons, most of them are transcriptionally silent in plants mainly because of the presence of stop codons and frameshifts in the coding regions. In the present study, our data showed that both group retrotransposon RT sequences possessed mutations accounted for 32.3 and 27.5% for *Ty1-copia* and *Ty3-gypsy*, respectively, which were obviously lower than those of persimmon (Nakatsuka et al., 2002), mungbean (Dixit et al., 2006) and apple (Sun et al., 2008). Important to note is that a striking similarity between putative amino acids of several Japanese apricot *Ty1-copia* RT sequences and two well-characterized active *Ty1-copia* retrotransposons *Tnt1* of tobacco (Grandbastien et al., 1989) and *Reme1* of melon (Ramallo et al., 2008) indicated the possibility of presence of related active retrotransposons in Japanese apricot genome. Absence of stop codons and/or frameshifts in the sequenced reading frames of most of *Ty1-copia* RT sequences further strengthened this opinion. Also, there might be potential active *Ty3-gypsy* retrotransposons in the *P. mume* genome, even though the clones of Japanese apricot were not grouped together with active *Ty3-gypsy* RT sequences reported previously (data not shown) (Fukai et al., 2008; Hansen and Sandmeyer, 1990; Lene Heegaard et al., 2005; Li et al., 2000; Marin and Llorens, 2000; Ohtsubo et al., 1999) due to the availability of very limited number of those sequences in other plant species. Moreover, the dN/dS analysis of each subgroup of the two types of retrotransposon RT sequences showed that purifying selection acted on these elements, and thus there is a good suggestion that most of the Japanese apricot retrotransposons analyzed in this study might be active (Hurst, 2002; Xu et al., 2006). However, our RT-PCR results with RNA isolated from young leaves of Japanese apricot plant, grown under normal growth conditions or treated with UV light and 2,4-D either individually or in both combinations, did not show any amplifications with *Ty1-copia* and *Ty3-gypsy* degenerate primers, indicating that both groups of sequences were transcriptionally

inactive under these conditions. It is likely that LTR retrotransposons in the Japanese apricot are difficult to be induced by treatments as described, which may be the evolution mechanism to minimize the possible deleterious effects of retrotransposons on the host. In addition, RT domains belong to only a small part of the POL genes of both groups of LTR retrotransposons and not completely represented in the characterization of full-length retrotransposons, therefore these results appears to conclude that there is no active retrotransposon in these sequences analyzed in the present study even though the ratios of dN/dS of the open reading frames among members of each subgroup of both group retrotransposons were less than 1.

Phylogenetic analysis revealed that each RT sequence subgroup of both types of LTR retrotransposons in Japanese apricot was more closely related to the representative sequences present in other plant species than to the other subgroups of Japanese apricot. In the case of *Ty1-copia*, the *P. mume* RT elements of subgroup SG1, supported by a high bootstrap probability value of 84%, were grouped with already deposited clones of apple and strawberry (*Fragaria* × *ananassa*), which belong to the *Rosaceae* family, and it indicated that these elements were present prior to the radiation of *Rosaceae*, which might be expected in line with vertical gene transmission. On the other hand, the sequences of the other five subgroups belonging to Japanese apricot grouped with unrelated plant species such as brown mustard (*Brassica nigra*), perennial kale (*Brassica oleracea*), grape (*Vitis vinifera*), pumello (*Citrus maxima*), chickpea (*Cicer arietinum*), trifoliate orange (*Poncirus trifoliate*), tomato and so on, suggested that different subgroups of Japanese apricot might have originated from different plant lineages of *Ty1-copia* retrotransposons and also that vertical and horizontal transmission as the source of *P. mume* RT sequences may not be mutually exclusive. Also, the *Ty3-gypsy* RT clones exhibited high levels of sequence homologies with those of other related or unrelated plant species, such as apple, strawberry, broomrape (*Orobanche crenata*), pineapple (*Ananas comosus*), maidenhair tree (*Ginkgo biloba*), grape, and so on. This spanning of species boundaries by *Ty3-gypsy* RT sequence subgroups also indicated that both vertical and horizontal transmission had played roles in *Ty3-gypsy* group retrotransposon evolution in the genome of *P. mume*. Alternatively, it is possible that these LTR retrotransposons might have originated from their common ancestor and diverged from that with gradual accumulation of sequence variations during the radiation of plant kingdom (Annu et al., 1998).

In conclusion, we firstly, revealed the existence of both *Ty1-copia* and *Ty3-gypsy* retrotransposons in Japanese apricot genome, and examined their sequence heterogeneity, phylogenetic relationships, copy numbers and transcriptional activity. In future, LTR sequences of retrotransposons in the Japanese apricot genome will be isolated by the chromosome walking technique based on RT sequences, and LTR retrotransposon-based molecular markers will be used to respond to the need for polymorphic markers to study the genetic relationships of *P. mume* populations.

REFERENCES

Annu S, Jaakko T, Alan HS (1998). Gypsy-like retrotransposons are widespread in the plant kingdom. Plant J. 13: 699-705.

Beguiristain T, Grandbastien MA, Puigdomenech P, Casacuberta JM (2001). Three *Tnt1* subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiol. 127: p. 212.

Bennetzen JL (2000). Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. 42: 251-269.

Bowen NJ, Jordan IK (2002). Transposable elements and the evolution of eukaryotic complexity. Curr. Issues Mol. Biol. 4: 65-76.

Chu MY (1999) China fruit flora. Japanese apricot. China Forest Press, Beijing

Chuda Y, Ono H, Ohnishi-Kameyama M, Matsumoto K, Nagata T, Kikuchi Y (1999). Mumefural, Citric Acid Derivative Improving Blood Fluidity from Fruit-Juice Concentrate of Japanese Apricot (*Prunus mume* Sieb. et Zucc). J. Agric. Food Chem. 47: 828-831.

Dixit A, Ma KH, Yu JW, Cho EG, Park YJ (2006). Reverse transcriptase domain sequences from Mungbean (*Vigna radiata*) LTR retrotransposons: sequence characterization and phylogenetic analysis. Plant Cell Rep. 25: 100-111.

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

Fukai E, Dobrowolska A, Madsen L, Madsen E, Umehara Y, Kouchi H, Hirochika H, Stougaard J (2008). Transposition of a 600 thousand-year-old LTR retrotransposon in the model legume *Lotus japonicus*. Plant Mol. Biol. 68: 653-663.

Gambino G, Perrone I, Gribaudo I (2008). A Rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. Phytochem. Anal. 19: 520-525.

Grandbastien M-A (1998). Activation of plant retrotransposons under stress conditions. Trends Plant Sci. 3: 181-187.

Grandbastien MA (1992). Retroelements in higher plants. Trends Genet. 8: 103-108.

Grandbastien MA, Lucas H, Morel JB, Mhiri C, Vernhettes S, Casacuberta JM (1997). The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. Genetica, 100:241-252.

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

Hansen LJ, Sandmeyer SB (1990). Characterization of a transpositionally active Ty3 element and identification of the Ty3 integrase protein. J. Virol. 64: 2599-2607.

Havecker ER, Gao X, Voytas DF (2004). The diversity of LTR retrotransposons. Genome Biol. 5: p. 225.

Hill P, Burford D, Martin DM, Flavell AJ (2005). Retrotransposon populations of *Vicia* species with varying genome size. Mol. Genet. Genomics, 273: 371-381.

Hurst LD (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 18: 486-487.

Jenog JT, Moon J-H, Park K-H, Shin CS (2006). Isolation and characterization of a new compound from *Prunus mume* fruit that inhibit cancer cells. J. Agric. Food Chem. 54: 2123-2128.

Kapitonov VV, Jurka J (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. Nat. Rev. Genet. 9: 411-412.

Kidwell MG, Lisch DR (2000). Transposable elements and host genome evolution. Trends Ecol. Evol. 15: 95-99.

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004). Retrotransposon-Induced Mutations in Grape Skin Color. Science, 304: p. 982.

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

Kumekawa N, Ohtsubo H, Horiuchi T, Ohtsubo E (1999). Identification and characterization of novel retrotransposons of the *gypsy* type in rice. Mol. General Genet. 260: 593-602.

Lene Heegaard M, Eigo F, Simona R, Christopher Karl Y, Niels S, Leif S, Jens S (2005). LORE1, an active low-copy-number TY3-*gypsy* retrotransposon family in the model legume *Lotus japonicus*. Plant J. 44: 372-381.

Li ZY, Chen SY, Zheng XW, Zhu LH (2000). Identification and chromosomal localization of a transcriptionally active retrotransposon of Ty3-*gypsy* type in rice. Genome, 43: 404-408.

Lodhi M, Ye G-N, Weeden N, Reisch B (1994). A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. Plant Mol. Biol. Rep. 12: 6-13.

Ma Y, Sun H, Zhao G, Dai H, Gao X, Li H, Zhang Z (2008). Isolation and characterization of genomic retrotransposon sequences from octoploid strawberry (*Fragaria × ananassa* Duch.). Plant Cell Rep. 27: 499-507.

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

Marin I, Llorens C (2000). *Ty3/Gypsy* retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. Mol. Biol. Evol. 17: 1040-1049.

McCarthy EM, Liu J, Lizhi G, McDonald JF (2002). Long terminal repeat retrotransposons of *Oryza sativa*. Genome Biol. 3: RESEARCH0053.

Nakatsuka A, Iwami N, Matsumoto S, Itamura H, Yamagishi M (2002). *Ty1-copia* group retrotransposons in persimmon (*Diospyros kaki* Thunb.). Genes Genet. Syst. 77: 131-136.

Ohtsubo H, Kumekawa N, Ohtsubo E (1999). *RIRE2*, a novel *gypsy*-type retrotransposon from rice. Genes Genet. Syst. 74: 83-91.

Park YJ, Dixit A, Yoo JW, Bennetzen J (2004). Further evidence of microcolinearity between barley and rice genomes at two orthologous regions. Mol. Cells 17: 492-502.

Poirot O, O'Toole E, Notredame C (2003). Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res. 31: 3503-3506.

Price Z, Dumortier F, MacDonald W, Mayes S (2002). Characterisation of copia-like retrotransposons in oil palm (*Elaeis guineensis* Jacq.). Theor. Appl. Genet. 104: 860-867.

Ramallo E, Kalendar R, Schulman A, Martínez-Izquierdo J (2008). *Reme1*, a Copia retrotransposon in melon, is transcriptionally induced by UV light. Plant Mol. Biol. 66: 137-150.

Rico-Cabanas L, Martinez-Izquierdo JA (2007). *CIRE1*, a novel transcriptionally active *Ty1-copia* retrotransposon from *Citrus sinensis*. Mol. Genet. Genomics, 277: 365-377.

Shi J, Gong J, Liu Je, Wu X, Zhang Y (2009). Antioxidant capacity of extract from edible flowers of *Prunus mume* in China and its active components. LWT-Food Sci. Technol. 42: 477-482.

Sun HY, Dai HY, Zhao GL, Ma Y, Ou CQ, Li H, Li LG, Zhang ZH (2008). Genome-wide characterization of long terminal repeat - retrotransposons in apple reveals the differences in heterogeneity and copy number between Ty1-*copia* and Ty3-*gypsy* retrotransposons. J. Integr. Plant Biol. 50: 1130-1139.

Suyama M, Torrents D, Bork P (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34: W609-612.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24: 1596-1599.

Vicient CM, Jaaskelainen MJ, Kalendar R, Schulman AH (2001). Active retrotransposons are a common feature of grass genomes. Plant Physiol. 125: 1283-1292.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8: 973-982.

Wilhelm M, Wilhelm FX (2001). Reverse transcription of retroviruses and LTR retrotransposons. Cell Mol. Life Sci. 58: 1246-1262.

Xu J, Pan G, Fang L, Li J, Tian X, Li T, Zhou Z, Xiang Z (2006). The varying microsporidian genome: Existence of long-terminal repeat retrotransposon in domesticated silkworm parasite *Nosema bombycis*. Int. J. Parasitol. 36: 1049-1056.

Yang Z, Nielsen R (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17: 32-43.

Zhang SL, Huang SX, Kitashiba H, Nishio T (2007). Identification of S-haplotype-specific F-box gene in Japanese plum (*Prunus salicina* Lindl.). Sexual Plant Reprod. 20: 1-8.