Characterization and chromosomal distribution of Ty3-*gypsy*-like retrotransposons in wax gourd (*Benincasa hispida*)

Biao Jiang[a,b]**, Wenrui Liu**[a,b]**, Xiaoming He**[a,b]**, Qingwu Peng**[a]**, Dasen Xie**[a,b,*]

[a] Vegetable Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China
[b] Guangdong Provincial Key Lab for New Technology Research on Vegetables, Guangzhou 510640, China

*Corresponding author, e-mail: xiedasen@126.com

ABSTRACT: Wax gourd ($2n = 2x = 24$) is an important vegetable species in the Cucurbitaceae family, yet little is known about its genome structure. As the largest component of most plant genomes, retrotransposons might play an important role in the genetic variation and genome evolution of wax gourd. In this study, the presence of Ty3-*gypsy*-like retrotransposons, their heterogeneity and chromosomal distribution in wax gourd have been investigated. The conserved domains of reverse transcriptase of Ty3-*gypsy*-like retrotransposons have been amplified, and this resulted in thirty-five unique sequences. Seventeen (49%) of the sequences were defective as they were disrupted by a premature stop codon and/or contained frameshift mutations. These 35 sequences were divided into four groups and showed low or even no homology with Ty3-*gypsy*-like sequences from other species, indicating lack of horizontal transfer. The Ty3-*gypsy*-like retrotransposons have a dispersed genomic organization, physically distributed on all chromosomes with clusters in the heterochromatin regions. These results suggest that retrotransposons of this kind are widely distributed in wax gourd with high copy numbers, and contributed to the formation of chromosomes with non-uniform insertion and amplification. This is the first report on the presence of reverse transcriptase sequences of Ty3-*gypsy*-like retrotransposons in wax gourd genome.

KEYWORDS: reverse transcriptase, heterogeneity, phylogenetic analysis, FISH

INTRODUCTION

Retrotransposons are ubiquitous in the plant kingdom and present in high copy numbers in most plants, which make them major components of plant genomes[1]. Long terminal repeat (LTR) retrotransposons, being one of the main groups of retrotransposons, are among the most abundant constituents of plant genomes[2]. They are further classified into the Ty1-*copia* and the Ty3-*gypsy* groups which differ from each other by sequence similarity and the order of their encoded gene products[1,2]. It is widely accepted that retrotransposons play an essential role in plant evolution[3]. They cannot only generate mutations by inserting themselves within or near genes, but also cause chromosomal rearrangements by serving as sites for ectopic recombination[4–6].

One particular effective strategy to detect LTR retrotransposons is to amplify the relevant sequences using degenerate primers for the reverse transcriptase domain[7–9]. This strategy has been employed to study the distribution of retrotransposons in many plant species, such as cereals[10], *Olea*[11], mung bean[12], strawberry[13], and *Cucumis hystrix*[14]. This approach is especially useful when genomic sequences for the species in question are not available. Wax gourd (*Benincasa hispida*, $2n = 2x = 24$) is an important vegetable crop in Cucurbitaceae, and is widely distributed in East and South Asia. The storage period of wax gourd is very long, making it a very important off-season vegetable. Its fruits are often consumed as fresh vegetables as well as candied, dried, medicinal, or processed products[15–17].

Despite being an important vegetable crop, little is known about wax gourd retrotransposons and their genetic distribution. The genome characterization of retrotransposons with reference to the content, heterogeneity, and chromosomal distribution may contribute to our understanding of the genome organization and evolution of wax gourd. Reverse transcriptase (RT) is the most conserved domain in retrotransposon sequences and commonly represents retrotransposon for phylogenetic analysis. Thus in the present study, we isolated the reverse transcriptase

sequences of Ty3-*gypsy*-like retrotransposons in wax gourd genome, and then investigated their sequence heterogeneity, phylogenetic relationship, and chromosomal distribution.

MATERIALS AND METHODS

Plant material and isolation of nucleic acids

One wax gourd cultivar 'B227' was used in this study. Seeds were dipped in water for 10 h, and then kept in a Petri dish with two layers of wet filter paper at 30 °C. The germinated seeds were grown in light incubator. Total DNA was extracted from fresh leaves of two-week-old seedlings using the CTAB method described by Murray and Thompson [18].

Polymerase chain reaction and cloning of PCR products

The internal region of RT gene of Ty3-*gypsy* retrotransposons was amplified by PCR using total genomic DNA of wax gourd as templates and degenerate primers Gy1 5′-AGM GRA TGT GYG TSG AYT AT -3′ and Gy2 5′-CAM CCM RAA MWC ACA MTT-3′ [9]. Reaction mixtures were 20 µl, containing 50 ng of DNA, 50 pmol of Gy1 and Gy2, 0.2 mM of dNTP, 2.5 mM of $MgCl_2$, and 1 U of *Taq* polymerase (TaKaRa, Japan). The PCR cycles included an initial denaturation step at 96 °C for 2 min, followed by 30 cycles at 96 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min, and then a final elongation step at 72 °C for 7 min.

The PCR products were separated by electrophoresis in 1% agarose gels in 1 × TAE buffer and visualized under UV light after staining with ethidium bromide. The amplified products with the size of about 430 bp were recovered and purified by DNA Gel Extraction Kit (Dingguo, China). The purified fragments were cloned into the vector pMD19-T (TaKaRa, Japan), and then transformed into *E. coli* strain DH5a. The positive colonies were selected to extract plasmid DNA. The insertions were confirmed by PCR amplification using recombinant clones as templates. The clones with expected size were then sequenced by Invitrogen Bio-Technology Co., Ltd., Shanghai, China.

Sequence analysis

Search for homologous sequences of our clones was performed with the BLASTN, BLASTX, and TBLASTX searching tools as implemented on the National Centre for Biotechnology Information (www. ncbi.nlm.nih.gov) non-redundant databases. Multiple sequence alignments were performed using

CLUSTALW [19]. Frameshifts were detected using ERRWISE (http://coot.embl.de/ERR_WISE/) and corrected to conserve the consensus peptide sequences of the sample data. Other RT sequences were used to create a comparative phylogenetic dendrogram (GenBank Accession No. is given in parentheses), including *Prunus mume* (ACU42183), *Cucumis melo* (ADN 34002), *Cucumis hystrix* (ADD 83130), *Orobanche ramose* (ABD 43151), *Orobanche owerinii* (ABD 43086), *Morus bombycis* (BAB 40830), *Chrysanthemum morifolium* (BAB 40831), *Solanum demissum* (AAT 38744), *Beta vulgaris* (AEV 42258), *Oryza sativa* (AAL 79340), *Eleocharis palustris* (ADF 46122), *Vitis vinifera* (CAN 66348), *Elaeis guineensis* (CAD 45567), and *Solanum lycopersicum* (ADU 56212). The construction of phylogenetic trees and bootstrap tests with 1000 replications were carried out using MEGA 5.0 using neighbour-joining method [20].

The mode of selection (positive selection, purifying selection, or neutral selection) among the RT sequences of Ty3-*gypsy* retrotransposons was detected. The Ka/Ks ratios were calculated according to Nei and Gojobori using K-ESTIMATOR 6.0 [21, 22].

Chromosome preparation and FISH

Roots of wax gourd were harvested and fixed in freshly prepared 3/1 mixture of 100% methanol/glacial acetic acid for at least 1 day. The roots were then rinsed for 5 min in distilled water for three times. Subsequently, the root-tips were cut off and transferred into enzyme mixtures containing 4% cellulose and 2% pectinase at 28 °C for 2.5 h. The enzyme solution were replaced using distilled water and left on ice for 30 min before transferring the root-tips into the fixative solution (100% methanol/glacial acetic acid, 3/1). The root tip was completely macerated on a slide and flame dried the slide. The quality of chromosomes was checked under a phase contrast microscope and stored in −70 °C freezer until use.

FISH assay was conducted essentially according to published procedures [23]. The total RT genes of Ty3-*gypsy* retrotransposons were used as probes for fluorescence in situ hybridization. The samples were labelled with digoxigenin-11-dUTP and detected with rhodamine-conjugated anti-digoxigenin antibody (Roche). Chromosomes were counterstained with DAPI in VectorShield anti-fade solution (Vector Laboratories, Burlingame, CA), and images were captured using a SenSys CCD camera attached to an Olympus BX60 microscope. The CCD camera was controlled using FISHVIEW 5.5 (Applied Spectral Imaging, Inc, USA). PHOTOSHOP was used to obtain optimal images.

RESULTS

Isolation and characterization of Ty3-*gypsy* RT sequences

Polymerase chain reaction (PCR) amplification with degenerate primers flanking the reverse transcriptase domains of Ty3-*gypsy*-like retrotransposons yielded the expected fragments from wax gourd. The targeted PCR products were recovered, purified and cloned. Forty clones were randomly selected for further sequencing. The sequencing data revealed that thirty-five clones were unique. The 5′ and 3′ terminal of all the 35 sequences matched the degenerate primers of *gypsy*-like retrotransposons. The conceptual translation products corresponded to the *gypsy* motifs. These sequences were designated as GYRT1 to GYRT3(5).

Among the 35 sequences, the pairwise similarity values ranged from 65% to 99.8% at the nucleotide sequence level. The putative *gypsy*-like RT sequences were translated into amino acids and analysed for the presence of stop codons and frameshifts in their coding regions. Among the 35 sequences, 17 sequences (49%) were 'defective', since each of them contained a stop codon, a frameshift mutation, or both. The remaining 18 sequences (51%) were 'intact,' which meant there was neither a stop codon nor a frameshift within their coding regions. Alignment of putative amino acid sequences showed that these 35 sequences contained the RMCVDYR motif at their N-termius (Fig. 1) and the similarity at the amino acid level ranged from 65% to 100%. The low sequence similarities observed among the sequences suggest that there are multiple Ty3-*gypsy* element families in the wax gourd genome.

Phylogenetic analysis of the Ty3-*gypsy* RT sequences

To study the relationships among the obtained RT sequences, a neighbour-joining tree was constructed by MEGA 5.0. These sequences were divided into four groups (Fig. 2). Group I is the largest group, which contained 17 sequences. Group II is the second largest group, containing 13 sequences.

For comparative purposes, sequences of *gypsy* group retrotransposons from other species were retrieved from the GenBank database, which were used in the sequence alignment (Fig. 3). The comparative tree sustained the four groups identified in Fig. 2. All the RT sequences amplified from wax gourd showed low or even no homology with those from other species.

The numbers of synonymous and non-synonymous substitutions per site were estimated for group
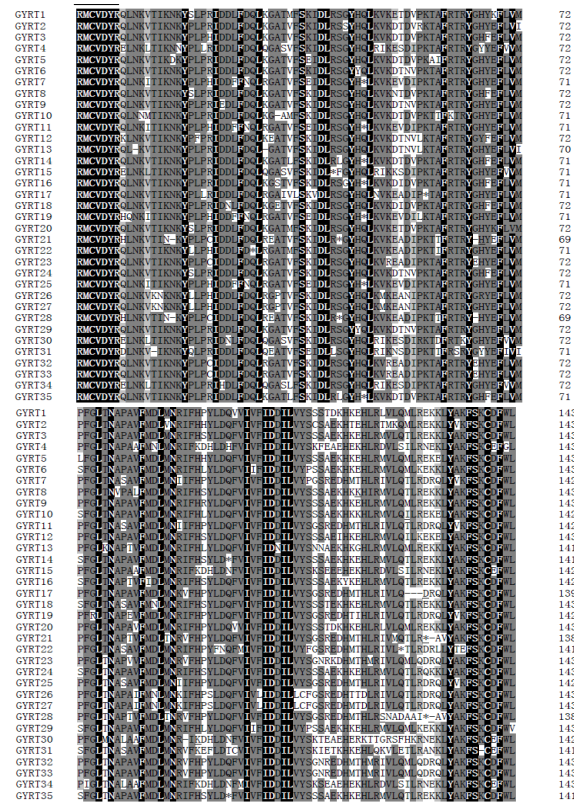

Fig. 1 Alignment of the conceptual translations of the nucleotide sequences representing the reverse transcriptase sequences of Ty3-gypsy retrotransposons isolated from wax gourd. The four shading levels indicate degree of residue conservation: black (100% conserved), dark grey (75% or greater conserved), light grey (50% or greater conserved), and no shading (< 50% conserved). Gaps are indicated as (–) and stop codons are presented as (*). The amino acid of frameshift is underlined. The numbers of amino acid residues are displayed on the right hand of each sequence. The conserved domains were indicated by black line upon the alignment.

Table 1 Number of synonymous and non-synonymous substitutions per site within RT gene domain of Ty3-*gypsy*-like retrotransposons from wax gourd.

Group	Ka	Ks	Ka/Ks
I	0.010 ± 0.008	0.049 ± 0.050	0.279
II	0.008 ± 0.001	0.014 ± 0.003	0.578
IV	0.018	0.034	0.529

I, group II, and group IV, since they contained 'intact' RT sequences (Table 1). The ratios of non-synonymous to synonymous substitutions (dN/dS) were 0.279, 0.578, and 0.529, respectively, which suggests
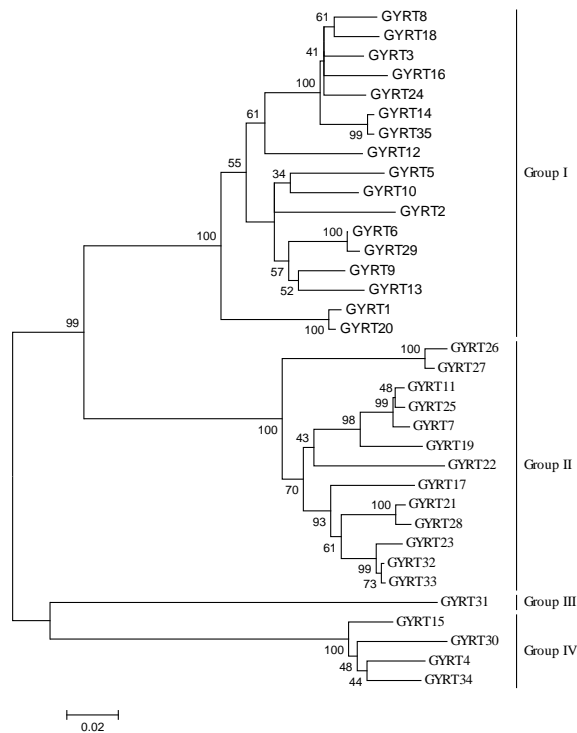
Fig. 2 Phylogenetic analysis of nucleotide sequences representing the Ty3-*gypsy* RT sequences from wax gourd. The tree was generated using MEGA 5.0. Numbers at the branch nodes indicate the bootstrap support out of 1000 replications. The branch lengths are proportional to the genetic distances as estimated by the neighbour-joining method[20].

that retrotransposons of these groups might have been under constraint- or purifying-selection.

Chromosomal distribution of Ty3-*gypsy* retrotransposons in wax gourd

The chromosomal distribution of Ty3-*gypsy*-like retrotransposons was investigated using fluorescence in situ hybridization (FISH). The total RT genes of Ty3-*gypsy* retrotransposons were used as probes to hybridize to chromosomes of wax gourd (Fig. 4). From the interphase nuclei, it was evident that the Ty3-*gypsy* retrotransposons are widely distributed on the chromatin with clusters in heterochromatic (DAPI-blight) regions (Fig. 4a–c). The distribution of Ty3-*gyspy* retrotransposons was more obvious on metaphase chromosomes, which showed that they were clustered in heterochromatic regions of all the chromosomes (Fig. 4d–f).
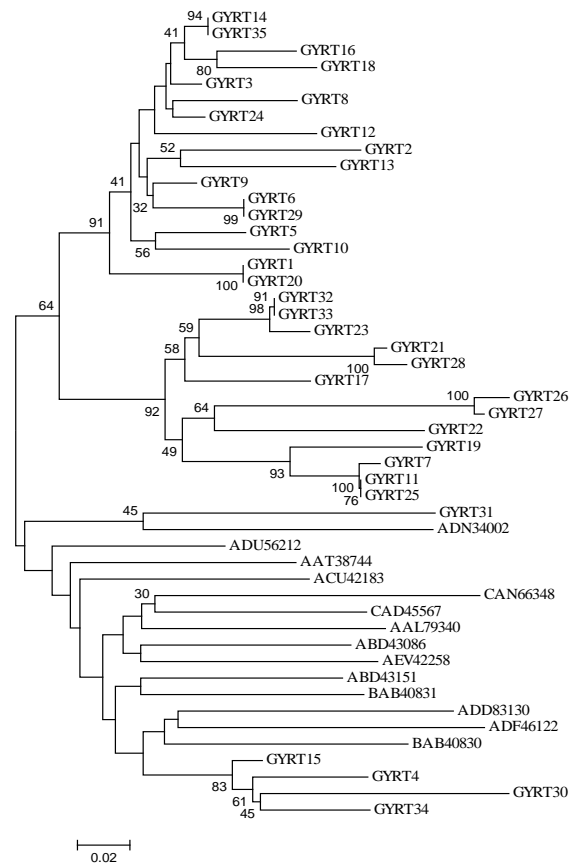

Fig. 3 Phylogenetic relationships among the deduced amino acid sequences for the reverse transcriptase genes of Ty3-*gypsy*-like retrotransposons from wax gourd and other related sequences using the neighbour-joining method[20]. The related sequences included: *Prunus mume* (ACU42183), *Cucumis melo* (ADN34002), *Cucumis hystrix* (ADD83130), *Orobanche ramose* (ABD43151), *Orobanche owerinii* (ABD43086), *Morus bombycis* (BAB40830), *Chrysanthemun morifolium* (BAB40831), *Solanum demissum* (AAT38744), *Beta vulgaris* (AEV42258), *Oryza sativa* (AAL79340), *Eleocharis palustris* (ADF46122), *Vitis vinifera* (CAN66348), *Elaeis guineensis* (CAD45567), *Solanum lycopersicum* (ADU56212).

DISCUSSION

To the best of our knowledge, this study is the first to investigate the presence of Ty3-*gypsy*-like retrotransposons in wax gourd genome. Reverse transcriptase domains of Ty3-*gypsy*-like retrotransposons in wax gourd genome were successfully amplified using degenerate primers, and eventually non-redundant 35 sequences were obtained. The Ty3-*gypsy*-like RT sequences exhibited significant levels of heterogeneity. The similarity at the amino acid level ranged from
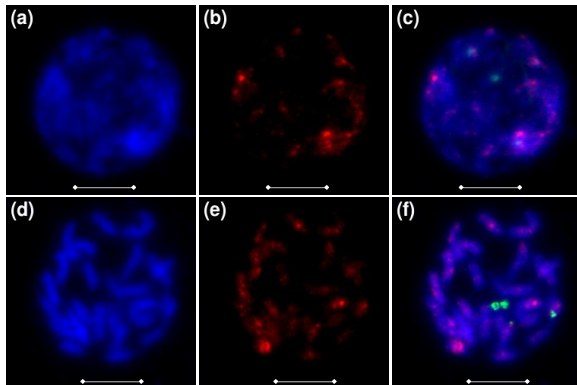
Fig. 4 Chromosomal distribution of Ty3-*gypsy*-like retrotransposons on interphase nuclei and metaphase chromosomes of wax gourd revealed by fluorescence in situ hybridization; (a), (d) interphase and metaphase chromosomes were counterstained with DAPI; (b), (e) total reverse transcriptase domains of Ty3-*gypsy* retrotransposons from wax gourd probed on interphase nucleus and metaphase chromosomes; (c), (f) chromosomes counterstained with DAPI and reverse transcriptase domains of Ty3-*gypsy* retrotransposons. Bars=5µm.

65% to 100%, which is in agreement with previous observations reported in other plant species, such as strawberry[13] and mungbean[12].

The classification of these sequences into four groups further illustrated the divergence of Ty3-*gypsy* retrotransposons in wax gourd (Fig. 2). Similar result was also reported in *Vicia* species[24]. Comparative analysis showed that all the sequences obtained from wax gourd had low or even no homology with other sequences from other species (Fig. 3). These results suggested that vertical transmission was the main transfer mode of Ty3-*gypsy*-like retrotransposons in wax gourd, and horizontal transmission might have little effect in the evolution of retrotransposons of this kind.

Although high copies of *gypsy*-like retrotransposons exist in plant genome, most of them are unable to transpose autonomously because of mutations such as stop codons and frameshifts in the coding regions. This might be an important mechanism to reduce potential deleterious effects of retrotransposons on plant genomes. Defective retrotransposons could be the result of mutations and/or errors made by the reverse transcriptase during reverse transcription, because the reverse transcriptase enzymes had no proof-reading function[25]. For example, about 65% of Ty3-*gypsy* reverse transcriptase domains in mung bean were defective because of a stop codon and/or a frameshift

mutation in their coding regions[12]. In the present study, about 49% of Ty3-*gypsy* reverse transcriptase coding sequences were found to be 'defective'. This is an underestimation because only a portion of the coding region was examined. The synonymous (dS) and non-synonymous substitution (dN) analysis of the 'intact' reverse transcriptase sequences implied that retrotransposons of this kind had been under constraint-or purifying-selection (Table 1).

Chromosomal location of Ty3-*gypsy* retrotransposons has been studied in several plant species[26–29]. In *Triticeae*, Ty3-*gypsy*-like retrotransposons were clustered in the subterminal as well as pericentromeric heterochromatic regions on chromosomes[28]. Ty3-*gypsy*-like retrotransposons in *Helianthus* were primarily localized to the pericentromeric regions of metaphase chromosomes[29]. In *C. hystrix*, retrotransposons of this type concentrated on the terminal heterochromatic regions of chromosomes (unpublished data). In our present study, Ty3-*gypsy* retrotransposons were widely dispersed over all the chromosomes in wax gourd, with clustering in some heterochromatin regions (DAPI-blight) (Fig. 4). This result suggested that Ty3-*gyspy* retrotransposons contribute to the formation of chromosomes with non-uniform insertion and amplification in wax gourd. In the meantime, these results also provide us important information for dissecting the genomic structure of wax gourd.

In conclusion, we firstly revealed the presence of Ty3-*gypsy*-like retrotransposons in wax gourd, and examined their sequence heterogeneity, phylogenetic relationship, and chromosomal distribution. These results provide basic genetic and evolutionary information for these elements in the wax gourd genome, and will also be of importance for the utilization of retrotransposons in wax gourd.

REFERENCES

1. Kumar A, Bennetzen JL (1999) Plant retrotransposons. *Annu Rev Genet* **33**, 479–532.
2. Havecker ER, Gao X, Voytas DF (2004) The diversity of LTR retrotransposons. *Genome Biol* **5**, 225.
3. Wessler SR, Bureau TE, White SE (1995) LRT-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr Opin Genet Dev* **5**, 814–21.

4. Yao JL, Dong YH, Bret AM (2001) Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. *Plant Biol* **98**, 1306–11.

5. Bennetzen JL (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**, 29–36.

6. Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. *Science* **204**, 982.

7. Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) *Ty1-copia* group retrotransposons are ubiquitous and heterogenous in higher plants. *Nucleic Acids Res* **20**, 3639–44.

8. 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 organization, evolution and use as molecular markers. *Genetica* **100**, 205–17.

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

10. Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P, Kumar A, Flavell AJ (1999) Phylogeny and transpositional activity of Ty1-*copia* group retrotransposons in cereal genomes. *Mol Gen Genet* **261**, 883–91.

11. Stergiou G, Katsiotis A, Hagidimitriou M (2002) Genomic and chromosomal organization of Ty1-*copia*-like sequences in *Olea europaea* and evolutionary relationships of *Olea* retroelements. *Theor Appl Genet* **104**, 926–33.

12. Dixit A, Ma KH, Yu JW, Cho EG, Park YJ (2006) Reverse transcriptase domain sequences from Mungbean (*Vigna radiate*) LTR retrotransposons: sequence characterization and phylogenetic analysis. *Plant Cell Rep* **25**, 100–11.

13. Ma Y, Sun HY, Zhao GL, Dai HY, Gao XY, Li H, Zhang ZH (2008) Isolation and characterization of genomic retrotransposon sequences from octoploid strawberry (*Fragaria × ananassa* Duch.). *Plant Cell Rep* **27**, 499–507.

14. Jiang B, Wu ZM, Lou QF, Wang D, Zhang WP, Chen JF (2010) Genetic diversity of Ty1-*copia* retrotransposons in a wild species of *Cucumis* (*C. hystrix*). *Sci Hort* **127**, 46–53.

15. Ramesh M, Gayathri V, Rao AV, Prabhakar M, Rao CS (1989) Pharmacological action of fruit juice of *Benincasa hispida*. *Fitoterapia* **60**, 241–7.

16. Grover JK, Adiga G, Vats V, Rathi SS (2001) Extracts of *Benincasa hispida* prevent development of experimental of experimental ulcer. *J Ethnopharmacol* **78**, 159–64.

17. Verma VK, Behera TK, Munshi AD, Parida SK, Mohapatra T (2007) Genetic diversity of ash gourd [*Benincasa hispida* (Thunb.) Cogn.] inbred lines based on RAPD and ISSR markers and their hybrid performance. *Sci Hort* **113**, 231–7.

18. Murray HG, Thompson WF (1980) Rapid isolation of higher weight DNA. *Nucleic Acids Res* **8**, 4321–6.

19. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–80.

20. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–25.

21. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418–26.

22. Comeron JM (1995) A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. *J Mol Evol* **41**, 1152–9.

23. Jiang JM, Gill BS, Wang GL, Ronald PC, Ward DC (1995) Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proc Natl Acad Sci USA* **92**, 4487–91.

24. Hill P, Burford D, Martin DMA, Flavell AJ (2005) Retrotransposon populations of *Vicia* species with varying genome size. *Mol Genet Genom* **273**, 209–16.

25. Steinhauer DA, Holland JJ (1986) Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *J Virol* **57**, 219–28.

26. Miller JT, Dong F, Jackson SA, Song J, Jiang J (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* **150**, 1615–23.

27. Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A *TY3-GYPSY* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *Plant J* **16**, 721–8.

28. Belyayev A, Raskina O, Nevo E (2000) Chromosomal distribution of reverse transcriptase-containing retroelements in two *Triticeae* species. *Chromosome Res* **9**, 129–36.

29. Staton SE, Ungerer MC, Moore RC (2009) The genomic organization of *Ty3/gypsy*-like retrotransposons in *Helianthus* (Asteraceae) homoploid hybrid species. *Am J Bot* **96**, 1646–55.