Tree Genetics & Genomes (2018) 14: 43
https://doi.org/10.1007/s11295-018-1257-x

ORIGINAL ARTICLE

# LTR retrotransposons from the *Citrus x clementina* genome: characterization and application

Dongliang Du[1] · Xiaoyun Du[2] · Matthew R. Mattia[1] · Yanbo Wang[2] · Qibin Yu[1] · Ming Huang[1] · Yuan Yu[1] · Jude W. Grosser[1] · Fred G. Gmitter Jr[1]

Received: 5 October 2017 / Revised: 15 May 2018 / Accepted: 17 May 2018 / Published online: 1 June 2018


## Abstract

Long terminal repeat retrotransposons (LTR-RTs) are a large portion of most plant genomes, and can be used as a powerful molecular marker system. The first citrus reference genome (*Citrus x clementina*) has been publicly available since 2011; however, previous studies in citrus have not utilized the whole genome for LTR-RT marker development. In this study, 3959 full-length LTR-RTs were identified in the *C. x clementina* genome using structure-based (LTR_FINDER) and homology-based (RepeatMasker) methods. LTR-RTs were first classified by protein domain into *Gypsy* and *Copia* superfamilies, and then clustered into 1074 families based on LTR sequence similarity. Three hundred fifty *Copia* families were grouped into four lineages: *Retrofit*, *Tork*, *Sire*, and *Oryco*. One hundred seventy-eight *Gypsy* families were sorted into six lineages: *Athila*, *Tat*, *Renia*, *CRM*, *Galadriel*, and *Del*. Most LTR-RTs (3218 or 81.3%) were anchored to the nine Clementine mandarin linkage groups, accounting for 9.74% of chromosomes currently assembled. Accessions of 25 Rutaceae species were genotyped using 17 inter-retrotransposon amplified polymorphism (IRAP) markers developed from conserved LTR regions. Sequence-specific amplified polymorphism (SSAP) makers were used to distinguish 'Valencia' and 'Pineapple' sweet oranges (*C. x sinensis*), and 24 sweet orange clones. LTR-RT markers developed from the Clementine genome can be transferred within the Rutaceae family demonstrating that they are an excellent tool for citrus and Rutaceae genetic analysis.

**Keywords** *Citrus* · LTR retrotransposons · *Gypsy* · *Copia* · IRAP · SSAP

## Introduction

Retrotransposons (RT) are a type of transposable element (TE) that moves through the genome via an RNA intermediate in a process that resembles "copy and paste" (Wicker et al. 2007). Retrotransposons can be separated into two major subclasses, long terminal repeat retrotransposons (LTR-RTs) and non-LTR retrotransposons, based on their structure and transposition mechanism (Todorovska 2007). A typical LTR-RT contains two highly similar long terminal repeats (LTRs), a primer-binding site (PBS), a polypurine tract (PPT), and two genes necessary for their retrotransposition, *gag* and *pol* (Du et al. 2010). The majority of LTR-RTs can be divided further into *Copia* and *Gypsy* superfamilies according to the order of proteinase (PR), integrase (IN), reverse transcriptase (RT), and RNase H (RH) domains in Pol (Domingues et al. 2012). The domains of *Gypsy* elements are arranged as LTR-GAG-PR-RT-RH-IN-LTR, whereas the *Copia* elements are organized as LTR-GAG-PR-IN-RT-RH-LTR (Wicker et al. 2007). *Gypsy* and *Copia* superfamilies can be further classified into lineages and families with phylogenetic analysis of protein domain sequences that are usually supported by differences in structure (such as size of LTRs and elements) (Wicker et al. 2007). In land plants, four *Copia* (*Retrofit*, *Tork*, *Sire*, and *Oryco*) and six *Gypsy* (*Athila*, *Tat*, *Renia*, *CRM*, *Galadriel*, and *Del*) lineages were reported (Llorens et al. 2011).

Dongliang Du and Xiaoyun Du contributed equally to this work.

Communicated by W.-W. Guo

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s11295-018-1257-x) contains supplementary material, which is available to authorized users.

✉ Fred G. Gmitter, Jr
fgmitter@ufl.edu

[1] Citrus Research and Education Center, Institute of Food and Agricultural Sciences, University of Florida, Lake Alfred, FL 33850, USA

[2] Yantai Academy of Agricultural Science, Yantai 265500, Shandong, China

LTR-RTs can be used as a molecular marker system because of their high copy number, widespread distribution, and high heterogeneity (Kumar and Hirochika 2001). Several types of LTR-RT molecular markers have been developed, such as inter-retrotransposon amplified polymorphism (IRAP), sequence-specific amplified polymorphism (SSAP), retrotransposon-microsatellite amplified polymorphism (REMAP), insertion site-based polymorphism (ISBP), and retrotransposon-based insertion polymorphism (RBIP) (Flavell et al. 1998; Kalendar et al. 1999; Paux et al. 2010; Waugh et al. 1997). Most notably, IRAP markers amplify the intervening region between two retrotransposons to show polymorphisms (Kalendar et al. 2011; Kalendar et al. 1999). IRAP is used frequently because of the easy development of one outward-facing LTR-derived primer and generation of marker bands without digestion and ligation. SSAP exploits LTR-RT polymorphisms by amplifying the region between a retrotransposon and adjacent restriction site in the genome creating additional polymorphisms that can be used to differentiate closely related accessions (Syed et al. 2005). SSAP is especially useful for clone identification (Bretó et al. 2001; Venturi et al. 2006; Zhao et al. 2010). LTR-RT markers have been used widely for pedigree analysis, population structure, fingerprinting, linkage, and genetic mapping in several plant species (Branco et al. 2007; Farouji et al. 2015; Huo et al. 2009; Jia et al. 2009; Kalendar et al. 2011; Mandoulakani et al. 2015; Queen et al. 2004; Smykal 2006; Sun et al. 2015). The development of next-generation sequencing technologies has allowed for huge numbers of retrotransposon sequences to be generated, providing new opportunities for molecular marker development (Barghini et al. 2014; Cossu et al. 2012; Du et al. 2010; Xu and Du 2013; Zhang et al. 2012).

Many types of molecular markers have been used to characterize the phylogenetic relationships of *Citrus* accessions and relatives. Previously used markers include random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), and sequence-related amplified polymorphism (Barkley et al. 2006; Federici et al. 1998; Uzun et al. 2009; Yamamoto et al. 1993). These marker types could not fully reveal the origin and taxonomy of citrus. There are few studies that use LTR-RT based makers in citrus (Asins et al. 1999; Bernet and Asins 2003; Biswas et al. 2010a; Biswas et al. 2010b; De Felice et al. 2009; Rico-Cabanas and Martinez-Izquierdo 2007). However, small numbers of LTR-RT based primers were developed in these studies. In this study, we identified and characterized full-length LTR-RTs in the *C. x clementina* genome for use as molecular markers. The LTR-RT markers were tested for transferability among and differentiation between Rutaceae species. The fluorescence-labeled SSAP system used in this study yields a high-throughput and highly efficient mutant identification system for *Citrus*. To our knowledge, this is the first detailed report on high-throughput fluorescent SSAP in *Citrus*.

## Material and methods

### Genomic sequences and plant materials

The genomic sequences of *C. x clementina* (v1.0) and *C. x sinensis* were obtained from Phytozome v10 (http://phytozome.jgi.doe.gov) (Wu et al. 2014) and the orange genome annotation project (Xu et al. 2013), respectively. Twenty-five accessions fromthe genus *Citrus* (Table 1) and related genera in Rutaceae family (Table 2) were used for IRAP analysis. Young leaves were collected at the Florida Citrus Arboretum (Winter Haven, Florida) and stored at − 80 °C until used. For SSAP analysis, 27 sweet orange accessions (*C. x sinensis* Osb.) were used, including 'Valencia', 'Pineapple', and one irradiated bud mutant of 'Valencia' or 'Pineapple' (denoted as OR); four bud-derived clones generated from irradiated OR (denoted as'B-' followed by a number that indicates a different bud mutant); and 20 tissue culture-derived somaclones of OR (denoted as 'OLL-' followed by a number that indicates a unique somaclone). Each OLL somaclone was an independent regeneration event from an embryogenic callus line of the OR sweet orange.

### LTR-RT identification

Considering that the primary objective of this study was to isolate LTR-RTs for molecular marker development, full-length LTR-RTs were sufficient for this study. Solo LTR-RTs and LTR-RT remnants containing only one complete or fragmented LTR were not included, because they were considered to be derived from the recombination of full-length LTR-RTs (Ma and Bennetzen 2004). Full-length LTR-RTs were limited to those that contain a pair of relatively identical LTRs at both ends and at least one typical LTR-RT feature (PPT, PBS, and target site duplications (TSD)). LTR-RTs containing two LTRs and all three features were defined as intact LTR-RTs (Cossu et al. 2012).

LTR_FINDER and RepeatMasker v-4.0.3 were used to identify the full-length LTR-RTs from *C. x clementina* (Tempel 2012; Xu and Wang 2007). LTR sequence lengths from 100 to 5000 bp and a maximum distance between LTRs of 25,000 bp were parameters used for LTR_FINDER. RepeatMasker v-4.0.3 was used to find loci with high similarity to TEs in the Repbase *Viridiplantae* repeat library (Jurka et al. 2005). Crossmatch was used as search engine with a Smith-Waterman cutoff of 225. A perl tool, "one code to find them all" was used to annotate RepeatMaster output (Bailly-Bechet et al. 2014).

**Table 1** *Citrus* species and accessions used for IRAP analysis

| Groups | Latin name | | Common name | Accession | No[c] |
|---|---|---|---|---|---|
| | Tanaka[a] | Swingle and Reece[b] | | | |
| Pummelo | *C. maxima* (Burm.) Merr. | *C. maxima* (Brum.) Merr. | Pummelo | Ling Ping Yau pummelo | 1 |
| | *C. maxima* (Burm.) Merr. | *C. maxima* (Brum.) Merr. | Pummelo | Siamese sweet pummelo | 2 |
| Grapefruit | *C. paradisi* Macf. | *C. paradisi* Macf. | Grapefruit | Duncan grapefruit | 3 |
| | *C. paradisi* Macf. | *C. paradisi* Macf. | Grapefruit | Ruby red grapefruit | 4 |
| Mandarin | *C. clementina* Hort.ex Tanaka | *C. reticulata* Blanco | Clementine | Nules Clementine | 5 |
| | *C. nobilis* Lour. | *C. reticulata* Blanco | King | King mandarin | 6 |
| Citron and its relatives | *C. medica* L. | *C. medica* L. | Citron | Citron | 7 |
| | *C. medica* var. *sarcodactylis* Swing. | *C. medica* var. *sarcodactylis* Swing. | Citron | Buddha's hand Citron | 8 |
| | *C. volkameriana* Tan. | *C. limon* (L.) Burm.f. | Volkamer lemon | Volkamer lemon | 9 |
| | *C. limon* (L.) Burm.f. | *C. limon* (L.) Burm.f. | Lemon | Eureka lemon | 10 |
| | *C. aurantifolia* (Cristm.) Swingle | *C. aurantifolia* (Cristm.) Swing. | Lime | Mexican lime | 11 |
| | *C. latifolia* Tanaka | *C. aurantifolia* (Cristm.) Swing. | Bearss lime | Persian lime | 12 |
| Sour and sweet oranges | *C. aurantium* L. | *C. aurantium* L. | Sour orange | Willowleaf SO | 13 |
| | *C. aurantium* L. | *C. aurantium* L. | Sour orange | Sour orange | 14 |
| | *C. sinensis* (L.) Osbeck | *C. sinensis* (L.) Osbeck | Sweet orange | Pineapple | 15 |
| | *C. sinensis* (L.) Osbeck | *C. sinensis* (L.) Osbeck | Sweet orange | Valencia | 16 |
| | *C. myrtifolia* Raf. | *C. aurantium* L. | Myrtle-leaf orange | Chinotto | 17 |
| Papeda | *C. hystrix* DC. | *C. hystrix* DC. | Mauritius Papeda | Kaffir lime | 18 |
| | *C. ichangensis* Swingle | *C. ichangensis* Swingle | Ichang papeda | Ichang Papeda | 19 |

[a] Latin name using Tanaka's system

[b] Latin name using Swingle's system

[c] Accession numberz

Sequences that aligned with the query that had greater than 80 bp, 80% of their respective length, and 80% identity were used for downstream analysis. Redundant sequences and those containing N symbols were discarded. The boundaries and structures of all LTR-RTs were manually confirmed. To check the accuracy of the predicted LTR-RT, 46,339 BAC end sequences were downloaded from NCBI (Terol et al. 2008) and aligned to the identified LTR-RTs with BLASTN using default parameters. Sequences that aligned with the query with more than 99% of their respective length, and more than 99% identity, were considered a match to identified LTR-RTs.

## LTR-RT classification

The identified LTR-RTs were classified according to the methods previously described by Wicker et al. (Wicker et al. 2007). All putative LTRs were clustered into families using the USEARCH program with a similarity cutoff of 80% (Edgar 2010). The sequences between two putative LTRs

**Table 2** Rutaceae Species (excluding *Citrus*) used in this study

| Subfamily | Tribe | Subtribe | Group | Latin name | Accession | No[a] |
|---|---|---|---|---|---|---|
| Aurantioideae | Citreae | Citrinae | True citrus fruit trees | *Poncirus trifoliata* (L.) Raf. | Trifoliate | 20 |
| | | | | *Fortunella crassifolia* Swing. | Meiwa kumquat | 21 |
| | | | | *Eremocitrus glauca* Swing. | Australian desert lime | 22 |
| | | | | *Microcitrus australasica* Swing. | Australian finger lime | 23 |
| | | Balsamocitrinae | Bael fruit group | *Afraegle gabonensis* (Swing.) Engl. | Gabon powder flask | 24 |
| Rutoideae | | | | *Zanthoxylum fagara* (L.) Sarg. | Lime prickly ash | 25 |

[a] Accession number

were analyzed with BLASTN and BLASTX against Repbase (Camacho et al. 2009). Coding domains were located using hmmsearch with sequence threshold (-E): 10 and domain threshold (-domE): 10 (Bateman et al. 2004). The following pfam profiles were used: PF03732 and PF00098 for GAG; PF07727, PF05380 and PF00078 for RT; PF03078 for envelope protein (ENV); PF02022, PF00665 and PF00552 for IN; PF00077 and PF00026 for PR; PF00075 for RH (Wang and Liu 2008).

## Phylogenetic analysis

The RT domains of 49 reference elements representing known *Viridiplantae* LTR-RT lineages were downloaded from the Gypsy Database (GyDB) (Llorens et al. 2011). Representative RT protein sequences of each citrus LTR-RT family (for 350 *Copia* and 178 *Gypsy* families) were randomly selected and aligned with the RT domains mentioned above using default parameters in Clustal Omega (Sievers et al. 2011). Two neighbor-joining phylogenetic trees were constructed separately for the *Gypsy* and *Copia* superfamilies using MEGA 5.05 with a *p* distance model and 100 bootstrap iterations (Tamura et al. 2007).

## Estimation of insertion time

The two LTRs of a LTR-RT are considered to be identical at the time of insertion. Therefore, the insertion time of intact LTR-RTs can be estimated from the sequence divergence of 5′ and 3′ LTRs (Ma and Jackson 2006). Two LTRs were aligned first using Clustal Omega, and the Jukes-Cantor distance ($k$) was calculated using the PHYLIP program dnadist (Retief 2000). The insertion time ($T$) of an intact LTR-RT was calculated using the formula: $T = k/2r$. A substitution rate of $2.4 \times 10^{-9}$ mutations per site per year, 2-fold higher than determined for the genes in poplar, was used in this study (Ma and Jackson 2006; Tuskan et al. 2006).

## IRAP primers design

To validate in silico analysis and to test the transferability of Clementine LTR-RT based markers, 17 IRAP primers were designed following the protocols published by Schulman AH (Kalendar et al. 1999; Kalendar and Schulman 2006). Twenty-five LTR-RT families including at least two members were randomly selected, and their homologs in sweet orange were identified using BLASTN (sequence identify > 90% and length > 90% of query LTRs). The LTRs within one family, along with their homologs from sweet orange, were aligned using Clustal Omega. At least one conserved region for 17 LTR-RT families was identified and used for primer design with Primer Premier 5.0 (Singh et al. 1998). Primer sequences are listed in Online Resource 1.

## DNA extraction and IRAP analysis

Total genomic DNA was extracted from leaves using the CTAB method (Doyle 1991). DNA concentration was measured with a Nanodrop spectrophotometer (Biotek Instruments, Winooski, VT). IRAP analysis was done as described by Kalendar et al. (1999). Amplification reactions were done in a 20 μl solution containing 0.25 mM dNTPs, 0.4 μM primer, 100 ng genomic DNA, 2.0 mM $MgCl_2$, 1 × PCR buffer, and 1 U Taq DNA polymerase (Promega, USA). The amplifications were carried out in a Bio-Rad thermal cycler using the following amplification profile: 1 cycle at 95 °C, 5 min; 30 cycles at 95 °C, 1 min; 55 °C 1 min; ramp + 0.5 °C $s^{-1}$ to 72 °C; 72 °C, 2 min + 5 s per cycle; 1 cycle at 72 °C, 8 min. PCR products were separated on a 2% agarose gel at 8 V/cm using 0.5X Tris-borate-EDTA buffer, stained with ethidium bromide, and photographed using SYNGENE Automated Gel Documentation System (Cambridge, USA). At least two PCR amplifications were conducted for each sample and only reproducible bands between 200 and 2000 bp were used for downstream analysis.

## SSAP analysis

Three SSAP forward primers were designed as mentioned above. Restriction, ligation, and pre-amplification reactions were done as described by Waugh et al. (1997). Selective amplification was conducted with a retrotransposon primer in combination with either *Mse* I + 3 or *Eco*R I + 3 (Online Resource 1). A total of 48 primer combinations were used (3 SSAP primers * 2 enzymes * 8 selective bases). SSAP analysis was repeated two times for each primer pair. Following selective amplification, 0.5 μl of PCR amplicons was added to a mixture with 9.25 μl Hi-Di formamide and 0.2 μl GenScan 500 LIZ molecular weight markers (Applied Biosystems, Foster City, CA) for 3 min at 94 °C and immediately placed on ice for 6 min. Samples were fluorescently labeled with an ABI PRISM 3130 xl Genetic Analyzer (Applied Biosystems, Foster City, CA) as follows: POP-7™ polymer at 63 °C, sample injection voltage was 1.6 kV with 12 s injection time, and 10 kV run voltage for 7200 s.

## Data analysis

Raw fluorescent SSAP data were analyzed and visualized using Genemarker v.4.0 software (SoftGenetics LLC®, State College, PA). A minimum fluorescence threshold value of 250 was chosen. Peaks between 60 and 500 bp were included in the analysis. The bin table output of peak areas called was transferred to an MS® Excel spreadsheet. The peak areas were converted to 0 and 1 scores indicating peak/marker absence and presence. For IRAP analysis, binary matrices (presence/absence) were prepared from electrophoretic patterns. The

simple matching (SM) similarity coefficient was calculated with the SIMQUAL module. Dendrograms were built by cluster analysis using the unweighted pair-group method with arithmetic averages (UPGMA) and the SAHN clustering program. The FIND module was used to identify all trees that resulted from different choices of tied similarity or dissimilarity values. The clustering goodness-of-fit to the data matrix was calculated by the programs COPH and MXCOMP. Figures were generated with the PROJECTIONS module. All analyses were performed with the software NTSYS-pc 2.10e (Rohlf 1992).

## Results

### Identification of full-length LTR-RTs

A total of 3959 full-length LTR-RTs were identified in the Clementine genome (Online Resources 2 and 3). LTR_FINDER and RepeatMasker identified 3791 and 1099 full-length LTR-RTs (931 in common), respectively. Intact LTR-RTs containing two LTRs and the three features (PPT, PBS, and TSD) comprised 40.7% (1612) of total LTR-RTs. Most of LTR-RTs in this study (3836) had at least two of the three features. Most full-length LTR-RTs (3593) were terminated by the highly conserved TG-CA boxes at both 5′ and 3′ ends of LTRs. The mean length of full-length LTR-RTs was 8.08 kb, with a standard deviation of 4.50 kb. For LTRs, the mean length was 781.69 bp, and the standard deviation was 572.57 bp. The accuracy of identified LTR-RTs was confirmed with the analysis of 28.6 Mb BAC end sequences, which corresponds to 8% of the Clementine genome (Terol et al. 2008). About half of full-length LTR-RTs (1738) matched at least one BAC end sequence. The total length of matched region was 3.5 Mb, corresponding to 11% of the identified LTR-RTs.

### Classification of full-length LTR-RTs

The 3959 full-length LTR-RTs were first classified into two superfamilies (*Gypsy*, *Copia*, or unknown) according to their protein domain organization. As shown in Table 3, 1285 and 1727 LTR-RTs were included in *Copia* and *Gypsy* superfamilies, respectively. Then, the full-length LTR-RTs were clustered into 1074 families based on LTR sequence similarity, including 386 *Copia* and 214 *Gypsy* families. The mean number of full-length LTR-RTs per family was 3.69, and the largest family had 282 LTR-RTs. The average size of the *Gypsy* families (8.07) was approximately two times higher than the *Copia* families (3.33).

To understand the evolutionary relationships among LTR-RT families, two phylogenetic trees were constructed separately for *Copia* and *Gypsy* superfamilies (Figs. 1 and 2). Three hundred fifty *Copia* families were grouped into four previously defined lineages: *Retrofit*, *Tork*, *Sire*, and *Oryco*. The *Sire* lineage and *Oryco* lineages were first clustered together, then grouped with *Tork* and *Retrofit* lineages. One hundred seventy-eight *Gypsy* families fell into six lineages: *Athila*, *Tat*, *Renia*, *CRM*, *Galadriel*, and *Del*. The four lineages-*Renia*, *CRM*, *Galadriel*, and *Del* of Branch 1, also called chromoviruses, were first clustered together, then grouped with *Athila* and *Tat*, two lineages of Branch 2. Family numbers of *Retrofit* and *Tork* lineages were almost equally represented within the *Copia* superfamily (Table 3). The *Athila* was the largest lineage and contained 814 LTR-RTs and was the most highly represented in *Gypsy* superfamily (Table 3).

The size distributions of LTR-RTs and LTR length varied among lineages (Table 3).The average LTR-RT length of *Sire* lineage within the *Copia* superfamily was greater than that of other lineages. The average LTR length in the *Sire* lineage was over three times larger than LTRs in the *Oryco* and *Retrofit* lineages. The LTR-RT sizes of *Athila*, *Del*, and *Tat* in the *Gypsy* superfamily were larger than the other three lineages. *Del* lineage was found to have the largest LTR length variation.

### Distribution of full-length LTR-RTs on citrus chromosomes

Of the 3959 full-length LTR-RTs identified in this study, 3218 (81.3%) were anchored to nine Clementine linkage groups (Fig. 3). These LTR-RTs occupied 28.1 Mb of genome sequence, accounting for 9.74% of the nine currently assembled chromosomes (288.6 Mb) (Online Resource 4). LTR-RTs were distributed throughout the genome and there was little variation in full-length LTR-RT density between the nine chromosomes. LTR-RTs, especially in the *Gypsy* superfamily, were more abundant in putative centromeric regions.

### Putative insertion time of intact LTR-RTs

The insertion time distribution of LTR-RTs fits an exponential decay curve ($r$ square = 0.96, $p < 0.01$) (Fig. 4a). This pattern was expected, because intact LTRs were rapidly changed to solo LTRs, truncated LTRs, or completely eliminated from the genome. The half-life of intact LTR-RTs in citrus was estimated to be 3.47 Myr. Most citrus LTR-RTs (73.6%) were amplified in the last 10 Myr, and 519 LTR-RTs were inserted within last 2.5 Myr. Significant “peaks” representing the insertion time of different LTR-RTs families showed that these families were active over a short period of time, especially for recently amplified families (Fig. 4b). Almost all LTR-RTs in family 271 were inserted within the last 2.5 Myr. Four of the most recently inserted 271 family members (12.5%) were found to have two identical LTRs, which suggests there was little time to accumulate evolutionary mutations (Online Resource 3).

**Table 3** General features of *C. x clementina* LTR-RT lineages

| | | | | Element | | Family | |
|---|---|---|---|---|---|---|---|
| Superfamily | Lineage | Size (kb)[a] | LTR-len (bp)[a] | No. | % | No. | % |
| Copia | | 7.02 (5.12–8.44) | 625 (294–783) | 1285 | 100 | 386 | 100 |
| | Oryco | 6.32 (4.61–5.24) | 312 (269–365) | 57 | 4.44 | 14 | 3.63 |
| | Retrofit | 5.87 (4.90–5.34) | 309 (225–321) | 380 | 29.57 | 165 | 42.75 |
| | Sire | 9.63 (8.53–10.2) | 1041 (966–1292) | 207 | 16.11 | 30 | 7.77 |
| | Tork | 6.5 (5.3–6.64) | 669 (472–667) | 525 | 40.86 | 141 | 36.53 |
| | Unknown | 8.82 (5.94–10.5) | 870 (424–1298) | 116 | 9.03 | 36 | 9.33 |
| Gypsy | | 9.54 (5.6–12.4) | 940 (472–1352) | 1727 | 100 | 214 | 100 |
| | Athila | 10.91 (8.42–12.5) | 1175 (1076–1383) | 814 | 47.13 | 64 | 29.91 |
| | CRM | 7.18 (5.43–6.83) | 509 (379–678) | 91 | 5.27 | 15 | 7.01 |
| | Del | 9.31 (7.93–10) | 1145 (303–2047) | 41 | 2.37 | 15 | 7.01 |
| | Galadriel | 5.79 (3.13–6.18) | 836 (363–1082) | 220 | 12.74 | 16 | 7.48 |
| | Renia | 5.62 (5.07–5.56) | 420 (342–473) | 107 | 6.2 | 40 | 18.69 |
| | Tat | 9.97 (7.99–11.13) | 716 (664–795) | 302 | 17.49 | 28 | 13.08 |
| | Unknown | 10.99 (5.81–13.74) | 838 (436–1312) | 152 | 8.8 | 36 | 16.82 |
| Unknown | Unknown | 6.85 (2.74–9.02) | 707 (212–938) | 947 | | 474 | |
| Total | | 8.08 (5.02–11.28) | 782 (329–1240) | 3959 | | 1074 | |

[a] Average, with 25th and 75th percentiles in parenthesis

## Polymorphisms of 25 Rutaceae accessions

All IRAP primers achieved successful amplification across the 25 Rutaceae accessions. In total, 209 reproducible and unambiguous bands were produced, ranging from 250 to 2000 bp in size. Most of these bands (205, 98.09%) were polymorphic. Six to 18 fragments were amplified from a single primer (Table 4). Representative patterns of four LTR primers are shown in Online Resource 5. The genetic similarity coefficients between the 25 accessions were calculated. The average similarity coefficients for the 25 accessions ranged from 0.493 to 0.997 with a mean of 0.650.

Using 0.659 as a threshold, the phylogenetic tree could be split into two major clusters (Fig. 5). Cluster 1 included all of accessions in *Citrus* except *C. hystrix*, indicating a greater genetic distance from the other *Citrus* species analyzed. All the other genera were included in Cluster 2. The two clusters could be further divided into four and two groups, respectively. Within Cluster 1, accessions from the same species were grouped together. *C. ichangensis* and *C. maxima* alone formed two single clusters: Cluster 1–1 and Cluster 1–2. *C. reticulata*, *C. sinensis*, *C. paradisi*, and *C. aurantium* were grouped in Cluster 1–3. *C. medica*, *C. limon*, and *C. aurantifolia* shared Cluster 1–4. In Cluster 2, four of the six *Citrus*-related genera examined (*Eremocitrus*, *Microcitrus*, *Afraegle*, and *Zanthoxylum*) were clustered together and formed an independent group (Cluster 2–2), which is in accordance with their remarkable phenotypic differences from *Citrus*. The other two genera, *Poncirus* and *Fortunella*, were grouped into Cluster 2–1. Comparatively speaking, Cluster 2–1 displayed a much closer relationship with *Citrus* genus than Cluster 2–2. Therefore, if 0.587 was used as a cut-off value for defining the clusters, Cluster 2–1 and Cluster 1 would be classified in the same group, clearly separated from Cluster 2–2. The correlation between the similarity coefficient matrix and the cophenetic matrix derived from the UPGMA tree was 0.94, corresponding to a good fit.

## Polymorphisms of 27 sweet orange accessions

For SSAP analysis, 24 out of 48 primer combinations generated easily readable patterns that were selected for the downstream analyses (Table 4). A single primer combination produced 17 to 198 bands. A total of 2156 amplification products were generated, of which 1518 (70%) were polymorphic. The primer combinations exhibited different levels of polymorphism ranging from 23.86 to 98%. Genetic similarity coefficients among 27 accessions were calculated using the SSAP analysis data. The distribution of genetic similarity coefficients between 'Pineapple', 'Valencia', and 24 clones (including four 'B-' accessions, and 20 'OLL-') were compared in Fig. 6. The 24 clones had a closer relationship to "Valencia" than "Pineapple" ($p < 0.01$, Student's $t$ test). The average genetic similarity coefficients among the 24 clones was 0.796, implying profound levels of genetic differentiation within these clones.

# Discussion

Different lineage classification systems were used for LTR-RTs in different studies. We chose to use the classification

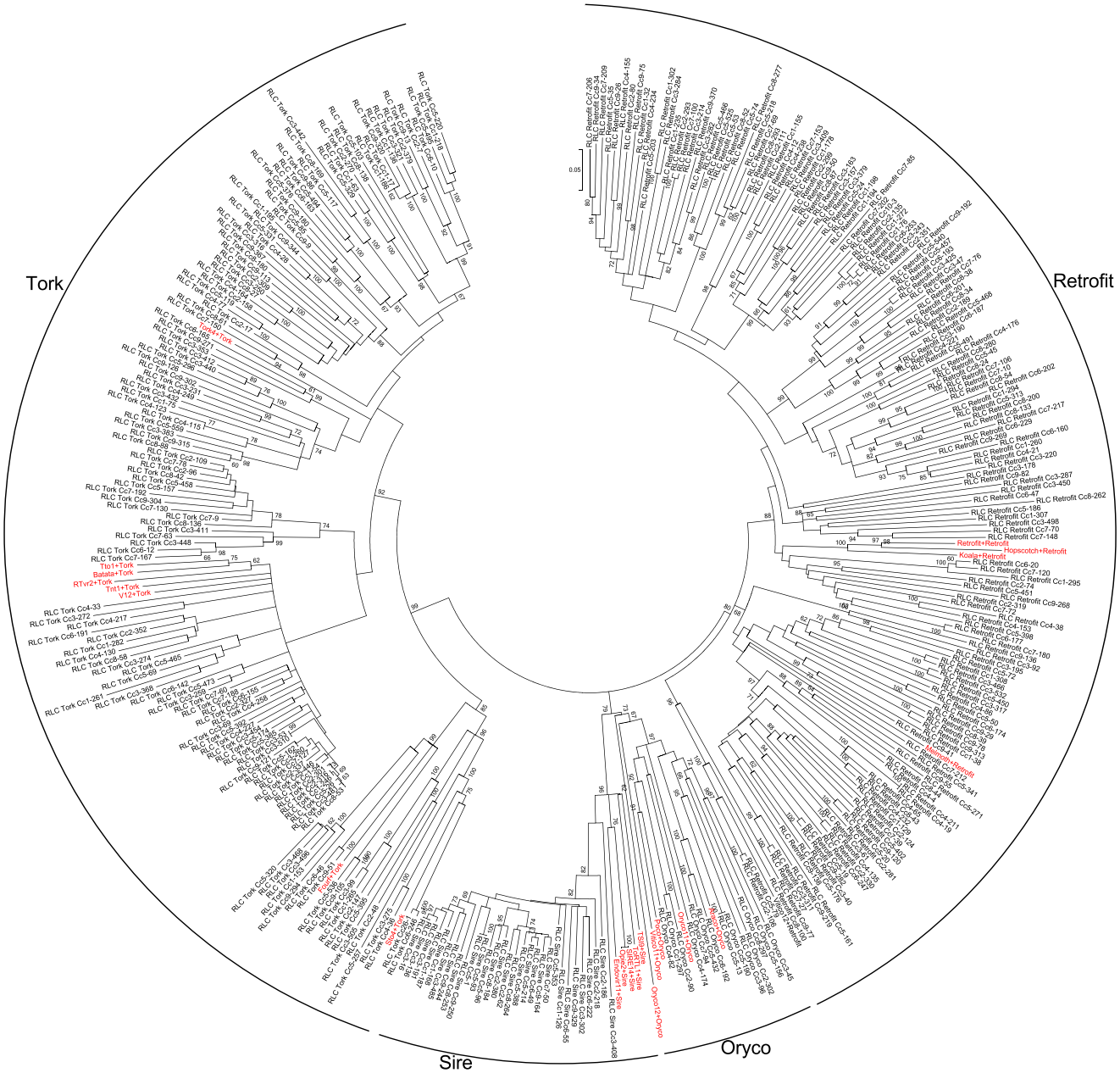


**Fig. 1** Neighbor-joining phylogeny of *Copia* families based on reverse transcriptase. One representative element that contains a complete RT domain was chosen for each of the 350 *Copia* families. Reference sequences from GyDB are denoted with a plus symbol and shown in red. Bootstrap values below 60% are not shown

system from GyDB, a research project focused on the phylogenetic classification of transposable elements (Llorens et al. 2011). Four and six lineages were reported in *Copia* and *Gypsy* superfamilies of land plants, respectively. All these lineages were found in the *C. x clementina* genome, but the full-length LTR-RT numbers within each lineage varied greatly. The two smallest lineages, *Del* and *Oryco*, were found in the *C. x clementina* genome, containing 41 and 57 full-length LTR-RTs, respectively. However, the largest lineage, *Athila*, included 814 full-length LTR-RTs and accounted for 47.13% of the *Gypsy* elements identified. Different species have distinct LTR-RT compositions. A pattern similar to citrus was found in Arabidopsis, where 59.3% of *Gypsy* elements belonged to the *Athila* lineage (Du et al. 2010; Marco and Marin 2008). However, only 0.1% of the *Gypsy* elements identified in rice were classified into the *Athila* lineage. The largest lineage in rice was *Tat*, accounting for 55.8% of its *Gypsy* elements (Du et al. 2010). The ratio of *Gypsy* to *Copia* elements in Clementine is 1.34:1, and is much lower than rice (4.9:1) (Tian et al. 2009) and sorghum (3.7:1) (Paterson et al. 2009), but is similar to soybean (1.4:1) (Du et al. 2010) and maize (1.6:1) (Baucom et al. 2009).

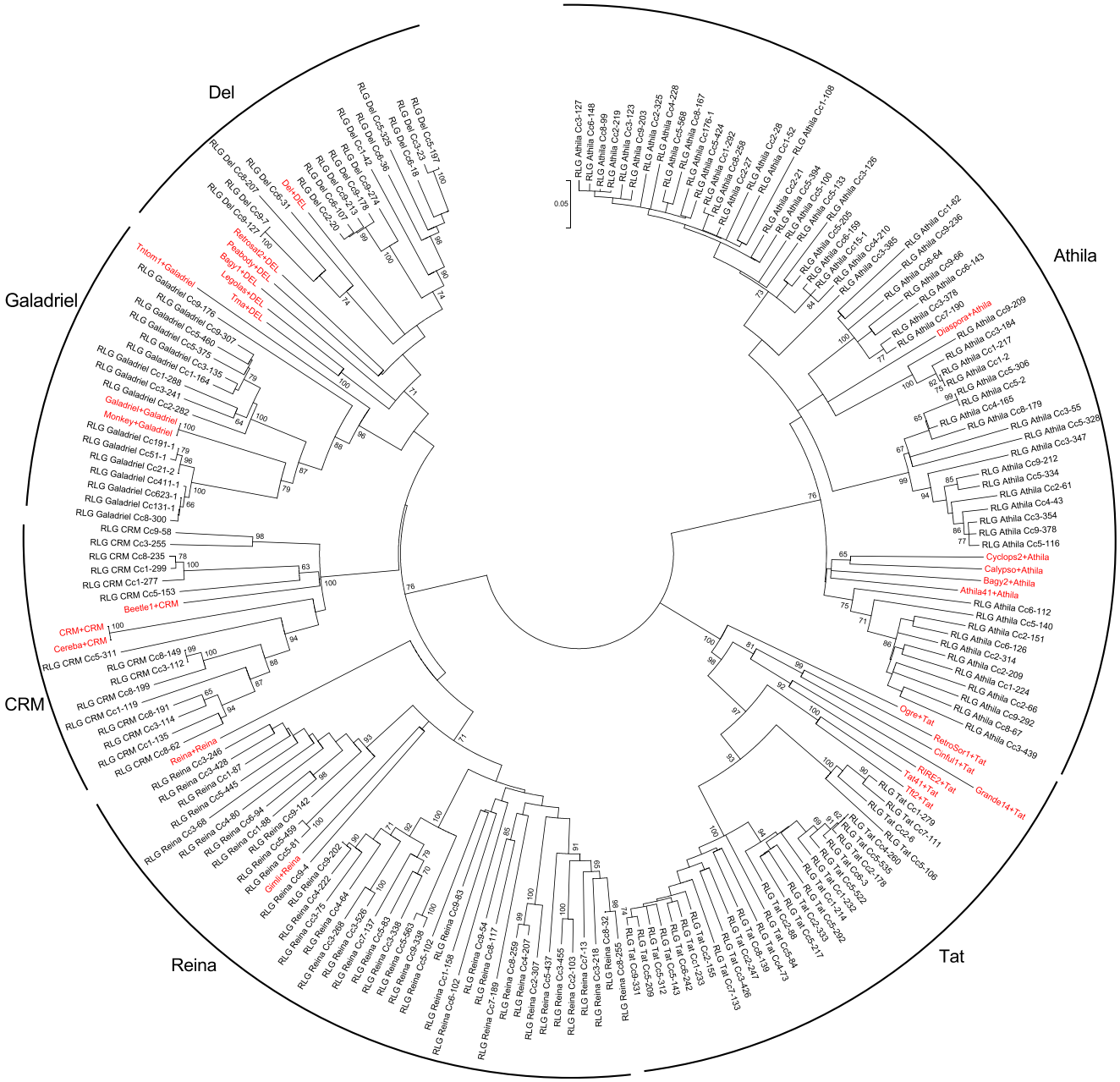


**Fig. 2** Neighbor-joining phylogeny of *Gypsy* families based on reverse transcriptase. One representative element that contains a complete RT domain was chosen for each of the 178 *Gypsy* families. Reference sequences from GyDB are denoted with a plus symbol and shown in red. Bootstrap values below 60% are not shown

Previous studies show that LTR-RT based primers can be used for closely related genera (Kalendar et al. 2011; Kalendar et al. 1999). In this study, all designed IRAP primers can be transferred across the 25 Rutaceae species. The transferability of pear IRAP to other Rosaceae species ranged from 87.5 to 100% (Sun et al. 2015). The transferability of IRAP markers is usually higher than other makers. For SSR markers, approximately half of the primers developed in sweet orange can be used in pummelo or lemon (Biswas et al. 2014). Similar results were reported in the Poaceae and Rosaceae species (Mamaghani et al. 2015; Sun et al. 2015). One reason for the high transferability of IRAP markers in this study may be that the primers were designed in conserved regions within LTR-RT families, not only in conserved regions of orthologous LTR-RTs. We selected LTR-RT families of different sizes for primer design contrary to an IRAP study in pear, where the largest sized LTR-RT families were selected for primer design (Sun et al. 2015). The largest sized full-length LTR-RT families are not

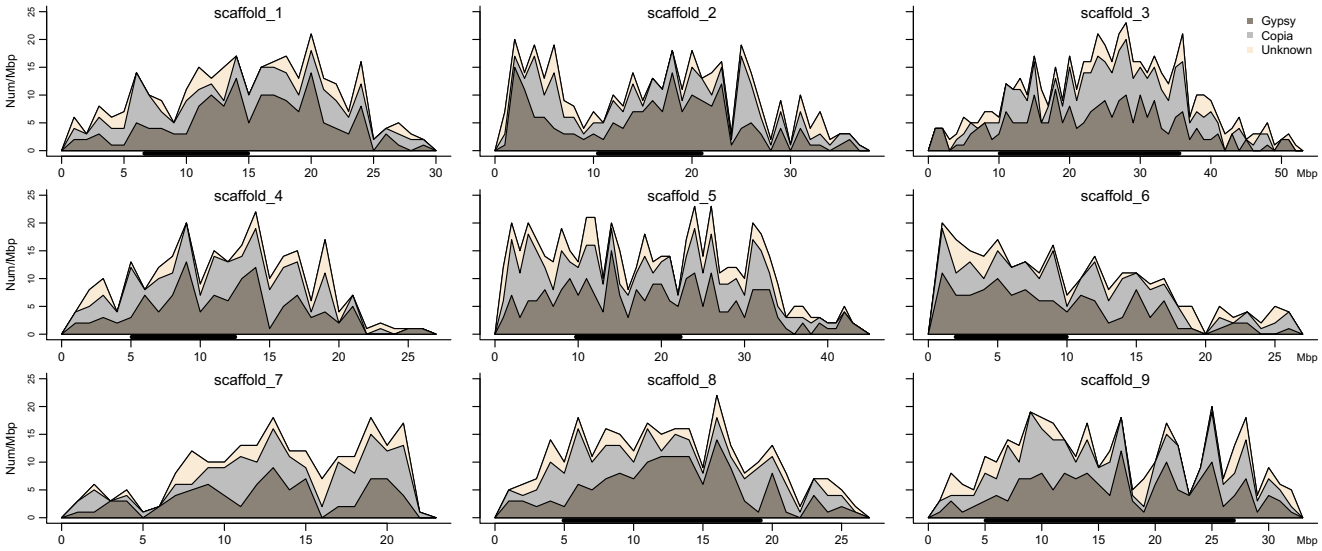


**Fig. 3** Distribution of full-length *Gypsy*, *Copia*, and unknown LTR-RTs in the 9 linkage groups of *C. x clementina*. The putative positions of centromeres are based on the results of Wu et al., and shown by black boxes

representative because they do not include homologous or clustered LTR-RTs in the genome.

According to the classification of Swingle and Reece (Swingle and Webber 1943), the "true citrus fruit trees" group (Citrinae) is divided into six genera: *Citrus*, *Fortunella*, *Poncirus*, *Clymenia*, *Eremocitrus*, and *Microcitrus*. *Citrus* is the most economically important, and is sexually compatible with all other genera. The UPGMA cluster analysis showed that *Fortunella* and *Poncirus* were closer to *Citrus* than were *Eremocitrus* and *Microcitrus*. This result was consistent with the findings by Pang et al. based on AFLP markers (Pang et al. 2007), but differ from the results obtained by Garcia-Lor et al. (2013) and Wu et al. (2018). The *Citrus* genus was divided into two subgenera *Citrus* and *Papeda*, according to the classification system of Swingle and Reece (Swingle and Webber 1943). In the UPGMA phylogenetic tree, *C. hystrix* formed one cluster separated from other *Citrus* species that suggests some *Papeda* species may be the most primitive *Citrus* (Nicolosi et al. 2000). *C. ichangnesis* was first clustered with pummelo and mandarin, and then clustered with citron supporting the hypothesis that *C. ichangnesis* could be an ancestor of mandarin, and is supported by the results of Xie et al. (2008) and Pang et al. (2007). According to our results, the citron is the most distantly related species among them, which is confirmed by the sequencing of their chloroplast genomes (Carbonell-Caballero et al. 2015).

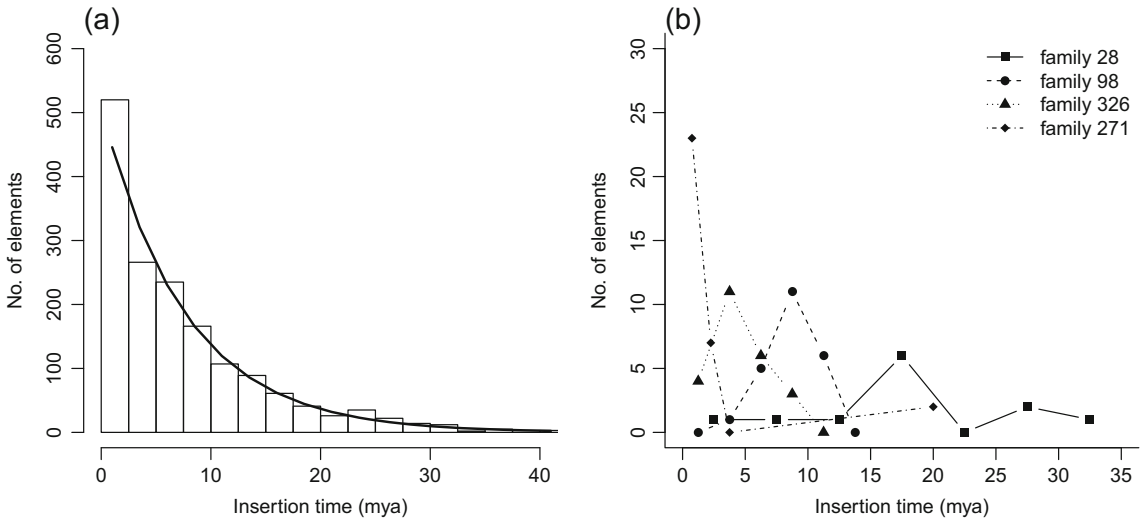


**Fig. 4** Insertion time of intact LTR-RTs in *C. x clementina*. **a** Insertion time of intact LTR-RTs (mya, million years ago). **b** Insertion time of intact LTR-RTs within four selected families

**Table 4** Summary of PCR amplification corresponding to individual primer in IRAP and SSAP analysis

| Primer | Total bands | Polymorphic bands | Polymorphic bands (%) | Primer | Total bands | Polymorphic bands | Polymorphic bands (%) |
|---|---|---|---|---|---|---|---|
| LTR03 | 11 | 11 | 100 | SSAP1/E-AGC | 198 | 163 | 82 |
| LTR05 | 6 | 5 | 83.33 | SSAP1/M-CAG | 33 | 22 | 66.67 |
| LTR06 | 12 | 12 | 100 | SSAP1/M-CTA | 17 | 9 | 53 |
| LTR07 | 10 | 9 | 90.00 | SSAP1/M-CTG | 26 | 16 | 61.54 |
| LTR08 | 10 | 10 | 100 | SSAP1/M-CTT | 42 | 29 | 69 |
| LTR09 | 8 | 7 | 87.50 | SSAP2/E-AAC | 144 | 92 | 63.89 |
| LTR10 | 11 | 11 | 100 | SSAP2/E-AAG | 135 | 80 | 59 |
| LTR11 | 14 | 14 | 100 | SSAP2/E-ACA | 135 | 121 | 90 |
| LTR12 | 13 | 13 | 100 | SSAP2/E-ACG | 159 | 153 | 96 |
| LTR13 | 13 | 13 | 100 | SSAP2/E-ACT | 159 | 123 | 77 |
| LTR14 | 18 | 18 | 100 | SSAP2/E-AGC | 141 | 138 | 98 |
| LTR17 | 12 | 12 | 100 | SSAP2/E-AGG | 176 | 147 | 84 |
| LTR19 | 13 | 13 | 100 | SSAP2/M-CAA | 77 | 45 | 58 |
| LTR20 | 17 | 17 | 100 | SSAP2/M-CAC | 95 | 53 | 56 |
| LTR21 | 13 | 13 | 100 | SSAP2/M-CAG | 31 | 25 | 81 |
| LTR25 | 12 | 12 | 100 | SSAP2/M-CAT | 182 | 75 | 41 |
| LTR29 | 16 | 15 | 93.75 | SSAP2/M-CTA | 95 | 73 | 76.84 |
| | | | | SSAP2/M-CTG | 88 | 21 | 23.86 |
| | | | | SSAP2/M-CTT | 27 | 14 | 52 |
| | | | | SSAP3/E-ACA | 19 | 11 | 58 |
| | | | | SSAP3/E-AGC | 44 | 34 | 77 |
| | | | | SSAP3/M-CAA | 65 | 46 | 70.77 |
| | | | | SSAP3/M-CTG | 34 | 17 | 50.00 |
| | | | | SSAP3/M-CTT | 34 | 11 | 32 |
| Total | 209 | 205 | 98.09 | Total | 2156 | 1518 | 70 |

The SSAP markers showed that the 24 sweet orange clones had patterns that were closer to 'Valencia' than 'Pineapple', and this was supported by phenotypic traits (data not shown). The SSAP markers also showed that the

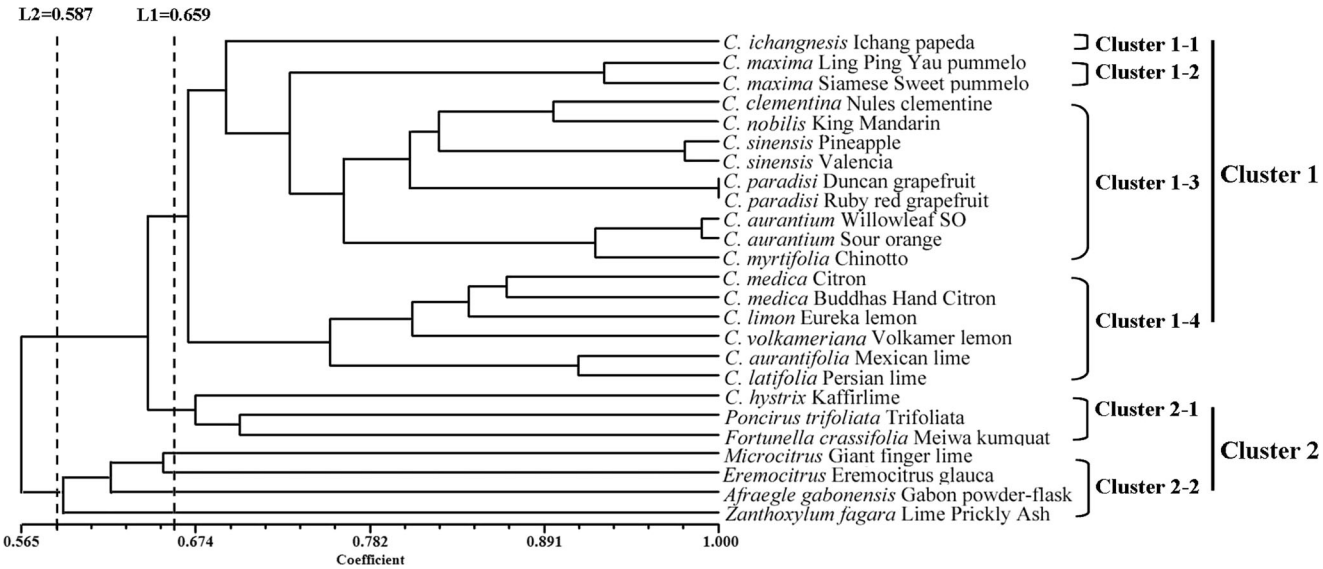


**Fig. 5** Dendrogram of 25 accessions by the UPGMA cluster analysis based on IRAP analysis

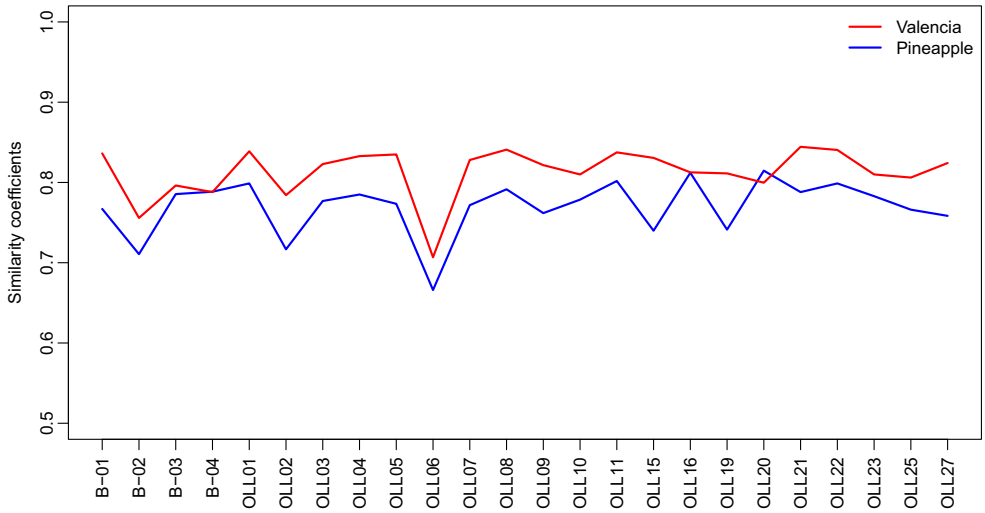


**Fig. 6** The distribution of genetic similarity coefficients between 'Valencia' or 'Pineapple' and 24 clones based on SSAP analysis

24 clones had abundant genetic variations which would be helpful for future variety registration and protection of intellectual property. Many widely grown citrus cultivars, such as sweet orange, grapefruit, lemon, and various clonal selections of Satsuma and Clementine mandarins, originated as either bud sport or apomictic seedling mutations (Rao et al. 2009). The use of traditional markers to distinguish such mutant clones is difficult because of lower genetic variability associated with these maker systems. The LTR-RT based SSAP markers were especially useful for distinguishing accessions with similar genetic backgrounds, like bud sport mutations in citrus (Venturi et al. 2006; Zhao et al. 2010). In our previous study, 18 SSR primer sets were used to distinguish these 27 accessions. However, all accessions showed similar PCR amplification patterns and were indistinguishable (data now shown). In *C. x clementina*, SSAP was successfully used to distinguish 24 accessions generated from bud mutations; other markers (ISSR, RAPD, AFLP, and SSR) tested could not (Bretó et al. 2001). Compared to other routinely used molecular markers such as SSR, LTR-RT based SSAP markers are better suited for genetic relationship analysis and phylogenetic analysis (Biswas et al. 2011; Schulman et al. 2012). An advantage of SSAP is that polymorphisms at multiple loci are detected in a single assay, while SSR usually detects polymorphisms at one locus (Powell et al. 1996). Retrotransposons are important sources of variation in *Citrus*, especially in the species mentioned above (Bretó et al. 2001; Wang et al. 2017). Several Sicilian blood oranges arose by insertion of a *Copia*-like retrotransposon (Tcs-1) adjacent to a MYB transcriptional activator (named Ruby) for anthocyanin production. The LTR of Tcs-1 provides a novel promoter that activates *Ruby* expression in the flesh of the fruit, resulting in distinctive red coloration in response to cold conditions. A Chinese blood orange ('Jingxian') was also found to contain an independent insertion of a similar retrotransposon that confers tissue-specific red coloration also in response to cold conditions (Butelli et al. 2012).

## Conclusion

Full-length LTR-RTs were mined from the Clementine genome and classified based on structural details. Randomly selected IRAP and SSAP markers were tested and showed that they could differentiate citrus accessions and mutant clones. Our findings indicate that LTR-RTs are an excellent molecular marker resource because they are easy to develop, polymorphic, widely distributed, and transferable within the Rutaceae family.

**Funding information** This work was supported by a grant from the Citrus Research and Development Foundation (CRDF-724) and the China Scholarship Council.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.



## References

Asins MJ, Monforte AJ, Mestre PF, Carbonell EA (1999) *Citrus* and *Prunus copia*-like retrotransposons. Theor Appl Genet 99:503–510. https://doi.org/10.1007/s001220051263

Bailly-Bechet M, Haudry A, Lerat E (2014) "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. Mob DNA 5:13. https://doi.org/10.1186/1759-8753-5-13

Barghini E, Natali L, Giordani T, Cossu RM, Scalabrin S, Cattonaro F, imkova H, Vrana J, Dole el J, Morgante M, Cavallini A (2014) LTR retrotransposon dynamics in the evolution of the olive (*Olea europaea*) genome. DNA Res 22:91–100. https://doi.org/10.1093/dnares/dsu042

Barkley NA, Roose ML, Krueger RR, Federici CT (2006) Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). Theor Appl Genet 112:1519–1531. https://doi.org/10.1007/s00122-006-0255-9

Bateman A et al (2004) The Pfam protein families database. Nucleic Acids Res 32:D138–D141. https://doi.org/10.1093/nar/gkh121

Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, Westerman RP, SanMiguel PJ, Bennetzen JL (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet 5:e1000732. https://doi.org/10.1371/journal.pgen.1000732

Bernet GP, Asins MJ (2003) Identification and genomic distribution of *gypsy* like retrotransposons in *Citrus* and *Poncirus*. Theor Appl Genet 108:121–130. https://doi.org/10.1007/s00122-003-1382-1

Biswas MK, Baig MNR, Cheng Y, Deng X (2010a) Retrotransposon based genetic similarity within the genus *Citrus* and its relatives. Genet Resour Crop Evol 57:963–972. https://doi.org/10.1007/s10722-010-9533-0

Biswas MK, Xu Q, Deng X (2010b) Utility of RAPD, ISSR, IRAP and REMAP markers for the genetic analysis of *Citrus* spp. Sci Hortic (Amsterdam) 124:254–261. https://doi.org/10.1016/j.scienta.2009.12.013

Biswas MK, Chai L, Amar MH, Zhang X, X-x D (2011) Comparative analysis of genetic diversity in Citrus germplasm collection using AFLP, SSAP, SAMPL and SSR markers. Sci Hortic (Amsterdam) 129:798–803. https://doi.org/10.1016/j.scienta.2011.06.015

Biswas MK, Xu Q, Mayer C, Deng X (2014) Genome wide characterization of short tandem repeat markers in sweet orange (*Citrus sinensis*). PLoS One 9(8):e104182. https://doi.org/10.1371/journal.pone.0104182

Branco CJS, Vieira EA, Malone G, Kopp MM, Malone E, Bernardes A, Mistura CC, Carvalho FIF, Oliveira CA (2007) IRAP and REMAP assessments of genetic similarity in rice. J Appl Genet 48:107–113. https://doi.org/10.1007/bf03194667

Bretó MP, Ruiz C, Pina JA, Asíns MJ (2001) The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species. Mol Phylogenet Evol 21:285–293. https://doi.org/10.1006/mpev.2001.1008

Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell 24:1242–1255. https://doi.org/10.1105/tpc.111.095232

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST plus: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421

Carbonell-Caballero J, Alonso R, Ibanez V, Terol J, Talon M, Dopazo J (2015) A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. Mol Biol Evol 32:2015–2035. https://doi.org/10.1093/molbev/msv082

Cossu RM, Buti M, Giordani T, Natali L, Cavallini A (2012) A computational study of the dynamics of LTR retrotransposons in the *Populus trichocarpa* genome. Tree Genet Genomes 8:61–75. https://doi.org/10.1007/s11295-011-0421-3

De Felice B, Wilson R, Argenziano C, Kafantaris I, Conicella C (2009) A transcriptionally active *copia*-like retroelement in *Citrus limon*. Cell Mol Biol Lett 14:289–304. https://doi.org/10.2478/s11658-008-0050-5

Domingues DS, Cruz GMQ, Metcalfe CJ, Nogueira FTS, Vicentini R, Alves CS, Van Sluys M-A (2012) Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. BMC Genomics 13:137. https://doi.org/10.1186/1471-2164-13-137

Doyle J (1991) DNA protocols for plants. In: Hewitt G, Johnston AB, Young JP (eds) Molecular techniques in taxonomy, vol 57. Springer Berlin Heidelberg, pp 283–293

Du J et al (2010) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. Plant J 63:584–598. https://doi.org/10.1111/j.1365-313X.2010.04263.x

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Farouji AE, Khodayari H, Saeidi H, Rahiminejad MR (2015) Genetic diversity of diploid *Triticum* species in Iran assessed using inter-retroelement amplified polymorphisms (IRAP) markers. Biologia 70:52–60. https://doi.org/10.1515/biolog-2015-0002

Federici CT, Fang DQ, Scora RW, Roose ML (1998) Phylogenetic relationships within the genus *Citrus* (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. Theor Appl Genet 96:812–822. https://doi.org/10.1007/s001220050807

Flavell AJ, Knox MR, Pearce SR, Ellis THN (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J 16:643–650. https://doi.org/10.1046/j.1365-313x.1998.00334.x

Garcia-Lor A, Curk F, Snoussi-Trifa H, Morillon R, Ancillo G, Luro F, Navarro L, Ollitrault P (2013) A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the "true citrus fruit trees" group (Citrinae, Rutaceae) and the origin of cultivated species. Ann Bot 111:1–19. https://doi.org/10.1093/aob/mcs227

Huo HQ, Conner JA, Ozias-Akins P (2009) Genetic mapping of the apospory-specific genomic region in *Pennisetum squamulatum* using retrotransposon-based molecular markers. Theor Appl Genet 119:199–212. https://doi.org/10.1007/s00122-009-1029-y

Jia H, Liu Y, Qu S, Gao Z, Wang K, Zhang Z (2009) Establishment and optimization of the inter-retrotransposon amplified polymorphism (IRAP) molecular marker system in apple. J Fruit Sci 26:254–257

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update: a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467. https://doi.org/10.1159/000084979

Kalendar R, Schulman AH (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nat Protoc 1:2478–2484. https://doi.org/10.1038/nprot.2006.377

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman A (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98:704–711. https://doi.org/10.1007/s001220051124

Kalendar R, Flavell AJ, Ellis THN, Sjakste T, Moisy C, Schulman AH (2011) Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106:520–530. https://doi.org/10.1038/hdy.2010.93

Kumar A, Hirochika H (2001) Applications of retrotransposons as genetic tools in plant biology. Trends Plant Sci 6:127–134. https://doi.org/10.1016/s1360-1385(00)01860-4

Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu JM, Tamarit D, Aguilar-Rodriguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, Latorre A, Moya A (2011) The *Gypsy* database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res 39:D70–D74. https://doi.org/10.1093/nar/gkq1061

Ma JX, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci U S A 101:12404–12410. https://doi.org/10.1073/pnas.0403715101

Ma JX, Jackson SA (2006) Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice. Genome Res 16:251–259. https://doi.org/10.1101/gr4583106

Mamaghani RA, Mohammadi SA, Aharizad S (2015) Transferability of barley retrotransposon primers to analyze genetic structure in Iranian *Hypericum perforatum* L. populations. Turk J Bot 39:664–672. https://doi.org/10.3906/bot-1405-76

Mandoulakani BA, Yaniv E, Kalendar R, Raats D, Bariana HS, Bihamta MR, Schulman AH (2015) Development of IRAP- and REMAP-derived SCAR markers for marker-assisted selection of the stripe rust resistance gene Yr15 derived from wild emmer wheat. Theor Appl Genet 128:211–219. https://doi.org/10.1007/s00122-014-2422-8

Marco A, Marin I (2008) How *Athila* retrotransposons survive in the Arabidopsis genome. BMC Genomics 9:219

Nicolosi E, Deng ZN, Gentile A, La Malfa S, Continella G, Tribulato E (2000) Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theor Appl Genet 100:1155–1166. https://doi.org/10.1007/s001220051419

Pang X, Hu C, Deng X (2007) Phylogenetic relationships within *Citrus* and its related genera as inferred from AFLP markers. Genet Resour Crop Evol 54:429–436. https://doi.org/10.1007/s10722-006-0005-5

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556. https://doi.org/10.1038/nature07723

Paux E, Faure SÃ©, Choulet FÃ©Ã©, Roger D, Gauthier VÃ©, Martinant JP, Sourdille P, Balfourier FÃ§, le Paslier MC, Chauveau AÃ©, Cakir M, Gandon BÃ©, Feuillet C (2010) Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnol J 8:196–210. https://doi.org/10.1111/j.1467-7652.2009.00477.x

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238. https://doi.org/10.1007/bf00564200

Queen RA, Gribbon BM, James C, Jack P, Flavell AJ (2004) Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Mol Gen Genomics 271:91–97. https://doi.org/10.1007/s00438-003-0960-x

Rao MN, Soneji JR, Gmitter FG (2009) Citrus breeding. In: Jain SM, Priyadarshan PM (eds) Breeding plantation tree crops: temperate species. Springer, New York

Retief JD (2000) Phylogenetic analysis using PHYLIP. Methods Mol Biol 132:243–258

Rico-Cabanas L, Martinez-Izquierdo JA (2007) CIRE1, a novel transcriptionally active *Ty1-copia* retrotransposon from *Citrus sinensis*. Mol Gen Genomics 277:365–377. https://doi.org/10.1007/s00438-006-0200-2

Rohlf FJ (1992) NTSYS-pc. Numerical taxonomy and multivariate analysis system, version 1.70. Setauket, New York,

Schulman AH, Flavell AJ, Paux E, Ellis THN (2012) The application of LTR retrotransposons as molecular markers in plants. In: Bigot Y (ed) Mob genet elements. Humana Press, Totowa, pp 115–153. https://doi.org/10.1007/978-1-61779-603-6_7

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75

Singh VK, Mangalam AK, Dwivedi S, Naik S (1998) Primer premier: program for design of degenerate primers from a protein sequence. BioTechniques 24:318–319

Smykal P (2006) Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. J Appl Genet 47:221–230. https://doi.org/10.1007/bf03194627

Sun J, Yin H, Li L, Song Y, Fan L, Zhang S, Wu J (2015) Evaluation of new IRAP markers of pear and their potential application in differentiating bud sports and other Rosaceae species. Tree Genet Genomes 11. https://doi.org/10.1007/s11295-015-0849-y

Swingle WT, Webber HJ (1943) The botany of citrus and its wild relatives of the orange subfamily (family Rutaceae, subfamily Aurantioideae). University of California Press, Berkeley

Syed NH, Sureshsundar S, Wilkinson MJ, Bhau BS, Cavalcanti JJV, Flavell AJ (2005) Ty1-*copia* retrotransposon-based SSAP marker development in cashew (*Anacardium occidentale* L.). Theor Appl Genet 110:1195–1202. https://doi.org/10.1007/s00122-005-1948-1

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. https://doi.org/10.1093/molbev/msm092

Tempel S (2012) Using and understanding RepeatMasker. Methods Mol Biol 859:29–51. https://doi.org/10.1007/978-1-61779-603-6_2

Terol J, Naranjo MA, Ollitrault P, Talon M (2008) Development of genomic resources for *Citrus clementina*: characterization of three deep-coverage BAC libraries and analysis of 46,000 BAC end sequences. BMC Genomics 9:423. https://doi.org/10.1186/1471-2164-9-423

Tian Z, Rizzon C, du J, Zhu L, Bennetzen JL, Jackson SA, Gaut BS, Ma J (2009) Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? Genome Res 19:2221–2230. https://doi.org/10.1101/gr.083899.108

Todorovska E (2007) Retrotransposons and their role in plant genome evolution. Biotechnol Biotechnol Equip 21:294–305

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa*. Science 313:1596–1604. https://doi.org/10.1126/science.1128691

Uzun A, Yesilo T, Aka-Kacar Y, Tuzcu O, Gulsen O (2009) Genetic diversity and relationships within *Citrus* and related genera based on sequence related amplified polymorphism markers (SRAPs). Sci Hortic (Amsterdam) 121:306–312. https://doi.org/10.1016/j.scienta.2009.02.018

Venturi S, Dondini L, Donini P, Sansavini S (2006) Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. Theor Appl Genet 112:440–444. https://doi.org/10.1007/s00122-005-0143-8

Wang H, Liu J-S (2008) LTR retrotransposon landscape in *Medicago truncatula*: more rapid removal than in rice. BMC Genomics 9: 382. https://doi.org/10.1186/1471-2164-9-382

Wang X, Xu Y, Zhang S, Cao L, Huang Y, Cheng J, Wu G, Tian S, Chen C, Liu Y, Yu H, Yang X, Lan H, Wang N, Wang L, Xu J, Jiang X, Xie Z, Tan M, Larkin RM, Chen LL, Ma BG, Ruan Y, Deng X, Xu Q (2017) Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat Genet 49:765–772. https://doi.org/10.1038/ng.3839

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

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. https://doi.org/10.1038/nrg2165

Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X, Ruiz M, Scalabrin S, Terol J, Takita MA, Labadie K, Poulain J, Couloux A, Jabbari K, Cattonaro F, del Fabbro C, Pinosio S, Zuccolo A, Chapman J, Grimwood J, Tadeo FR, Estornell LH, Muñoz-Sanz JV, Ibanez V, Herrero-Ortega A, Aleza P, Pérez-Pérez J, Ramón D, Brunel D, Luro F, Chen C, Farmerie WG, Desany B, Kodira C, Mohiuddin M, Harkins T, Fredrikson K, Burns P, Lomsadze A, Borodovsky M, Reforgiato G, Freitas-Astúa J, Quetier F, Navarro L, Roose M, Wincker P, Schmutz J, Morgante M, Machado MA, Talon M, Jaillon O, Ollitrault P, Gmitter F, Rokhsar D (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat Biotechnol 32:656–662. https://doi.org/10.1038/nbt.2906

Wu GA, Terol J, Ibanez V, López-García A, Pérez-Román E, Borredá C, Domingo C, Tadeo FR, Carbonell-Caballero J, Alonso R, Curk F, du D, Ollitrault P, Roose ML, Dopazo J, Gmitter FG, Rokhsar DS, Talon M (2018) Genomics of the origin and evolution of *Citrus*. Nature 554:311–316. https://doi.org/10.1038/nature25447

Xie R, Zhou Z, Deng L (2008) Taxonomic and phylogenetic relationships among the genera of the true Citrus fruit trees group (Aurantioideae, Rutaceae) based on AFLP markers. J Syst Evol 46:682–691. https://doi.org/10.3724/sp.j.1002.2008.07135

Xu Y, Du J (2013) Genome-wide identification and evolutionary analysis of LTR-retrotransposons in potato. Genomics Appl Biol 32:734–742

Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35:W265–W268. https://doi.org/10.1093/nar/gkm286

Xu Q, Chen LL, Ruan X, Chen D, Zhu A, Chen C, Bertrand D, Jiao WB, Hao BH, Lyon MP, Chen J, Gao S, Xing F, Lan H, Chang JW, Ge X, Lei Y, Hu Q, Miao Y, Wang L, Xiao S, Biswas MK, Zeng W, Guo F, Cao H, Yang X, Xu XW, Cheng YJ, Xu J, Liu JH, Luo OJ, Tang Z, Guo WW, Kuang H, Zhang HY, Roose ML, Nagarajan N, Deng XX, Ruan Y (2013) The draft genome of sweet orange (*Citrus sinensis*). Nat Genet 45:59–66

Yamamoto M, Kobayashi S, Nakamura Y, Yamada Y (1993) Phylogenic relationships of citrus revealed by RFLP analysis of mitochondrial and chloroplast DNA. JPN J Breed 43:355–365

Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, Tao Y, Wang J, Yuan Z, Fan G, Xing Z, Han C, Pan H, Zhong X, Shi W, Liang X, du D, Sun F, Xu Z, Hao R, Lv T, Lv Y, Zheng Z, Sun M, Luo L, Cai M, Gao Y, Wang J, Yin Y, Xu X, Cheng T, Wang J (2012) The genome of *Prunus mume*. Nat Commun 3:1318. https://doi.org/10.1038/ncomms2290

Zhao G, Dai H, Chang L, Ma Y, Sun H, He P, Zhang Z (2010) Isolation of two novel complete Ty1-*copia* retrotransposons from apple and demonstration of use of derived S-SAP markers for distinguishing bud sports of *Malus domestica* cv. Fuji. Tree Genet Genomes 6:149–159. https://doi.org/10.1007/s11295-009-0236-7