Cytogenetic and Genome Research

Original Article

Cytogenet Genome Res 2020;160:272–282
DOI: 10.1159/000506222

Accepted: January 21, 2020
Published online: June 10, 2020

# Sequence Evolution, Abundance, and Chromosomal Distribution of Ty1-copia Retrotransposons in the *Saccharum spontaneum* Genome

Shan Yang[a] Kai Zeng[a] Ke Chen[a] Xinwang Zhao[a] Jiayun Wu[b] Yongji Huang[a] Muqing Zhang[c] Zuhu Deng[a, c]

[a]National Engineering Research Center for Sugarcane, Fujian Agriculture and Forestry University, Fuzhou, China; [b]Guangdong Provincial Bioengineering Institute, Guangzhou Sugarcane Industry Research Institute, Guangzhou, China; [c]State Key Laboratory for protection and utilization of subtropical agro-bioresources, Guangxi University, Nanning, China

**Keywords**
Evolutionary analysis · Reverse transcriptase · *Saccharum spontaneum* · Sugarcane · Ty1-copia retrotransposons

**Abstract**
*Saccharum spontaneum* is a wild germplasm resource of the genus *Saccharum* that has many valuable traits. Ty1-copia retrotransposons constitute a large proportion of plant genomes and affect genome sequence organization and evolution. This study aims to analyze the sequence heterogeneity, phylogenetic diversity, copy number, and chromosomal dispersion patterns of Ty1-copia retrotransposons in *S. spontaneum*. A total of 44 Ty1-copia reverse transcriptase subclones isolated from *S. spontaneum* showed a range of heterogeneity, and all sequences were A-T rich, averaging approximately 54.59%. Phylogenetic analysis divided the 44 reverse transcriptase sequences into 5 distinct lineages (Retrofit/Ale, Sire/Maximus, Bianca, Tork/TAR, and Ty1-copia like). Dot-blot hybridization revealed that Ty1-copia retrotransposons consisted of a significant component of approximately 38,900 copies and 16,300 copies per genome in the accessions YN82–114 (2n = 10x = 80) and AP85–441 (2n = 4x = 32), respectively. The results of a local blast analysis showed that there are 15,069 Ty1-copia retrotransposon copies in the genome of AP85–441, of which the Retrofit/Ale lineage had the highest copy number, followed by the Tork/TAR, Sire/Maximus, and Bianca lineages. Furthermore, both FISH and the local blast analysis with AP85–441 genomic data demonstrated that the Ty1-copia retrotransposons were unevenly distributed throughout the chromosomes. Taken together, this study provides insights into the role of Ty1-copia retrotransposons in the evolution and organization of the *S. spontaneum* genome. 

Sugarcane (*Saccharum* spp. L.) is one of the most important sugar and bio-energy crops in the world [Li et al., 2017]. There are 6 species in the genus *Saccharum*, including *S. officinarum* (2n = 80), *S. sinense* (2n = 112–120), *S. barberi* (2n = 82–124), *S. edule* (2n = 60, 70, 80), and 2 wild species, *S. robustum* (2n = 60–120) and *S. spontaneum* (2n = 40–128) [Irvine, 1999]. The commercial sugarcane cultivars are multiple interspecific hybrids consisting of 10–20% chromosomes from *S. spontaneum* and 80–90% chromosomes from *S. officinarum* [D'Hont

S.Y. and K.Z. contributed equally to this work.

karger@karger.com
www.karger.com/cgr



Zuhu Deng
National Engineering Research Center for Sugarcane
Fujian Agriculture and Forestry University
Fuzhou, 350002 (China)
E-Mail dengzuhu@163.com

et al., 1996]. Therefore, sugarcane has a highly polyploid and aneuploid genome with a complete set of homologous genes predicted to range from 10 to 12 copies and to include representatives from each of 2 different species [Souza et al., 2011]. The chromosome number in *S. spontaneum* ranges from 40 to 128 with a basic chromosome number of x = 8 [D'Hont et al., 1998; Zhang et al., 2012]. *S. spontaneum* provides a vast gene pool for many important agronomic traits that are not available in modern commercial sugarcane, including an advanced root system, wide adaptability, tillering ability, and strong disease resistance, as well as many other desirable properties [Liu et al., 2014].

Transposable elements (TEs) are a major component of eukaryotic genomes. TEs can move and replicate within their host genomes to induce chromosomal rearrangements and influence genome evolution [Flavell, 1992; Kumar and Bennetzen, 1999]. Retrotransposons are the largest class of TEs, which are ubiquitous in most plant genomes. They play significant roles in the evolution of gene content, gene order, and gene expression through the introduction of extensive mutations and genome rearrangements [Arumuganathan and Earle, 1991; Flavell et al., 1997; Feschotte et al., 2002]. Retrotransposons can be divided into 2 major groups: long terminal repeats (LTRs) and non-LTR retrotransposons according to their structural features and evolutionary origin [Temin, 1981; Kumar and Bennetzen, 1999]. The structure of retrotransposons is similar to that of retroviruses, consisting of LTRs, a primer binding site, a polypurine tract coding group, and specific antigen and polyprotein structures [Temin, 1981]. Additionally, based on sequence similarity and the order of polyprotein region domains, retrotransposons can be subdivided into 2 major types, Ty1-copia and Ty3-gypsy. Genes in the polyprotein region of Ty1-copia are ordered as protease, integrase, reverse transcriptase (RT), and ribonuclease H, whereas the order for Ty3-gypsy is protease, RT, ribonuclease H, and integrase [Kumar and Bennetzen, 1999]. Ty1-copia families can be divided into 6 major common evolutionary lineages, namely Tork/TAR, Tork/Angela, Sire/Maximus, Retrofit/Ale, Oryco/Ivana, and Bianca [Wicker and Keller, 2007].

In plants, the copy number of retrotransposons is generally high, and they typically account for more than half of the nuclear genome [SanMiguel et al., 1996; Baucom et al., 2009]. Moreover, it has been reported that amplification of retrotransposons makes a strong contribution to genome evolution [Sormacheva and Blinov, 2011; Kawahara et al., 2013]. For example, through recent bursts of the LTR-RT family, the *Oryza granulata* genome size increased by 50% [El Baidouri et al., 2014]. Retrotransposons can also act as potential controlling elements by inserting near or within genes and can also generate mutations that alter gene expression or the structure of the resulting gene product [Kashkush et al., 2003; Kawakami et al., 2004]. In addition, LTR retrotransposon genes tend to accumulate with preference in the pericentromeric regions of host genomes where they play an important role in maintaining chromatin structure or promoting chromosomal rearrangements, as well as regulating centromere function and gene expression [Havecker et al., 2004]. With the development of plant genome sequencing, unprecedented opportunities are now available for the structural and evolutionary analysis of LTR retrotransposons in plants.

In the past few decades, *S. spontaneum* germplasm resource research had mainly focused on improvements in the breeding of cultivated sugarcane through the utilization of superior genes from *S. spontaneum*. A major advance forward occurred with the availability of reference genomic sequences and information concerning the evolution of the *S. spontaneum* genome (AP85–441, 2n = 4x = 32) in 2018 [Zhang et al., 2018]. However, many anti-reverse genes remain undiscovered in *S. spontaneum*, and details for retrotransposons in the *S. spontaneum* genome are lacking. Here, we describe a comprehensive study of Ty1-copia retrotransposon sequences in *S. spontaneum* and explore their diversity, heterogeneity, phylogenetic relationship, copy number, and chromosomal distribution patterns to gain new insights into the genome of *S. spontaneum*.

## Materials and Methods

*Plant Material and DNA Isolation*

*S. spontaneum* plants, YN82–114 (2n = 10x = 80) and AP85–441 (2n = 4x = 32), were obtained from the Fujian Agricultural and Forestry University (Fuzhou, China). Genomic DNA (gDNA) from young leaves of the plants was extracted using a cetyltrimethyl/ammonium bromide (CTAB) standard protocol [Porebski et al., 1997].

*PCR Amplification of Retrotransposon RT Fragments*

The degenerate primers Ty1-F: 5'-ACNGCNTTYYTNCAYGG-3', Ty1-R: 5'-ARCATRTCRTCNACRTA-3' were used to amplify RT domains of Ty1-copia group LTR retrotransposons [Flavell, 1992]. The PCR reactions included 20 μM of each primer, 0.2 mM of each dNTP, 1× ExTaq buffer, 2.5 U ExTaq polymerase (Takara, Bio Inc, Tokyo, Japan), and 50 ng gDNA with sterile distilled water added to a final volume of 50 μL. PCR reactions involved denaturation at 94°C for 3 min, followed by 35 cycles of 30 s at

94°C, 45 s at 45°C and 30 s at 72°C, with a final elongation step of 10 min at 72°C. PCR products were separated on 1.5% agarose gels and visualized under UV light after ethidium bromide staining.

*Cloning and Sequencing*

The target PCR products were recovered and purified using a DNA purification kit (Takara, Japan). The purified fragments were then cloned into the pMD19-T vector (Takara, Japan), which was used to transform competent *Escherichia coli* DH5α cells. Plasmid DNA was extracted from 49 colonies, and the presence of inserts of desired length was verified by PCR with M13 primers [Park et al., 2007]. The cloned fragments were sequenced by the Beijing Genomics Institute Co., Ltd. (Beijing, China). RT sequences were termed Ssp-Ty1–1 to Ssp-Ty1–44 and deposited in the GenBank database under accession numbers MH743729–MH743772.

*DNA Sequence Analysis*

RT sequences were initially searched against published sequences (BLAST) in the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) to identify homologies with retrotransposons of other graminaceous species. Ty1-copia RT nucleotide sequences were then translated into putative amino acid sequences, which were compared with multiple homologous sequences from graminaceous species and available conserved plant retrotransposon peptide sequences for Ty1-copia (TAFLHG and YVDDML) using a MACSE tool [Flavell et al., 1992]. RT domain frameshift mutations were detected using MUSCLE [Wu, 1978]. Whenever necessary, gaps were introduced to retain open reading frames (ORFs), and PCR primer regions were excluded from the multiple sequence alignment analysis. Finally, according to the amino acid sequence for the Ty1-copia LTR retrotransposon domain, all phylogenetic analyses were constructed according to the p-distance using Neighbor-Joining methods in Molecular Evolutionary Genetics Analysis (MEGA) 6.0 software. The pairwise similarity matrix of sequences were evaluated by BioEdit 7.0.9.0 software.

*Southern Hybridization and Dot-Blot Analysis*

Copy numbers of the Ty1-copia retrotransposon in the *S. spontaneum* genome were determined using dot-blot hybridization by applying a series of dilutions of gDNA and the complete heterogeneous population of sequences. The gDNA was diluted to 300, 200, 150, 100, 75, 50, 37.5, and 25 ng, and PCR products of the Ty1-copia RT sequences were diluted to 5, 2.5, 1.25, 0.8, 0.6, 0.4, 0.2, and 0.1 ng, denatured in a 100°C water bath for 5 min, and then transferred to an ice water bath for 10 min. The denatured gDNA and PCR products were spotted onto Amersham HybondNC nylon membranes (GE Healthcare). The gDNA was immobilized on the membranes by UV cross-linking, and then the membrane was washed with high pressure distilled water and air dried. DIG-labeled Ty1-copia genomic DNA of *S. spontaneum* was used as a probe with the DIG Nick Translation Kit (Roche Diagnostics) following the manufacturer's instructions. Hybridization was performed at 42°C with immersion for 18 h in the DIG DNA hybridization probe mixture. The membranes were washed and signals were quantified with a ChemiDocXRS instrument (Bio-Rad, Hercules, CA, USA). The hybridization intensity was converted to optical density using ImageJ software [Schneider et al., 2012]. The quantity of Ty1-copia and host gDNA was calculated and named 'Q1' and 'Q2', respectively, when the optical density of Ty1-copia and host gDNA were postulated the same value. The mean DNA content of YN82–114 and AP85–441 were estimated as 8.19 pg/2C and 3.44 pg/2C, respectively [Zhang et al., 2012]. The quantity of Ty1-copia RT sequences could be estimated and named 'M' using the calculation equation: *Q1/Q2 = M/(Mean DNA content per genome)*. The Ty1-copia copy number per genome was estimated by dsDNA copy number calculator (http://cels.uri.edu/gsc/cndna.html) when the 'M' value and the average length of Ty1-copia RT sequences were provided.

*Preparation of Chromosome Slides*

Root tips (2–3 cm) were harvested and placed in a saturated p-dichlorobenzene solution at 25°C for 2 h to accumulate and concentrate chromosomes. The roots were then fixed in an ethanol:acetic acid (3:1) solution. The fixed root tips were washed in citric acid buffer (6 mM sodium citrate, 4 mM citric acid, pH 4.6) and stored in 70% ethanol solution at –20°C until use. The cell walls of root tips were digested with cellulose and pectinase (8% Onozuka R10 cellulose, 2% pectolyase Y-23, and 2% pectinase) at 37°C for 120 min. The meristematic cells of root tips were squashed on a clean slide in a drop of fixation solution with a ratio of ethanol to acetic acid of 3:1 (v/v), then air dried, and stored at –20°C until use.

*FISH Analysis*

Probes were made from a mixture of PCR-amplified Ty1-copia RT fragments and labelled with digoxigenin-11-dUTP. The probe mixture per slide, consisting of 50–100 ng of the labeled Ty1-copia probe, 50% deionized formamide, 0.5% sodium dodecyl sulphate (SDS), 10% dextran sulfate, and 2× SSC, was denatured at 97°C for 10 min and then incubated in ice water for 5 min. The chromosomal slides were denatured in a Hybaid Omnislide (Hybaid) at 80°C for 3 min. Subsequently, the slides were dehydrated using a series of ice-cold ethanol solutions (70, 95, and 100% ethanol) and air dried before application of the heat-denatured hybridization mixture for 5 min at each stage, followed by hybridization overnight at 37°C. The digoxigenin-labeled probe was detected by Rhodamine anti-DIG-sheep. Finally, chromosomes were counterstained with 30 μL/slide Vectashield antifade solution (Vector Laboratories, concentration of 10 μg/mL DAPI) and mounted with a coverslip. The FISH signal was examined with an AxioScope A1 Imager fluorescent microscope (Carl Zeiss, Gottingen, Germany). Images were analyzed for DAPI blue and Texas red channel using Axiovision software.

*Blast with Local Sequences*

The genomic data of AP85–441 (2n = 4x = 32) were downloaded from http://www.life.illinois.edu/ming/downloads/Spontaneum_genome/. We then built a blast database with AP85–441 genomic data using blast+/2.7.1. A total of 44 Ty1-copia RT sequences of YN82–114 as the query sequences were searched in the AP85–441 genome using blast+/2.7.1. The blast results were visualized using a web tool called Map Gene 2 Chromosome V2.0 (http://mg2c.iask.in/mg2c_v2.0/). The copy number of each Ty1-copia RT sequence was calculated in the AP85–441 genome using Microsoft Office Excel 2013 after blast analysis with a local AP85–441 genomic database.

## Results

*Isolation and Sequence Characterization of Ty1-copia RT Fragments*

The ~260 bp long Ty1-copia RT-specific PCR fragments were cloned, and 49 individual subclones were randomly selected for further sequencing. After aligning the sequences, 44 unique clones termed Ssp-Ty1–1 to Ssp-Ty1–44 were obtained for the final analyses (online suppl. Table 1; see www.karger.com/doi/10.1159/000506222 for all online suppl. material). The fragment sizes for the 44 sequences ranged from 246 to 265 bp. These cloned nucleotide sequences had a high A-T content with an average of 54.59%, and the pairwise similarity among these sequences ranged from 43.6 to 100% (online suppl. Tables 1, 2). Thus, the 44 Ty1-copia RT sequences showed high heterogeneity.

The Ty1-copia sequences were translated into amino acids and analyzed for the presence of stop codons and frameshifts in the coding regions. Among the 44 unique sequences, 28 sequences (64%) were 'defective' due to the presence of premature stop codons and/or frameshift mutations. The alignment of the putative amino acid sequences showed that the conceptual ORFs of the 44 sequences corresponded to the conserved peptide motifs that are expected to lie upstream of TAFLHG and downstream of YVDDML. Pairwise comparisons showed that the similarity of the amino acid sequences ranged from 25 to 100% (online suppl. Table 3). The low level of amino acid sequence similarity indicated the presence of multiple Ty1-copia element sequences in the *S. spontaneum* genome.

For further visual representation of the sequence type, we studied the phylogenetic relationships using the nucleotide sequences. The results showed that these sequences were divided into 4 subgroups (Fig. 1). There were 23 sequences, 11 sequences, 4 sequences, and 6 sequences in subgroup I, subgroup II, subgroup III, and subgroup IV, respectively. However, based on phylogenetic analysis using Ty1-copia RT amino acid sequences, the 44 sequences can be divided into 3 subgroups, whereby subgroup I, subgroup II, and subgroup III contained 18 sequences, 7 sequences, and 19 sequences, respectively (Fig. 2). Therefore, the phylogenetic relationships had some differences by using nucleotide and amino acid sequences.

*Phylogenetic Analysis of Ty1-copia RT Sequences*

To further investigate the Ty1-copia type of *S. spontaneum*, we conducted evolutionary analyses using various types of Ty1-copia amino acid sequences from different plants and Ty1-copia amino acid sequences of *S. spontaneum* (Fig. 3). The amino acid sequences were separated into 7 distinct subgroups (I–VII). There were 12 Ty1-copia RT sequences (27.3%) in subgroup I that clustered with *Saccharum* hybrid Retrofit/Ale 5.1. Subgroup III contained 6 Ty1-copia RT sequences (13.6%) that clustered with *S. arundinaceus* Sire/Maximus, and these sequences were very similar to each other, of which the SSP-TY1–14 sequence was more closely related to the *Saccharum* hybrid cultivar R570. Three Ty1-copia RT sequences (6.8%) in subgroup IV clustered with *Eleocharis quinqueflora* Bianca and *Sorghum bicolor*. Four Ty1-copia RT sequences formed 2 clades in group VI, of which 1 clade included SSP-TY1–8, SSP-TY1–21 and the *Saccharum* hybrid Tork/TAR, while the other contained SSP-TY1–32, SSP-TY1–44, and *Oryza sativa* (M94492.1). There were 19 Ty1-copia RT sequences (43.2%) in subgroup VII, in which most of Ty1-copia RT sequences clustered with the Ty1-copia-like sequence of *Saccharum* hybrid cultivar SP80–3280 (MF737008.1). However, one sequence (SSP-TY1–41) was closely related to the Ty1-copia-like RT sequence of *Setaria verticillata* (AF227058.1). Subgroup II (Oryco/Ivana) and subgroup V (Tork/Angela) were not detectable in the *S. spontaneum* genome.

*Abundance of Ty1-copia Retrotransposons in the S. spontaneum Genome*

According to the signal intensity from the dot-blot hybridization, approximately 38,900 and 16,300 copies of Ty1-copia retrotransposons were present per genome of YN82–114 and AP85–441, respectively (online suppl. Fig. 1, 2). Compared with these results, the total copy number of Ty1-copia retrotransposons in the YN82–114 genome was 2.38 times higher than that of AP85–441, indicating that the copy number of Ty1-copia retrotransposons had a close relationship with the ploidy of *S. spontaneum*. The total copy number of the 44 Ty1-copia retrotransposons was 15,069 in the AP85–441 genome based on the local blast analysis with AP85–441 genomic data (Table 1). Moreover, the copy number for the Retrofit/Ale lineage of Ty1-copia retrotransposons had the highest in the AP85–441 genome, followed by the Ty1-copia like, Tork/TAR lineage and Bianca lineage.

*Localization of Ty1-copia Retrotransposons in the S. spontaneum Genome*

FISH of labelled Ty1-copia PCR products demonstrated an uneven distribution of hybridization signals along all chromosomes of *S. spontaneum* (Fig. 4). The results of

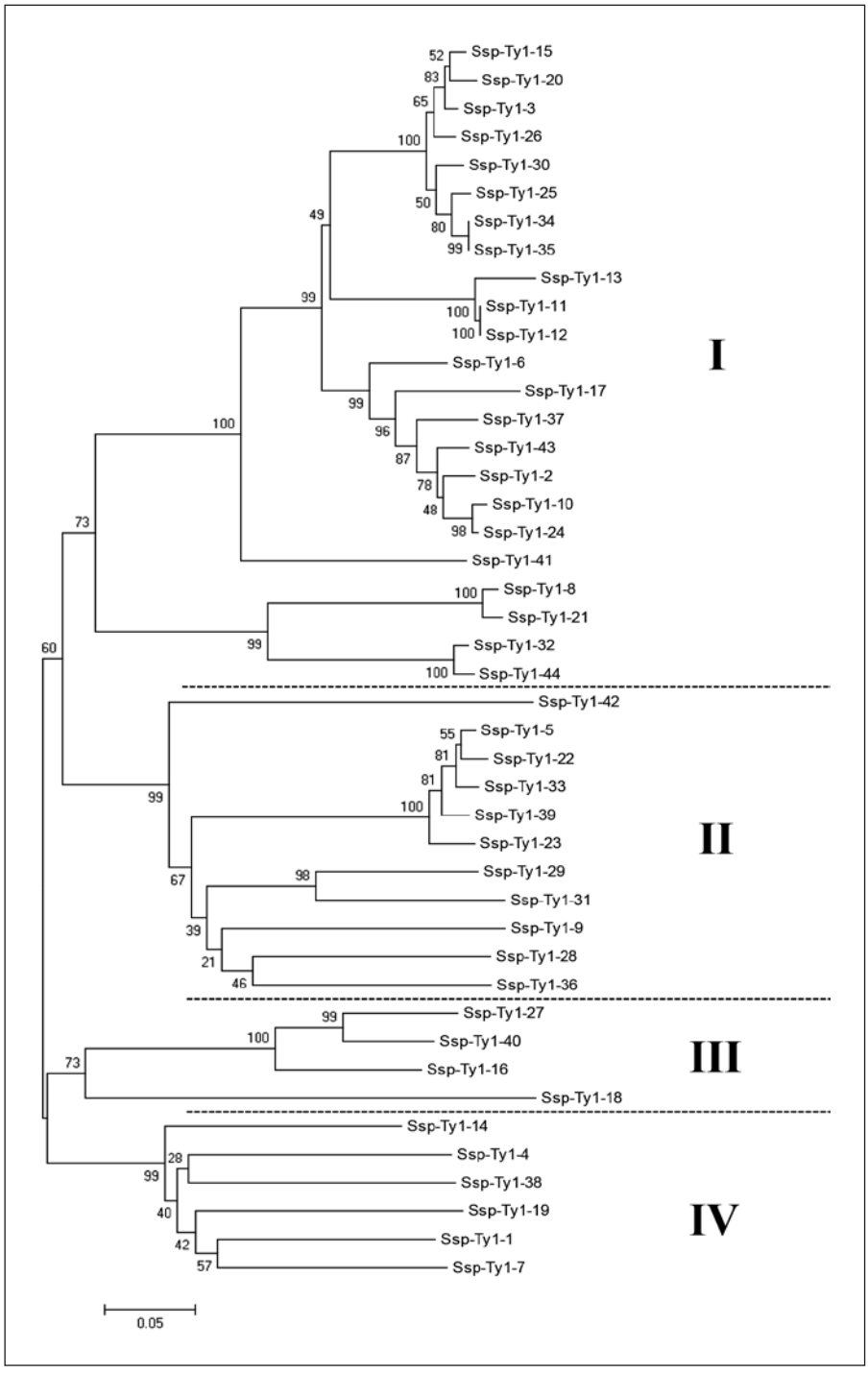


**Fig. 1.** Phylogenetic analysis of Ty1-copia RT nucleotide sequences.

a local blast analysis with AP85–441 genomic data showed that 44 Ty1-copia RT sequences were dispersed and unevenly located in the whole genome (online suppl. Material File 1). In addition, the copy number for Ty1-copia retrotransposons that had a copy number less than 100, such as the Sire/Maximus, Bianca, Tork/TAR, and Ty1-copia like lineages, were not located on all of chromosomes of AP85–441 (Table 1; online suppl. Fig. 4–7). Most of these Ty1-copia retrotransposons were located on the chromosomal arms or telomeres (online suppl. Fig. 4–7). However, the Retrofit/Ale lineage had the highest copy number in the AP85–441 genome and was lo-


Cytogenet Genome Res 2020;160:272–282
DOI: 10.1159/000506222

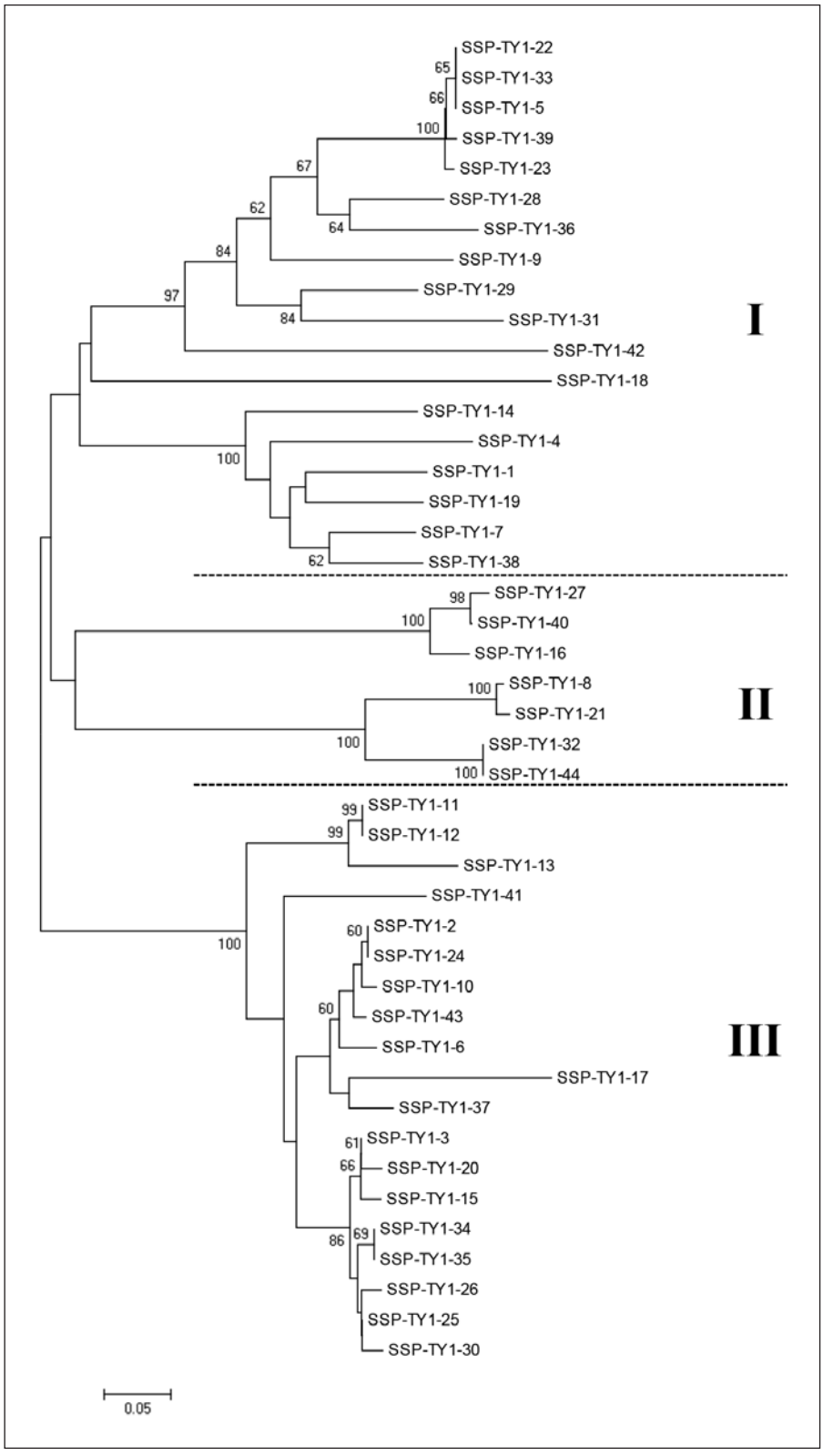


**Fig. 2.** Phylogenetic analysis of Ty1-copia RT amino acid sequences.

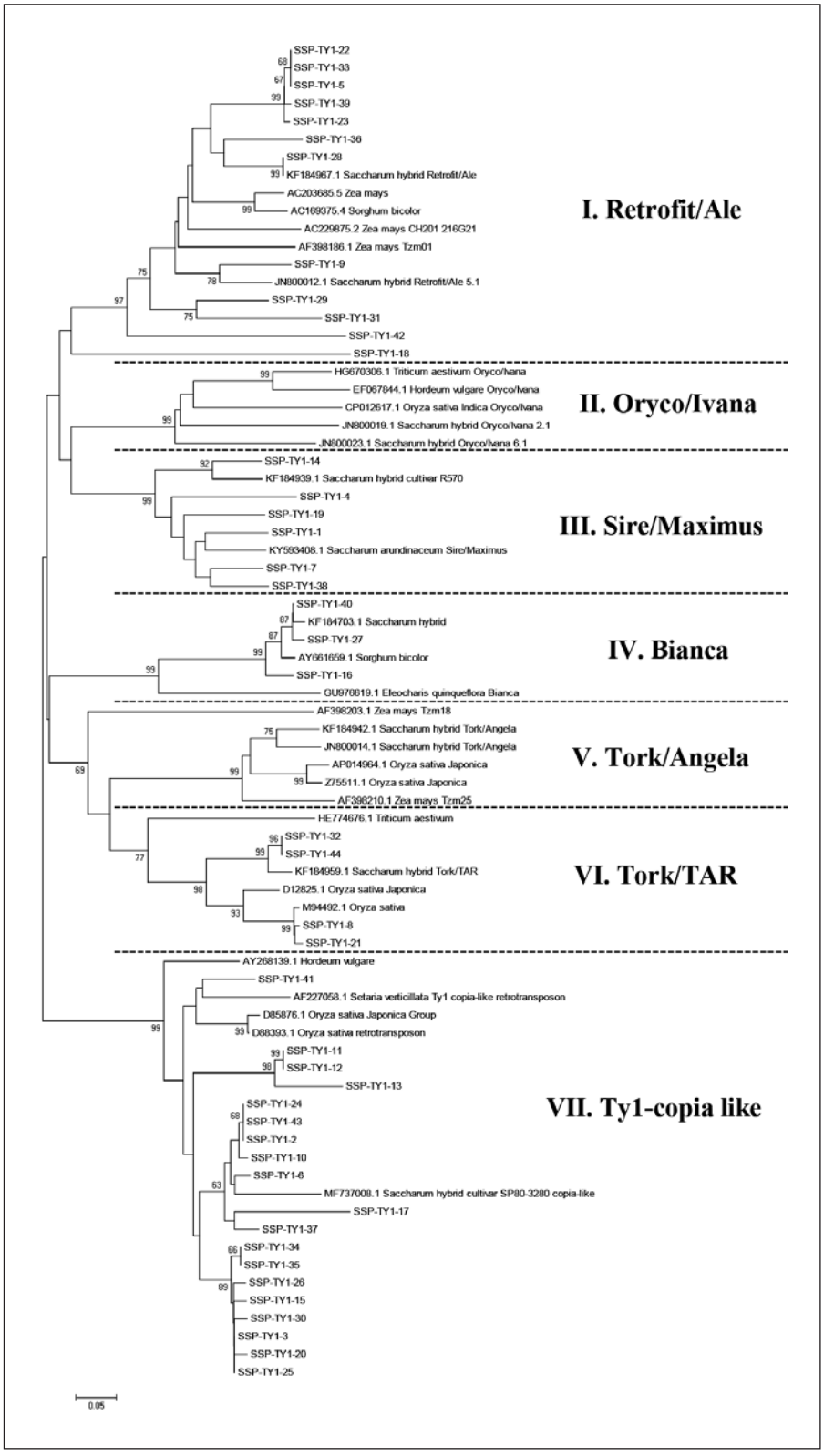


**Fig. 3.** Phylogenetic analysis of Ty1-copia RT amino acid sequences from *S. spontaneum* and other plants.

**Table 1.** The copy number of Ty1-copia RT nucleotide sequences in the AP85-441 genome

| No. | RT sequence name | Copy number | Style | Shown in online suppl. Figure No. |
|---|---|---|---|---|
| 1 | Ssp-Ty1-18 | 4,253 | Retrofit/Ale | 3 |
| 2 | Ssp-Ty1-22 | 1,897 | Retrofit/Ale | |
| 3 | Ssp-Ty1-5 | 1,897 | Retrofit/Ale | |
| 4 | Ssp-Ty1-33 | 1,895 | Retrofit/Ale | |
| 5 | Ssp-Ty1-39 | 1,894 | Retrofit/Ale | |
| 6 | Ssp-Ty1-23 | 1,891 | Retrofit/Ale | |
| 7 | Ssp-Ty1-36 | 66 | Retrofit/Ale | |
| 8 | Ssp-Ty1-29 | 29 | Retrofit/Ale | |
| 9 | Ssp-Ty1-28 | 20 | Retrofit/Ale | |
| 10 | Ssp-Ty1-31 | 9 | Retrofit/Ale | |
| 11 | Ssp-Ty1-42 | 6 | Retrofit/Ale | |
| 12 | Ssp-Ty1-9 | 5 | Retrofit/Ale | |
| 13 | Ssp-Ty1-14 | 15 | Sire/Maximus | 4 |
| 14 | Ssp-Ty1-19 | 10 | Sire/Maximus | |
| 15 | Ssp-Ty1-38 | 4 | Sire/Maximus | |
| 16 | Ssp-Ty1-7 | 4 | Sire/Maximus | |
| 17 | Ssp-Ty1-4 | 3 | Sire/Maximus | |
| 18 | Ssp-Ty1-1 | 3 | Sire/Maximus | |
| 19 | Ssp-Ty1-40 | 11 | Bianca | 5 |
| 20 | Ssp-Ty1-27 | 7 | Bianca | |
| 21 | Ssp-Ty1-16 | 1 | Bianca | |
| 22 | Ssp-Ty1-44 | 78 | Tork/TAR | 6 |
| 23 | Ssp-Ty1-32 | 78 | Tork/TAR | |
| 24 | Ssp-Ty1-21 | 53 | Tork/TAR | |
| 25 | Ssp-Ty1-8 | 53 | Tork/TAR | |
| 26 | Ssp-Ty1-43 | 79 | Ty1-copia like | 7 |
| 27 | Ssp-Ty1-25 | 78 | Ty1-copia like | |
| 28 | Ssp-Ty1-10 | 78 | Ty1-copia like | |
| 29 | Ssp-Ty1-24 | 77 | Ty1-copia like | |
| 30 | Ssp-Ty1-37 | 74 | Ty1-copia like | |
| 31 | Ssp-Ty1-2 | 70 | Ty1-copia like | |
| 32 | Ssp-Ty1-6 | 65 | Ty1-copia like | |
| 33 | Ssp-Ty1-17 | 55 | Ty1-copia like | |
| 34 | Ssp-Ty1-35 | 35 | Ty1-copia like | |
| 35 | Ssp-Ty1-34 | 35 | Ty1-copia like | |
| 36 | Ssp-Ty1-26 | 33 | Ty1-copia like | |
| 37 | Ssp-Ty1-30 | 32 | Ty1-copia like | |
| 38 | Ssp-Ty1-20 | 32 | Ty1-copia like | |
| 39 | Ssp-Ty1-15 | 32 | Ty1-copia like | |
| 40 | Ssp-Ty1-3 | 32 | Ty1-copia like | |
| 41 | Ssp-Ty1-13 | 24 | Ty1-copia like | |
| 42 | Ssp-Ty1-12 | 22 | Ty1-copia like | |
| 43 | Ssp-Ty1-11 | 22 | Ty1-copia like | |
| 44 | Ssp-Ty1-41 | 12 | Ty1-copia like | |

cated on all of the AP85–441 chromosomes (online suppl. Fig. 3). Therefore, the results of FISH were similar to the local blast analysis, indicating Ty1-copia retrotransposon sequences were unevenly distributed along all chromosomes of *S. spontaneum*.

## Discussion

Retrotransposition is a source of genetic diversity in plants and can affect gene expression, location, and organization. Retrotransposons are also important contributors to genome plasticity and evolution [Heslop-Harrison et al., 1997; Kidwell and Lisch, 1997]. To the best of our knowledge, few studies have characterized Ty1-copia retrotransposons that are present in *S. spontaneum*. In this study, 64% of the Ty1-copia RT elements from the *S. spontaneum* genome were defective and incomplete due to the presence of premature stop codons and/or frameshift mutations. *Erianthus arundinaceus* is a species close to the genus *Saccharum* in which approximately 80% of Ty1-copia RT sequences contained premature stop codons and/or frameshifts [Huang et al., 2017]. A large number of these defective elements will lead to the accumulation of mutations in plants, creating lots of heterogeneous RT sequences. Our results showed that Ty1-copia RT sequences in the *S. spontaneum* genome had significant heterogeneity, which is a typical feature of retrotransposons in most plant species [Flavell and Smith, 1992; Pearce et al., 1996; Yanez et al., 1998]. Although the specific mechanisms that contribute to this sequence heterogeneity are not fully known, there are several possible explanations for these observations [Flavell and Smith, 1992]: (a) High numbers of mutations are introduced during reverse transcription by RT, which lacks a proofreading function; (b) RT-mediated cDNA synthesis from RNA is error-prone such that the mutation rate increases with each replication cycle [Pearce et al., 1996; Jiang et al., 2010]; and (c) High heterogeneity of RT sequences influences horizontal and vertical transmission among distantly related species [Woodrow et al., 2012].

In this study, *S. spontaneum* RT sequences were divided into 5 major evolutionary lineages (Retrofit/Ale, Sire/Maximus, Bianca, Tork/TAR, and Ty1-copia like) (Fig. 3). The Retrofit/Ale was the largest Ty1-copia retrotransposon lineage in the *S. spontaneum* genome, followed by Tork/TAR, Sire/Maximus, and Bianca (Fig. 3). This observation is different from soybean, since the Tork/TAR lineage was the largest number of LTR retrotransposon families in soybean [Du et al., 2010]. However, only a sin-

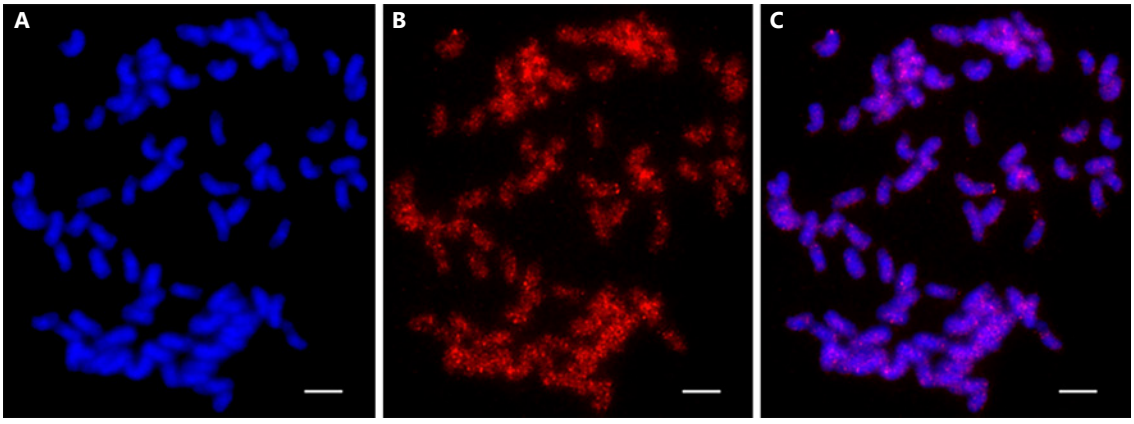


**Fig. 4.** Localization of Ty1-copia RT sequences on the metaphase chromosome of YN82–114. **A** Metaphase chromosomes were stained with DAPI. **B** Disperse signals for Ty1-copia retrotransposons are shown by red fluorescence. **C** Merged images of blue and red signal. The Ty1-copia probes were a PCR-generated genomic mix of RT sequences. Scale bars, 5 μm.

gle intact element for the Tork/TAR lineage was found in the *Arabidopsis* genome [Du et al., 2010]. Hence, differences exist for LTR retrotransposons in different species. Interestingly, we did not detect the Oryco/Ivana and Tork/Angela lineages of Ty1-copia retrotransposons in *S. spontaneum*. One possible explanation for this is that some types of Ty1-copia retrotransposons were lost during the process of evolution. For instance, the Bianca lineage was not found in sugarcane (accession R570) [Domingues et al., 2012] and the Oryco/Ivana and Bianca lineages were not detected in the *E. arundinaceus* genome [Huang et al., 2017]. The other possible explanation was due to amplification bias of degenerate primers used for PCR, resulting in inefficient amplification of targeted regions of the Ty1-copia RT sequences from other lineages.

Previous studies indicated that high copy number and sequence heterogeneity are typical features of Ty1-copia group retrotransposons in maize genomes [Meyers et al., 2001]. In our study, the copy number of the Ty1-copia retrotransposons was approximately 38,900 per genome and 16,300 per genome in YN82–114 and AP85–441, respectively (online suppl. Fig. 1, 2), which is markedly higher than the *E. arundinaceus* genome [Huang et al., 2017] and less than the *Secale cereale* genome [Pearce et al., 1997; Francki, 2001]. This result indicated that Ty1-copia retrotransposons contained a substantial proportion of nuclear genomes leading them to play significant roles in the genome evolution of *S. spontaneum*. Furthermore, our dot blot hybridization analysis used to compare the copy number for Ty1-copia retrotransposons among 2 different *S. spontaneum* genotypes with different ploidies and showed that, even in the same species, the copy number for Ty1-copia retrotransposons is positively related with ploidy. The phenomenon raises an interesting question of how high copy numbers for retrotransposons were proliferated within these polyploid genomes, which might shape the genome structures and complexity in polyploid species. Moreover, different lineages of Ty1-copia retrotransposons had different copy numbers in the *S. spontaneum* genome (Table 1). For the first time, we found that the Retrofit/Ale lineage had the highest copy number and that the Sire/Maximus and Bianca lineages had lower copy numbers in the *S. spontaneum* genome. Previous studies indicated that the FISH signal was detected for the Ale and Maximus lineages, and no detectable FISH signal was obtained for the Angela and Ivana lineages in sugarcane (R570), which suggests that the Ale and Maximus lineages have high copy numbers [Domingues et al., 2012]. We inferred that the Retrofit/Ale lineage of Ty1-copia retrotransposons might play a main role in the evolution process of polyploid *S. spontaneum*. Therefore, a biological function study of the Retrofit/Ale lineage of Ty1-copia retrotransposons will be critical for a better understanding in this area. The different lineages of Ty1-copia retrotransposons exhibited different distribution patterns and different copy numbers (online suppl. Fig. 3–6). For example, the Retrofit/Ale lineage was located on all chromosomes of AP85–441, and the Bianca lineage was only located on 9 chromosomes with focus on the chromosomal arm region. This distribution patterns were similar to the results shown for *Saccharum* hybrids [Domingues et al., 2012]. D'Hont et al. [1996] reported that the *Saccharum* hybrids (the modern sugarcane cultivars) possess approxi-

mately 120 chromosomes, with 70–80% derived from *S. officinarum*, 10–20% derived from *S. spontaneum*, and a few chromosomes derived from interspecific recombination. Consequently, we deduced that Ty1-copia retrotransposons had similar distribution patterns in the genome of *S. officinarum* and *S. robustum*, since the *Saccharum* hybrids contained the blood of *S. officinarum*, and *S. robustum* is an ancestor of *S. officinarum* [Grivet et al., 2004].

In conclusion, 44 Ty1-copia RT subclones with a range of heterogeneity were isolated from *S. spontaneum*. Phylogenetic analysis showed that these sequences were divided into 5 distinct lineages (Retrofit/Ale, Sire/Maximus, Bianca, Tork/TAR, and Ty1-copia like). Dot-blot hybridization analysis showed that there were approximately 38,900 and 16,300 copies of Ty1-copia retrotransposons per genome in YN82–113 and AP85–441, respectively. A local blast analysis indicated that there were 15,069 copies of Ty1-copia retrotransposons in the AP85–441 genome, of which the Retrofit/Ale lineage had the highest copy number, followed by Tork/TAR, Sire/Maximus, and Bianca lineages. Moreover, both FISH and the local blast analysis demonstrated that the Ty1-copia retrotransposons were unevenly distributed throughout the chromosomes.


## Acknowledgements

We thank the Sugarcane Research Institute of the Yunnan Academy of Agricultural Sciences for providing the plant materials used in this study. We greatly appreciate Bioscience Editing Solutions for critically reading this paper and providing helpful suggestions.


## Statement of Ethics

The authors have no ethical conflicts to disclose.

## Disclosure Statement

The authors have no conflicts of interest to declare.


## Funding

This work was funded by the National Natural Science Foundation of China (31571730 and 31771863) and supported by the Graduate School of Fujian Agriculture and Forestry University (KFA17168A). This project was also supported by the earmarked fund for Modern Agricultural Technology of China (CARS-170106) and Doctoral Students of Fujian Agriculture and Forestry University Going Abroad to Cooperative Research (324–112110082). These fund institutions did not play a role in study design, data collection or analysis, or preparation of the manuscript.


## Author Contributions

S.Y., K.Z., Z.D., and F.Y. designed the research. K.Z. and K.C. performed the experiments. S.Y., J.W., Y.H., F.Y, and M.Z. analyzed the results. F.Y., K.C. and J.W. contributed materials. K.Z., Z.D., S.Y., X.Z., and M.Z. wrote and revised the manuscript, and all authors read and approved the final manuscript.

## References


Arumuganathan K, Earle ED: Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218 (1991).

Baucom RS, Estill JC, Leebens-Mack J, Bennetzen JL: Natural selection on gene function drives the evolution of LTR retrotransposon families in the rice genome. Genome Res 19:243 (2009).

D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, et al: Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Mol Gen Genet 250:405–413 (1996).

D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC: Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. Genome 41:221–225 (1998).

Domingues DS, Cruz GM, Metcalfe CJ, Nogueira FT, Vicentini R, et al: Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. BMC Genomics 13:137 (2012).

Du J, Tian Z, Hans CS, Laten HM, Cannon SB, et al: 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 (2010).

El Baidouri M, Carpentier MC, Cooke R, Gao D, Lasserre E, et al: Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res 24:831–838 (2014).

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

Flavell AJ: Ty1-copia group retrotransposons and the evolution of retroelements in the eukaryotes. Genetica 86:203–214 (1992).

Flavell AJ, Smith DB: A Ty1-copia group retrotransposon sequence in a vertebrate. Mol Gen Genet 233:322–326 (1992).

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

Flavell AJ, Pearce SR, Heslop-Harrison P, Kumar A: The evolution of Ty1-copia group retrotransposons in eukaryote genomes. Genetica 100:185–195 (1997).

Francki MG: Identification of Bilby, a diverged centromeric Ty1-copia retrotransposon family from cereal rye (*Secale cereale* L.). Genome 44:266–274 (2001).

Grivet L, Daniels C, Glaszmann J-C, D'Hont A: A review of recent molecular genetics evidence for sugarcane evolution and domestication. Ethnobotany Res Appl 2:9–17 (2004).

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

Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Vershinin AV, et al: The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. Genetica 100: 197–204 (1997).
Huang Y, Luo L, Hu X, Yu F, Yang Y, et al: Characterization, genomic organization, abundance, and chromosomal distribution of Ty1-copia retrotransposons in *Erianthus arundinaceus*. Front Plant Sci 8:924 (2017).
Irvine JE: *Saccharum* species as horticultural classes. Theor Appl Genet 98:186–194 (1999).
Jiang B, Wu ZM, Lou QF, Dong W, Zhang WP, et al: Genetic diversity of Ty1-copia retrotransposons in a wild species of Cucumis (*C. hystrix*). SCI Hortic-Amsterdam 127:46–53 (2010).
Kashkush K, Feldman M, Levy AA: Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet 33:102–106 (2003).
Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, et al: Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice 6:4 (2013).
Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, et al: A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. Dev Cell 7: 133–144 (2004).
Kidwell MG, Lisch D: Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA 94:7704–7711 (1997).
Kumar A, Bennetzen JL: Plant retrotransposons. Annu Rev Genet 33:479–532 (1999).
Li XY, Sun HD, Rott PC, Wang JD, Huang MT, et al: Molecular identification and prevalence of *Acidovorax avenae* subsp. avenae causing red stripe of sugarcane in China. Plant Pathol 67: 929–937 (2017).
Liu J, Bai C, Yan L, Chen S, Zhang Y, et al: Nutritional Value Evaluation of *Saccharum spontaneum* L. Germplasm Resources. Agr Sci Tech 15:1981–1986 (2014).
Meyers BC, Tingey SV, Morgante M: Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res 11:1660–1676 (2001).
Park JM, Schneeweiss GM, Weiss-Schneeweiss H: Diversity and evolution of Ty1-copia and Ty3-gypsy retroelements in the non-photosynthetic flowering plants Orobanche and Phelipanche (Orobanchaceae). Gene 387:75–86 (2007).
Pearce SR, Harrison G, Li D, Heslop-Harrison J, Kumar A, et al: The Ty1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 250:305–315 (1996).
Pearce SR, Harrison G, Heslop-Harrison PJ, Flavell AJ, Kumar A: Characterization and genomic organization of Ty1-copia group retrotransposons in rye (*Secale cereale*). Genome 40:617–625 (1997).
Porebski S, Bailey LG, Baum BR: Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Rep 15:8–15 (1997).
SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, et al: Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768 (1996).
Schneider CA, Rasband WS, Eliceiri KW: NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671 (2012).
Sormacheva I, Blinov A: LTR retrotransposons in plants. Russ J Genet 1:540–564 (2011).
Souza GM, Berges H, Bocs S, Casu R, D'Hont A, et al: The sugarcane genome challenge: strategies for sequencing a highly complex genome. Trop Plant Biol 4:145–156 (2011).
Temin HM: Structure, variation and synthesis of retrovirus long terminal repeat. Cell 27:1–3 (1981).
Wicker T, Keller B: Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. Genome Res 17:1072–1081 (2007).
Woodrow P, Ciarmiello LF, Fantaccione S, Annunziata MG, Pontecorvo G, et al: Ty1-copia group retrotransposons and the evolution of retroelements in several angiosperm plants: evidence of horizontal transmission. Bioinformation 8:267–271 (2012).
Wu R: DNA sequence analysis. Annu Rev Biochem 47:607–634 (1978).
Yanez M, Verdugo I, Rodriguez M, Prat S, Ruiz-Lara S: Highly heterogeneous families of Ty1/copia retrotransposons in the *Lycopersicon chilense* genome. Gene 222:223–228 (1998).
Zhang J, Nagai C, Yu Q, Pan Y-B, Ayala-Silva T, et al: Genome size variation in three *Saccharum* species. Euphytica 185:511–519 (2012).
Zhang J, Zhang X, Tang H, Zhang Q, Hua X, et al: Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. Nat Genet 50:1565–1573 (2018).