Isolation and characterization of Ty1-*copia* retrotransposons in *Saccharum officinarum*

Ke Chen [1], **Fan Yu** [1], **Yongji Huang** [1], **Xianglin Wu** [1], **Shan Yang** [1], **Xueting Li** [1], **Kai Zeng** [1], **Xuguang Hu** [1], **Fei Huang** [1], **Zuhu Deng** [Corresp., 1, 2, 3], **Muqing Zhang** [3]

[1] National Engineering Research Center for Sugarcane, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

[2] Key Lab of Sugarcane Biology and Genetic Breeding, Ministry of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

[3] State Key Laboratory for protection and utilization of subtropical agro-bioresources Guangxi University, Nanning 530004, China

Corresponding Author: Zuhu Deng
Email address: dengzuhu@163.com

Background. *Saccharum officinarum* is the most significant resource for sugar and high-yield genes in sugarcane breeding programs. However, the unknown information of evolution and genome organization remain largely in the sugarcane, which has limited progress in sugarcane breeding. Retrotransposons occupy a large proportion of the plant genome; therefore, characterization of Ty1-*copia* retrotransposons will improve understanding of the evolution and organization of plant genomes.

Methods. The present study isolated conserved domains of Ty1-*copia* retrotransposon-encoded reverse transcriptase genes from *S. officinarum* to characterize their phylogenetic diversity, genomic abundance, and chromosomal distribution.

Results. In total, 42 Ty1-*copia* reverse transcriptase sequences with 35-100% similarity and high levels of heterogeneity were obtained. Of them, 11 (26%) were disrupted by stop codons and/or frameshift mutations. Phylogenetic analysis revealed these sequences could be split into four distinct evolutionary lineages (Tork/TAR, Tork/Angela, Sire/Maximus, and Retrofit/Ale). Dot blot analysis showed that Ty1-*copia* retrotransposons represent a significant portion of the *S. officinarum* genome, with copy numbers as high as 1.7×10^5. Fluorescence *in situ* hybridization revealed that Ty1-*copia* retrotransposons were dispersed within heterochromatic regions among all *S. officinarum* chromosomes, with around 30 obvious signals clustering in terminal regions. However, Ty1-*copia* retrotransposons were not found in nucleolar organizing regions of 45S rDNA.

Discussion. These results serve to enhance our understanding of the chromosomal distribution and evolution of the *S. officinarum* genome as well as promote possible utilization of retrotransposons in sugarcane breeding programs.

Isolation and Characterization of Ty1-*copia* Retrotransposons in *Saccharum officinarum*

Ke Chen[1], Fan Yu[1], Yongji Huang[1],Xianglin Wu[1], Shan Yang[1], Xueting Li[1],Kai Zeng[1], Xuguang Hu[1], Fei Huang[1], Zuhu Deng[1,2,3#], and Muqing Zhang[3]

[1]National Engineering Research Center for Sugarcane, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

[2]Key Lab of Sugarcane Biology and Genetic Breeding, Ministry of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, China

[3]State Key Laboratory for protection and utilization of subtropical agro-bioresources Guangxi University, Nanning, 530004,China

Corresponding author:

Zuhu Deng[1,2,3#]

[#]Corresponding author: dengzuhu@163.com

Abstract

Background. *Saccharum officinarum* is the most significant resource for sugar and high-yield genes in sugarcane breeding programs. However, the unknown information of evolution and genome organization remain largely in the sugarcane, which has limited progress in sugarcane breeding. Retrotransposons occupy a large proportion of the plant genome; therefore, characterization of Ty1-*copia* retrotransposons will improve understanding of the evolution and organization of plant genomes.

Methods. The present study isolated conserved domains of Ty1-*copia* retrotransposon-encoded reverse transcriptase genes from *S. officinarum* to characterize their phylogenetic diversity, genomic abundance, and chromosomal distribution.

Results. In total, 42 Ty1-*copia* reverse transcriptase sequences with 35-100% similarity and high levels of heterogeneity were obtained. Of them, 11 (26%) were disrupted by stop codons and/or frameshift mutations. Phylogenetic analysis revealed these sequences could be split into four distinct evolutionary lineages (Tork/TAR, Tork/Angela, Sire/Maximus, and Retrofit/Ale). Dot blot analysis showed that Ty1-*copia* retrotransposons represent a significant portion of the *S. officinarum* genome, with copy numbers as high as 1.7×10^5. Fluorescence *in situ* hybridization revealed that Ty1-*copia* retrotransposons were dispersed within heterochromatic regions among all *S. officinarum* chromosomes, with around 30 obvious signals clustering in terminal regions. However, Ty1-*copia* retrotransposons were not found in nucleolar organizing regions of 45S rDNA.

Discussion. These results serve to enhance our understanding of the chromosomal distribution and evolution of the *S. officinarum* genome as well as promote possible utilization of retrotransposons in sugarcane breeding programs.

Keywords: *Saccharum officinarum*, Ty1-*copia* retrotransposons, phylogenetic diversity, chromosomal organization, genome

Introduction

Sugarcane (*Saccharum* spp.) is an important agricultural crop that accounts for 80% of the world's raw sugar production and is being increasingly utilized as a source of renewable energy (D'Hont et al. 2008; Lam et al. 2009). *Saccharum officinarum* is a noble cane with large stems, wide leaves, and a high sugar content (Irvine 1999). Hence, *S. officinarum* is considered the most significant germplasm resource for sugar and high-yield genes in sugarcane breeding programs. Modern sugarcane cultivars are aneuploid or polyploid, with highly complex genomes that are 80-90% *S. officinarum* and 10-20% *S. spontaneum* in origin (D'Hont 2005; Grivet & Arruda 2002). *S. officinarum* (2n = 80) is polyploid, with a highly complex genome; its genomic size is reportedly about 7.88 Gb, with 8.42 pg DNA content (Oliveira et al. 2015; Zhang et al. 2012). Due to its complicated genome contains a large number of high copy repetitive nucleotide sequences and polyploid genomes, sugarcane genome sequencing is an enormous challenge (Okura et al. 2016). The lack of genomic sequence data resulting in research of the evolution and sequence organization in the *S. officinarum* genome is weak.

Transposable elements are mobile, plastic genetic elements that are ubiquitous and abundant in all eukaryotes. These elements are able to replicate substantial fractions of host genomes (Bowen & Jordan 2002; Feschotte et al. 2002). Transposable elements are divided into two main classes (DNA transposons and retrotransposons) based on their DNA sequence, structural similarity, and/or transposition mechanism. Retrotransposons are particularly abundant in the plant kingdom. They are transposed via RNA intermediates and amplified, eventually inserting into target gene fragments in the host genome (Wicker et al. 2007). Retrotransposons are distributed on chromosomes with a high copy number and heterogeneity. In plants, they are widely distributed, mobile genetic elements that represent a considerable portion of the dispersed repeats in the genome. Meanwhile, retrotransposons has a profound influence on changes of plant genome size and structure. (Brookfield 2005; Domingues et al. 2012; Tsukahara et al. 2009; Vitte & Panaud 2005; Vonholdt et al. 2012; Zedek et al. 2010). In recent years, genomic sequence of various higher plants has demonstrated that retrotransposons are an important

component of the plant genome (Choulet et al. 2014; Garsmeur et al. 2018; Initiative 2000; Mayer et al. 2014; Paterson et al. 2009). It has been suggested that genomic expansion is largely attributed to retrotransposon amplification. At present, retrotransposons are a main focus in plant structural and evolutionary genomic research.

Depending on whether they have a long terminal repeat (LTR), retrotransposons can be further classified into two distinct types: LTR and non-LTR retrotransposons. LTR retrotransposons play especially important roles in the plant kingdom (Kumar & Bennetzen 2000). LTR retrotransposons are further subdivided into five groups based on their domain structure in the polyprotein region. In particular, Ty1-*copia* retrotransposons exist in most higher plants, including algae, bryophytes, gymnosperms, and angiosperms (Wicker & Keller 2007). The highly conserved reverse transcriptase (RT) domains of Ty1-*copia* retrotransposons can be used to study evolutionary dynamics and phylogenetic relationships, both within and among related groups of taxa (Dixit et al. 2006; Flavell et al. 1992; Goodwin & Poulter 2002; Heslop-Harrison et al. 1997; Khaliq et al. 2012; Kumar et al. 1997; Lee et al. 2013; Ma et al. 2008; Santini et al. 2002). While study of Ty1-*copia* retrotransposon RTs have provided insight into to the genomic organization and evolution of many plants, very little information is known about these retrotransposons in *S. officinarum*.

In the present work, Ty1-*copia* RT sequences were isolated from *S. officinarum* and their heterogeneity, phylogenetic relationships, abundance, and chromosomal distribution were investigated for the first time. The results will provide greater insight into the genomic structure and evolution of *S. officinarum* that can be utilized in sugarcane breeding programs.

Materials & Methods

Plants and plant DNA

Badila cultivar *S. officinarum* plants were grown in the Fujian Agriculture and Forestry University (Fuzhou, China) greenhouse under field conditions. Total genomic DNA was extracted from fresh young *S. officinarum* leaves following the cetyltrimethyl ammonium

bromide method (Doyle 1987). DNA quality was evaluated by 1% agarose gel electrophoresis.

Polymerase chain reaction (PCR) and cloning

To amplify highly conserved partial sequences of Ty1-*copia* retrotransposon RT genes, degenerate primers (forward: 5'-ACNGCNTTYYTNCAYGG-3'; reverse: 5'-ARCATRTCRTCNACRTA-3') were used according to a previous report (Kumar et al. 1997). PCR amplification was carried out in a BIO-RAD T100™ Thermal Cycler (Bio-Rad, Hercules, CA, USA). In a total reaction volume of 25 μL, containing 30-50 ng of DNA, 10 nM of each of the primers forward and reverse, 0.2 mmol/l of dNTP, 2.5 mmol/l of MgCl2 and 1 U of Taq polymerase (Takara, Tokyo, Japan). PCR conditions included an initial denaturation at 94 ℃ for 4 min, followed by 35 cycles of 94 ℃ for 30 s, annealing at 45 ℃ for 45 s, extension at 72 ℃ for 20 s, with final elongation step at 72 ℃ for 5 min. PCR products were purified using a Qiaquick Gel Extraction Kit (Qiagen, Germany), cloned in a pMD19-T vector (Takara, Tokyo, Japan), and then transformed into the DH5a strain of Eschericia coli. Positive colonies were further confirmed by PCR, then all results were sequenced by Beijing Genomics Institute Co., Ltd. (Shenzhen, China). In total, 42 Ty1-*copia* RT sequences were deposited in the GenBank database (accession no. MH603333-MH603374) and designated as SoffTy1-*copia*-1 to -42.

Phylogenetic analysis

Ty1-*copia* retrotransposon RT sequences from *S. officinarum* were assembled using DNAMAN (Lynnon BioSoft). BLASTN from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) databases was used to survey their homology to previously characterized plant Ty1-*copia* retroelement lineages, such as graminaceous species. Ty1-copia RT sequences from S. officinarum were translated to amino acid sequences by the Transeq tool in EMBOSS package2 (https://www.ebi.ac.uk/Tools/st/emboss_transeq/). Multiple nucleotide and amino acid sequences were aligned using BioEdit software (Alzohairy 2011). Ty1-*copia* RT amino acid sequences were aligned against other similar sequences and frameshift

mutations were detected using ERRWISE (http://coot.embl.de/ERR_WISE/). ERRWISE can detect interruptions in open reading frames (ORFs), gaps were introduced to retain ORFs in multiple sequences. Multiple sequence alignment of Ty1-*copia* RT sequences from S. *officinarum* and Nucleotide sequences of other species RT sequences were used to create a comparative phylogenetic dendrogram using the neighbor joining method in MEGA 7.0 (Kumar et al. 2016). All sequences was aligned by MUSCLE(Edgar 2004), phylogenetic analysis of the aligned sequences based on p-distance and supported with 1000 bootstrap replicates, the Pairwise deletion of missing data (gaps) was used to compute the distance matrices.

Dot blot hybridization

All purified plasmids clones and PCR products Ty1-*copia* RT sequences in *S. officinarum* was estimated by quantitative dot blot hybridization protocol followed the procedure of (Huang et al. 2017). 42 clones were diluted to a final concentration of 20 ng/μL, serial dilutions of *S. officinarum* genomic DNA (500, 400, 250, 200, 125, and 100 ng) and PCR products Ty1-*copia* RT sequences (1.8, 1.2, 0.9, 0.6, 0.45, and 0.3 ng). The copy number per genome was estimated by determining the hybridization intensity using ImageJ software (Schneider et al. 2012).

Chromosome preparation and fluorescence in situ hybridization (FISH)

For mitotic chromosome accumulation, the fresh root were harvested and treated with 2 mM 8-hydroxyquinoline at 30 °C for 18 h, 2.5 μM Amiprophos-methyl at 30 °C for 2h, and ddH_2O at 30 °C for 6h and then fixed in 3:1 (v/v) ethanol:acetic acid. The meristematic cells of root tips were digested in an enzyme solution containing 3% Onozuka R10 cellulose, 0.5% pectolyase Y-23, and 1% pectinase at 37 ℃ for 90 min. Then the meristematic cells of no wall were squashed on a clean slide, quality of cells was checked under a phase contrast microscope and stored in at -20 °C until use. FISH according to the procedures were described by (Jiang et al. 1995) with minor modifications. The PCR products of Ty1-*copia* RT domain were labeled with digoxigenin-11-dUTP (DIG). 45S ribosomal DNA (rDNA) were labeled with biotin-16-dUTP

(BIO). FISH signals were visualized by An AxioScope A1 Imager fluorescent microscope (Carl Zeiss, Gottingen, Germany). Photographs were captured and analyzed using Axio imaging software, Photoshop CS6 software was used to obtain optimal experimental images.

Results

Cloning and Analysis of Ty1-*copia* retrotransposon RT sequences in *S. officinarum*

Degenerate primers have been widely used to amplify fragment of theTy1-*copia* RT domain in many plant species (Flavell et al. 1992; Kumar et al. 1997). We designed a pair of degenerate primers to the highly conserved sequence were used to amplify the Ty1-*copia* RT sequences. After purification and cloning, a total of 42 independent clones were randomly selected for sequencing. Sequences were found to have lengths around 260 bp. Ty1-*copia* RT sequences were AT-rich, with 59% mean A/T content (Table 1). All fragments were the same size, initially demonstrating that Ty1-*copia* RT sequences are abundant in *S. officinarum* genomes.

BLAST comparisons of the obtained 42 Ty1-*copia* RT sequences with those in GenBank revealed their homology with known Ty1-*copia* RT sequences from other plants. High nucleotide sequence similarity (range, 39-100%; mean, 77%) was observed among isolated RT sequences (Table 1). All Ty1-copia RT sequences were translated into amino acid sequences with sequences exist conserved domain motifs upstream TAFLHG, central SLYGLKQ, and downstream YVDDM. Analysis of the amino acid homology of the amplified RT fragments also revealed higher homology (mean, 66%) among Ty1-*copia* RT fragments (Fig 1).

Phylogenetic analysis of Ty1-*copia* retrotransposon RT sequences

To study the relationships among the 42 obtained Ty1-*copia* RT sequences in the *S. officinarum* genome, a neighbor joining tree was constructed. After translation and further alignment of the amino acid sequence to observe divergence, Ty1-*copia* RT sequences were edited for frameshift mutations within their coding regions, and gaps were introduced to retain ORFs. Eleven of the 42 sequences (26%) contained premature stop codons and/or

insertions/deletions, disrupting the ORF (Table 1). Primer binding were conservative regions, which manually edited to remove fragments for phylogenetic analysis of Ty1-*copia* RT sequences. In addition, relationships among the isolated *S. officinarum* Ty1-*copia* RT sequences were compared with related Ty1-*copia* retrotransposons from other graminaceous species (*Aegilops, Hordeum, Oryza, Saccharum, Sorghum, Triticum*, and *Zea*). Most of the Ty1-*copia* RT sequences from 31 species were obtained from GenBank and included in the phylogenetic analysis (Fig 1) and the amino acid sequences divided into four evolutionary lineages (Fig 1). The constructed tree showed the largest lineage to be Tork/TAR (62%) in *S. officinarum*, followed by Sire/Maximus (24%), Tork/Angela (7%), and Retrofit/Ale (7%) (Table 1). Tork/TAR sequences showed high similarity (95%) (Table 1). Moreover, the neighbor joining algorithm showed very high sequence heterogeneity between the predicted amino acids in *S. officinarum*.

Copy number of Ty1-*copia* retrotransposon RT sequences in *S. officinarum*

Reverse dot blot hybridization analysis was performed to examine the relative abundance of isolated Ty1-*copia* RT sequence clones from *S. officinarum*. Hybridization of all purified plasmids containing Ty1-*copia* RT domain clones to *S. officinarum* genomic DNA revealed no obvious signals, indicating single Ty1-*copia* RT sequences have low copy numbers in the *S. officinarum* genome. Total Ty1-*copia* sequences, on the other hand, were found to have very strong hybridization signals (Fig 2), confirming that total Ty1-*copia* retrotransposons are quite abundant in the *S. officinarum* genome. To estimate the total copy number of these Ty1-*copia* retrotransposons in the *S. officinarum* genome, quantitative dot blot assay using serial dilutions of total Ty1-*copia* RT sequences from PCR-amplified clones as probes and genomic DNA from *S. officinarum* were used. The results indicated the total copy number was approximately 1.7×10^5 per genome (Fig 2).

Chromosomal distribution of Ty1-*copia* retrotransposon RT sequences in *S. officinarum*

The chromosomal distribution of Ty1-*copia* elements was determined using FISH. Ty1-*copia* element hybridization signals have been shown to be unevenly distributed on metaphase chromosomes and interphase nuclei. (Brookfield 2005; Heslop-Harrison et al. 1997; Khaliq et al. 2012; Vitte & Panaud 2005; Zedek et al. 2010). In the genome of *S. officinarum*, approximately 30 telomere regions of chromosomes were found to have very strong hybridization signals (Fig 3b). Moreover, FISH signals in interphase nuclei mostly colocalized with euchromatic regions, while heterochromatic regions of interphase nuclei showed fewer and fainter hybridization signals (Fig 3f).

Further analysis of Ty1-*copia* retrotransposon distribution patterns in the *S. officinarum* genome and selection of potential chromosomal markers was done using biotin-labeled 45S rDNA as a probe. FISH results showed no colocalization between 45S rDNA and Ty1-*copia* element hybridization signals in metaphase chromosomes of *S. officinarum* (Fig 4). These observations suggest that Ty1-*copia* retrotransposons are dispersed within euchromatic regions and enriched in about 30 telomere regions of chromosomes but do not exist in the nucleolus organizer region 45S rDNA.

Discussion

Retrotransposons constitute a significant portion of most plant genomes and are an important source of genetic diversity, organization, and evolution (Goodwin & Poulter 2002; Kumar & Bennetzen 2000; Santini et al. 2002; Vitte & Panaud 2005; Vonholdt et al. 2012). The present study is the first attempt to survey the heterogeneity, phylogenetic relationships, abundance, and chromosomal distribution of Ty1-*copia* retrotransposon RT sequences in *S. officinarum*, an important germplasm resource for sugar and high-yield genes in sugarcane breeding. Although study of Ty1-*copia* retrotransposon RTs have provided insight into to the genomic organization and evolution of many plants, very little information is known about these retrotransposons in *S. officinarum*.

Ty1-*copia* retrotransposons have been widely researched in monocotyledonous to dicotyledonous plant taxa via PCR (Flavell et al. 1992; Heslop-Harrison et al. 1997; Jiang et al. 2010; Khaliq et al. 2012; Kumar et al. 1997; Lee et al. 2013) and other molecular cytology tools (Huang et al. 2017; Khaliq et al. 2012; Kolano et al. 2013; Kumar et al. 1997; Santini et al. 2002). Herein, the conserved domains of Ty1-*copia* sequences were amplified using a pair of carefully selected degenerate primers and three independent rounds of amplification and cloning to help alleviate bias from methodological effects (Park et al. 2007). This approach increased our chances of obtaining a more representative genomic sample of the analyzed fragments and allowed for broader sampling of their diversity in the *S. officinarum* genome. We obtained 42 sequences with variable homogeneity ranging from 39% to 100%, indicating that Ty1-*copia* retrotransposons in *S. officinarum* are highly heterogeneous. Similar results have also been reported in other species (Dixit et al. 2006; Jiang et al. 2010; Kolano et al. 2013; Ma et al. 2008; Sun et al. 2013), suggesting that heterogeneity is a natural consequence of the presence of retrotransposons. The high heterogeneity of Ty1-*copia* retrotransposons is likely due to a number of factors. First, retrotransposition entails a high mutation rate, which increases the mutation frequency with every replication cycle (Steinhauer & Holland 1986). Vertical transmission of retrotransposons within plant lineages and horizontal transmission between distantly species have played roles in the evolution of retrotransposons in plants (Flavell et al. 1992; Kumar & Bennetzen 2000). In the genome expansion period, both illegitimate recombination and unequal homologous recombination are driving force with the major retrotransposons are efficient removed from the host genome (Devos et al. 2002; Vonholdt et al. 2012). Furthermore, the mean A/T content in Ty1-*copia* retrotransposons is 59% in *S. officinarum*; being AT-rich can increase DNA flexibility, and genomes are more prone to mutations when self-amplifying.

LTR retrotransposons in most plants define six major common evolutionary Ty1-*copia* lineages: Tork/TAR, Tork/Angela, Sire/Maximus, Oryco/Ivana, Retrofit/Ale, and Bianca (Domingues et al. 2012; Llorens et al. 2009; Wicker et al. 2007). The present phylogenetic analysis revealed four lineages within the *S. officinarum* genome (Tork/TAR, Tork/Angela,

Sire/Maximus, and Oryco/Ivana (Fig 1). Although Oryco/Ivana and Bianca lineages were not found in newly amplified Ty1-*copia* RT fragments of *S. officinarum* here, they have been shown to be minor components of Ty1-*copia* retrotransposons in other sugarcane species (Grivet & Arruda 2002). This suggests demonstrates horizontal transmission and vertical transmission of Ty1-copia retrotransposons in the S. *officinarum* genome. Tork/TAR contained the largest Ty1-*copia* retrotransposons in *S. officinarum*, suggesting this lineage is the most abundant in *S. officinarum*, followed by Retrofit/Ale, Tork/Angela, and Sire/Maximus. Tork/TAR has the largest number of LTR retrotransposon families in soybean but may be facing extinction in *Arabidopsis*, and Sire/Maximus and Bianca are the only Ty1-*copia* lineages not found in quinoa (Kolano et al. 2013). The number of Ty1-*copia* retrotransposons within each lineage may vary tremendously by species, some lineages are lost on the evolutionary process and numbers gradually decrease to extinction in genomes. Fortunately, the next generation sequence can greatly enhance this retrotransposon of research when the evolutionary process gradually eliminates certain lineages.

Dot blot hybridization allowed relatively accurate measurement of Ty1-*copia* retrotransposon copy number in the genome. It is now well-established that retrotransposons are key drivers in the evolution of plant genome size (Devos et al. 2002; Tsukahara et al. 2009). Plant genomes either undergo downsizing by elimination of transposed copies or increase through bursts of retrotransposition. In different host genomes, the copy number of retrotransposons can vary from hundreds of elements to over 1 million per genome. For example, *Arabidopsis* and rice have relatively small genomes and therefore, lower retrotransposon copy numbers, while plants with medium to large genomes have high retrotransposon copy numbers (Paterson et al. 2009; Pereira 2004; Zhang & Gao 2017). The copy number of Ty1-*copia* RT sequences in *S. officinarum* was found to be approximately 1.7×10^5 per genome in the present study, implying that retrotransposons occupy a significant position in the genomic evolution of *S. officinarum*. Although Ty1-*copia* retrotransposons are abundant in *S. officinarum,* we found that about 26% were defective sequences, while the remaining 31 fragments (74%) were integrity

297 ORFs (Table 1). The presence of mutations, such as frameshift and those introducing stop
298 codons, in Ty1-*copia* retrotransposon coding regions results in their being unable to
299 autonomously transpose in host genomes. (Ma et al. 2008; Navarro-Quezada & Schoen 2002).
300 The chromosomal distribution of Ty1-*copia* retrotransposons has been studied in many plant
301 species in metaphase and interphase chromosomes using Ty1-*copia* RT sequences as FISH
302 probes (Huang et al. 2017; Jiang et al. 2010; Khaliq et al. 2012; Kolano et al. 2013; Lee et al.
303 2013). Ty1-*copia* retrotransposons are most commonly found to be distributed along the entire
304 length of chromosomes, with a possible exception of some chromosomal landmarks, such as
305 NORs, centromeres, and telomeres (Kumar et al. 1997). Ty1-*copia* elements are also reported to
306 be widely dispersed over all chromosomes and clusters in heterochromatin regions (Friesen et al.
307 2001; Jiang et al. 2010), with some on both ends of the chromosome (Huang et al. 2017). In the
308 present study, Ty1-*copia* retrotransposons in *S. officinarum* were found to be dispersed in
309 heterochromatic regions, with strong signals at the terminal regions of most chromosomes (Fig
310 3). However, Ty1-*copia* retrotransposons were not found in nucleolar organizing regions of 45S
311 rDNA (Fig 4). This indicates that Ty1-*copia* retrotransposons have higher copy numbers in distal
312 chromosome regions. This unique distribution pattern could be used as potential chromosomal
313 marker, providing important information for dissecting the genomic structure of *S. officinarum*.
314

315 **Conclusions**

316 The present study is the first to extensively study the phylogenetic diversity, genomic
317 abundance, and chromosomal distribution of Ty1-*copia* retrotransposon RT sequences in *S.*
318 *officinarum*. A total of 42 Ty1-*copia* RT sequences were isolated and characterized, with high
319 levels of heterogeneity and four evolutionary lineages (Tork/TAR, Tork/Angela, Sire/Maximus,
320 and Retrofit/Ale). There copy numbers were found to be as high as 1.7×10^5 in the genome of *S.*
321 *officinarum*, and 26% of sequences were disrupted by stop codons and/or frameshift mutations.
322 Ty1-*copia* retrotransposons were dispersed throughout heterochromatic regions of chromosomes,
323 with around 30 obvious signals clustering in terminal regions. However, Ty1-*copia*

retrotransposons were not found in nucleolar organizing regions of 45S rDNA. These results provide key chromosomal organization and evolutionary information on Ty1-*copia* retrotransposons in the *S. officinarum* genome that will be important for further studies on *Saccharum* spp. and aid the advancement of sugarcane breeding programs.

References

Alzohairy AM. 2011. BioEdit: An important software for molecular biology. *Gerf Bulletin of Biosciences* 2:60-61. 2 (1) :60-61

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

Brookfield JF. 2005. Evolutionary forces generating sequence homogeneity and heterogeneity within retrotransposon families. *Cytogenet Genome Res* 110:383-391. 10.1159/000084970

Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Le Gouis J, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Safar J, Dolezel J, Rogers J, Vandepoele K, Aury JM, Mayer K, Berges H, Quesneville H, Wincker P, and Feuillet C. 2014. Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345:1249721. 10.1126/science.1249721

D'Hont A. 2005. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenet Genome Res* 109:27-33. 10.1159/000082378

D'Hont A, Souza GM, Menossi M, Vincentz M, Van-Sluys M-A, Glaszmann JC, and Ulian E. 2008. *Sugarcane: A major source of sweetness, alcohol, and bio-energy*: Springer, 233 Spring Street, New York, Ny 10013, United States.

Devos KM, Brown JK, and Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* 12:1075-1079. 10.1101/gr.132102

Dixit A, Ma KH, Yu JW, Cho EG, and Park YJ. 2006. Reverse transcriptase domain sequences from Mungbean (*Vigna radiata*) LTR retrotransposons: sequence characterization and phylogenetic analysis. *Plant Cell Rep* 25:100-111. 10.1007/s00299-005-0008-2

Domingues DS, Cruz GM, Metcalfe CJ, Nogueira FT, Vicentini R, Alves Cde S, and Van Sluys MA. 2012. Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. *BMC Genomics* 13:137. 10.1186/1471-2164-13-137

Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11-15.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797. 10.1093/nar/gkh340

Feschotte C, Jiang N, and Wessler SR. 2002. Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3:329-341. 10.1038/nrg793

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

Friesen N, Brandes A, and Heslop-Harrison JS. 2001. Diversity, origin, and distribution of retrotransposons (*gypsy*

and *copia*) in conifers. *Mol Biol Evol* 18:1176-1188. 10.1093/oxfordjournals.molbev.a003905

Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, Jenkins J, Martin G, Charron C, Hervouet C, Costet L, Yahiaoui N, Healey A, Sims D, Cherukuri Y, Sreedasyam A, Kilian A, Chan A, Van Sluys MA, Swaminathan K, Town C, Berges H, Simmons B, Glaszmann JC, van der Vossen E, Henry R, Schmutz J, and D'Hont A. 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat Commun* 9:2638. 10.1038/s41467-018-05051-5

Goodwin TJ, and Poulter RT. 2002. A group of deuterostome Ty3/ *gypsy*-like retrotransposons with Ty1/ *copia*-like pol-domain orders. *Mol Genet Genomics* 267:481-491. 10.1007/s00438-002-0679-0

Grivet L, and Arruda P. 2002. Sugarcane genomics: depicting the complex genome of an important tropical crop. *Curr Opin Plant Biol* 5:122-127.

Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Vershinin AV, Alkhimova EG, Kamm A, Doudrick RL, Schwarzacher T, Katsiotis A, Kubis S, Kumar A, Pearce SR, Flavell AJ, and Harrison GE. 1997. The chromosomal distributions of Ty1-*copia* group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* 100:197-204.

Huang Y, Luo L, Hu X, Yu F, Yang Y, Deng Z, Wu J, Chen R, and Zhang M. 2017. Characterization, Genomic Organization, Abundance, and Chromosomal Distribution of Ty1-*copia* Retrotransposons in *Erianthus arundinaceus*. *Front Plant Sci* 8:924. 10.3389/fpls.2017.00924

Initiative AG. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815. 10.1038/35048692

Irvine JE. 1999. *Saccharum* species as horticultural classes. *Theoretical and Applied Genetics* 98:186-194. 10.1007/s001220051057

Jiang BA, Wu ZM, Lou QF, Wang D, Zhang WP, and Chen JF. 2010. Genetic diversity of Ty1-*copia* retrotransposons in a wild species of *Cucumis* (*C hystrix*). *Scientia Horticulturae* 127:46-53. 10.1016/j.scienta.2010.09.013

Jiang J, Gill BS, Wang GL, Ronald PC, and Ward DC. 1995. Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proc Natl Acad Sci U S A* 92:4487-4491.

Khaliq I, Khan MA, and Pearce S. 2012. Ty1-*Copia* retrotransposons are heterogeneous, extremely high copy number and are major players in the genome organization and evolution of *Agave tequilana*. *Genetic Resources and Crop Evolution* 59:575-587. 10.1007/s10722-011-9705-6

Kolano B, Bednara E, and Weiss-Schneeweiss H. 2013. Isolation and characterization of reverse transcriptase fragments of LTR retrotransposons from the genome of *Chenopodium quinoa* (Amaranthaceae). *Plant Cell Rep* 32:1575-1588. 10.1007/s00299-013-1468-4

Kumar A, and Bennetzen JL. 2000. Retrotransposons: central players in the structure, evolution and function of plant genomes. *Trends Plant Sci* 5:509-510.

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

Kumar S, Stecher G, and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology & Evolution* 33:1870.

Lam E, Shine J, Da Silva J, Lawton M, Bonos S, Calvino M, Carrer H, Silva-Filho MC, Glynn N, Helsel Z, Ma J, Richard E, Souza GM, and Ming R. 2009. Improving sugarcane for biofuel: engineering for an even better feedstock. *Global Change Biology Bioenergy* 1:251-255. 10.1111/j.1757-1707.2009.01016.x

Lee SI, Park KC, Son JH, Hwang YJ, Lim KB, Song YS, Kim JH, and Kim NS. 2013. Isolation and characterization of novel Ty1-*copia*-like retrotransposons from lily. *Genome* 56:495-503. 10.1139/gen-2013-0088

Llorens C, Munoz-Pomer A, Bernad L, Botella H, and Moya A. 2009. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biol Direct* 4:41. 10.1186/1745-6150-4-41

Ma Y, Sun H, Zhao G, Dai H, Gao X, Li H, and Zhang Z. 2008. Isolation and characterization of genomic retrotransposon sequences from octoploid strawberry (*Fragaria* x ananassa Duch.). *Plant Cell Rep* 27:499-507. 10.1007/s00299-007-0476-7

Mayer KFX, Rogers J, Doležel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, and Sourdille P. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788. 10.1126/science.1251788

Navarro-Quezada A, and Schoen DJ. 2002. Sequence evolution and copy number of Ty1-*copia* retrotransposons in diverse plant genomes. *Proc Natl Acad Sci U S A* 99:268-273. 10.1073/pnas.012422299

Okura VK, de Souza RS, de Siqueira Tada SF, and Arruda P. 2016. BAC-Pool Sequencing and Assembly of 19 Mb of the Complex Sugarcane Genome. *Front Plant Sci* 7:342. 10.3389/fpls.2016.00342

Oliveira AC, Pasqual M, Bruzi AT, Pio LA, Mendonca PM, and Soares JD. 2015. Flow cytometry reliability analysis and variations in sugarcane DNA content. *Genet Mol Res* 14:7172-7183. 10.4238/2015.June.29.11

Park JM, Schneeweiss GM, and Weiss-Schneeweiss H. 2007. Diversity and evolution of Ty1-*copia* and Ty3-*gypsy* retroelements in the non-photosynthetic flowering plants *Orobanche* and *Phelipanche* (Orobanchaceae). *Gene* 387:75-86. 10.1016/j.gene.2006.08.012

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 R, Ware D, Westhoff P, Mayer KF, Messing J, and Rokhsar DS. 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature* 457:551-556. 10.1038/nature07723

Pereira V. 2004. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biol* 5:R79. 10.1186/gb-2004-5-10-r79

Santini S, Cavallini A, Natali L, Minelli S, Maggini F, and Cionini PG. 2002. Ty1 /*copia*- and Ty3 /*gypsy*-like DNA sequences in *Helianthus* species. *Chromosoma* 111:192-200. 10.1007/s00412-002-0196-2

Schneider CA, Rasband WS, and Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9:671-675.

Steinhauer DA, and Holland JJ. 1986. Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *J Virol* 57:219-228.

Sun J, Huang Y, Zhou JY, Guo J, and Sun QB. 2013. LTR-retrotransposon diversity and transcriptional activation under phytoplasma stress in *Ziziphus jujuba*. *Tree Genetics & Genomes* 9:423-431. 10.1007/s11295-012-0563-y

Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, and Kakutani T. 2009. Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* 461:423-426. 10.1038/nature08351

Vitte C, and Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet Genome Res* 110:91-107. 10.1159/000084941

Vonholdt BM, Takuno S, and Gaut BS. 2012. Recent retrotransposon insertions are methylated and phylogenetically clustered in *japonic*a rice (*Oryza sativa spp. japonica*). *Mol Biol Evol* 29:3193-3203. 10.1093/molbev/mss129

444 Wicker T, and Keller B. 2007. Genome-wide comparative analysis of *copia* retrotransposons in Triticeae, rice, and
445 Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia
446 families. *Genome Res* 17:1072-1081. 10.1101/gr.6214107
447 Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E,
448 SanMiguel P, and Schulman AH. 2007. A unified classification system for eukaryotic transposable elements.
449 *Nat Rev Genet* 8:973-982. 10.1038/nrg2165
450 Zedek F, Smerda J, Smarda P, and Bures P. 2010. Correlated evolution of LTR retrotransposons and genome size in
451 the genus *Eleocharis*. *BMC Plant Biol* 10:265. 10.1186/1471-2229-10-265
452 Zhang JS, Nagai C, Yu QY, Pan YB, Ayala-Silva T, Schnell RJ, Comstock JC, Arumuganathan AK, and Ming R. 2012.
453 Genome size variation in three *Saccharum* species. *Euphytica* 185:511-519. 10.1007/s10681-012-0664-6
454 Zhang QJ, and Gao LZ. 2017. Rapid and Recent Evolution of LTR Retrotransposons Drives Rice Genome Evolution
455 During the Speciation of AA-Genome *Oryza* Species. *G3 (Bethesda)* 7:1875-1885. 10.1534/g3.116.037572
456

Table 1(on next page)

Saccharum officinarum Ty1-*copia* retrotransposon reverse transcriptase sequence characterization.

[#]sequence similarity given as minimum (mean) maximum,

A/T, adenine/thymine nucleotides; ORF, open reading frame.

Ty1-*copia* sequence	number	length（bp）range	similarily#	A/T	intact ORF
Ty1-*copia*	42	239-263/263(83%)	39(77)100	59%	74%
Tork/TAR	26	260-263/263(83%)	67(95)100	63%	77%
Tork/Angela	3	263,263,263	48(70)65	60%	33%
Retrofit/Ale	3	239,240,263	62(86)96	48%	100%
Sire/Maximus	10	258-263/263（92%）	55(76)93	58%	70%

1

Figure 1

Neighbor joining phylogenetic tree of amino acid sequences based on alignment of Ty1-*copia* retrotransposon RT sequences from *S. officinarum* with those from other graminaceous species.

Graminaceous species (*Aegilops, Hordeum, Oryza, Saccharum, Sorghum, Triticum*, and *Zeas*). Bootstrap values over 60 are indicated at the nodes.

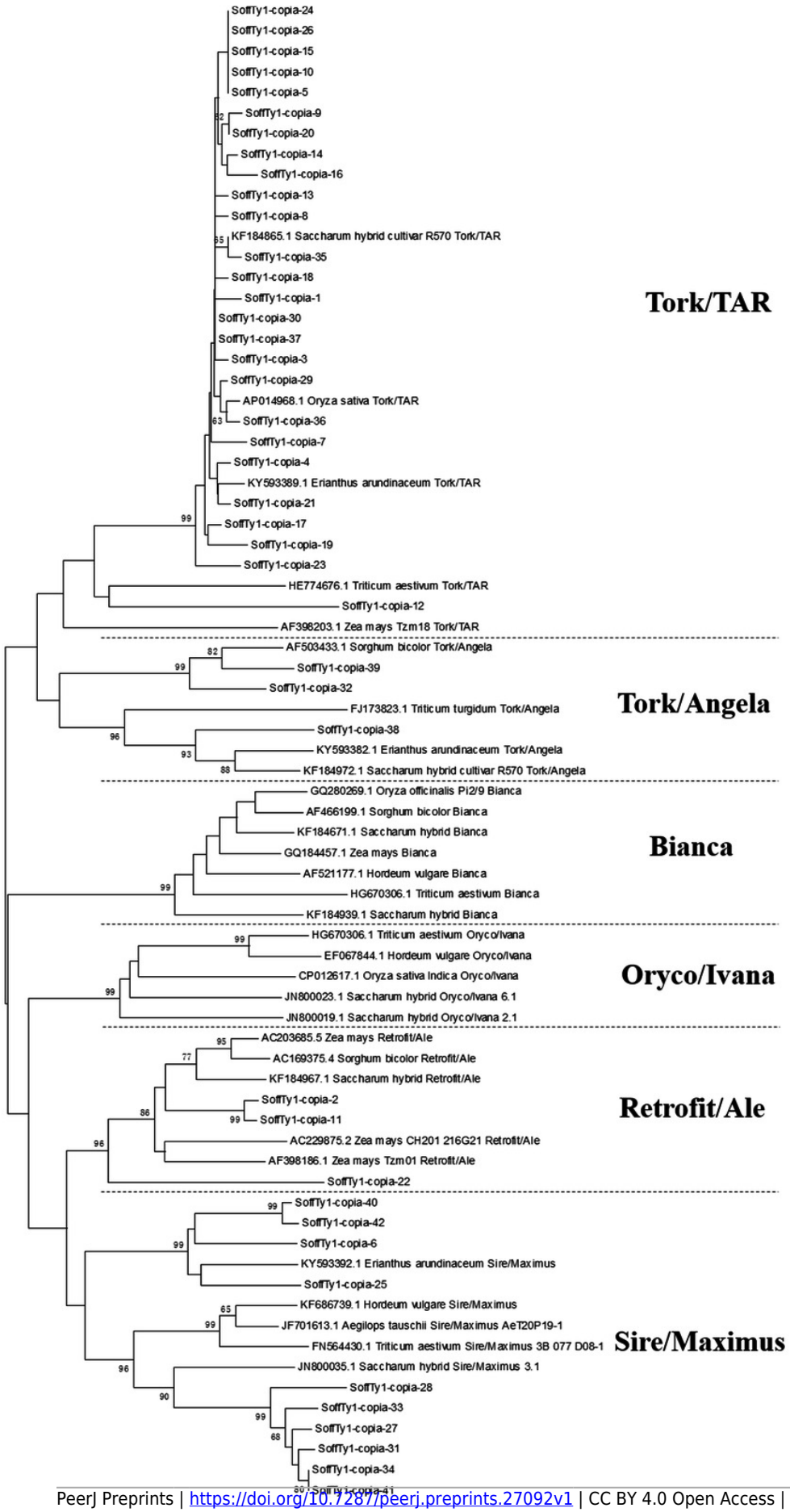

Figure 2

Dot blot estimation of the total copy number of Ty1-*copia* retrotransposon RT sequences in the *S. officinarum* genome.

Serial dilutions of genomic DNA from *S. officinarum* (row A) and PCR clones of Ty1-*copia* RT sequences (row B) were dot blotted on a membrane that was hybridized with a labeled PCR probe containing all Ty1-*copia* RT sequence clones.

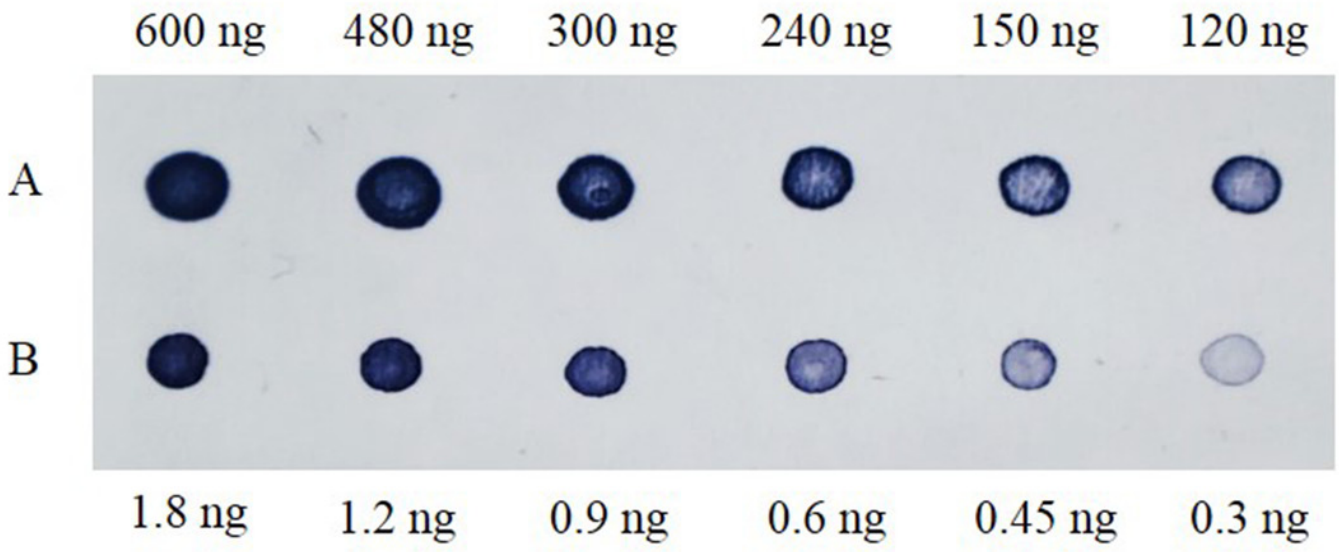

Figure 3

Localization of Ty1-*copia* retrotransposon RT sequences on metaphase and interphase chromosomes of *S. officinarum* (2n = 8*x* = 80) by fluorescence in situ hybridization (FISH).

Metaphase and interphase chromosomes stained with 4′,6-diamidino-2-phenylindole (DAPI) (a and d). Total Ty1-*copia* RT PCR products from *S. officinarum* were used as probes on metaphase chromosomes (b) and interphase nuclei (e). Scale bars = 5 mm.

**Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.*

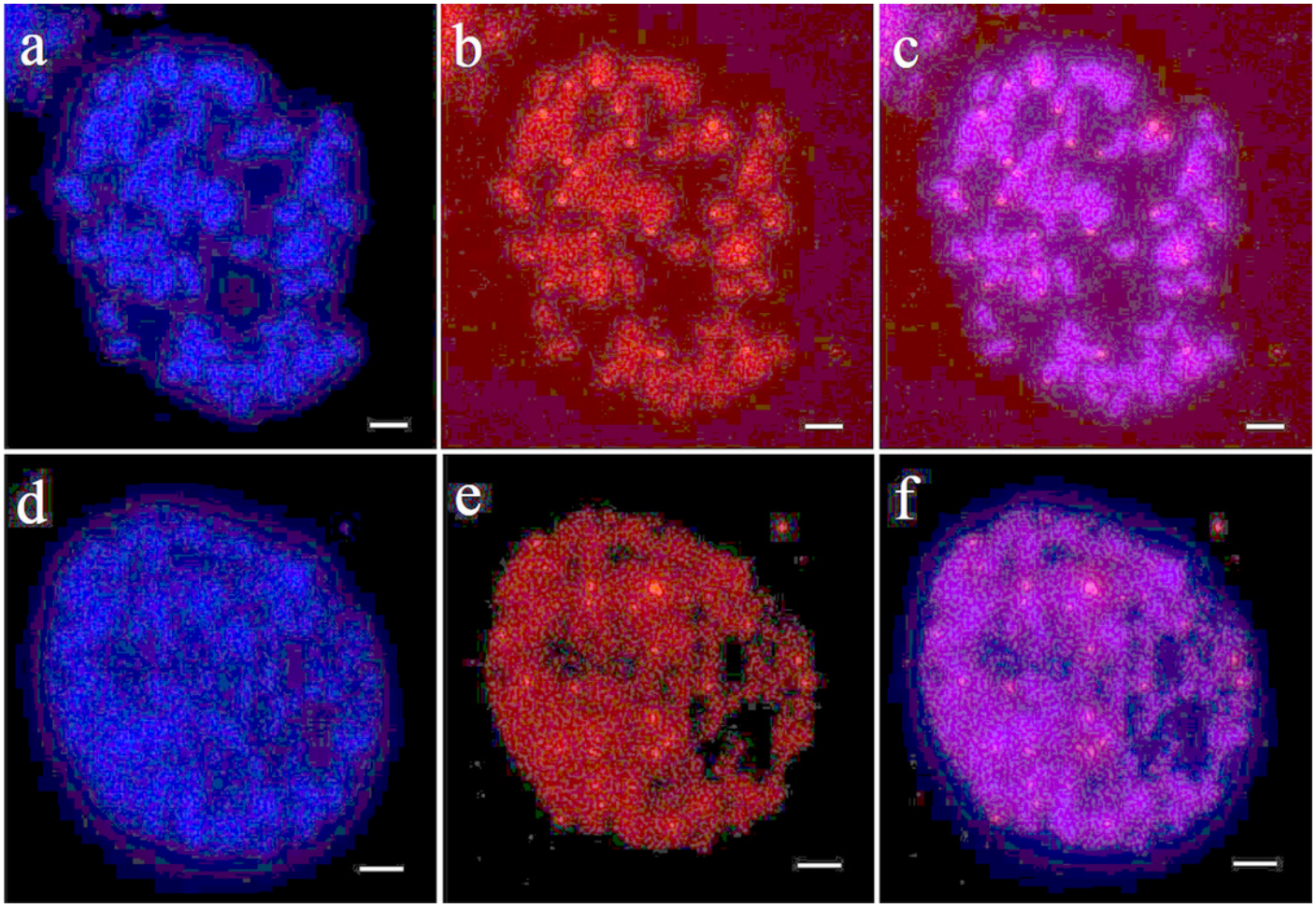

Figure 4

Localization of Ty1-*copia* retrotransposon RT sequences and 45S rDNA probes on metaphase chromosomes of *S. officinarum* (2n = 8*x* = 80) by fluorescence in situ hybridization (FISH).

Metaphase chromosomes stained with 4′,6-diamidino-2-phenylindole (DAPI). Red signals indicate Ty1-*copia*; green (arrows) signals indicate 45S ribosomal DNA. Scale bars = 5 mm.

**Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.*