1 Introduction

Transposable elements (TEs), which were first discovered in maize by Barbara McClintock [1], are mobile genomic DNA sequences that are ubiquitous and abundant components of all eukaryotic genomes, and play important roles in the structural organization and evolution of genes and genomes [2-6]. TEs occupy a considerable proportion of host genomes in eukaryotic species, especially in plants, such as wheat (80%) [7] and maize (85%) [8]. They can be classified into two categories according to the presence or absence of an RNA intermediate, retrotransposons (Class I) or DNA transposons (Class II) [9-11]. Retrotransposons (Class I) insert a copy of themselves into new chromosomal locations by an RNA intermediate (commonly referred to as a “copy and paste” mechanism) [6,9]. Due to a consequence of this mechanism, they are particularly abundant in plant genomes [7,8,12]. They can be further divided into LTR or non-LTR retrotransposons, depending on the presence or absence of an RNA intermediate, retrotransposons (Class I) or DNA transposons (Class II) [9-11]. Retrotransposons (Class I) insert a copy of themselves into new chromosomal locations by an RNA intermediate (commonly referred to as a “copy and paste” mechanism) [6,9]. Due to a consequence of this mechanism, they are particularly abundant in plant genomes [7,8,12]. They can be further divided into LTR or non-LTR retrotransposons, depending on the presence or absence of long terminal repeats (LTRs) [11].

Within the LTR retrotransposable elements, two subclasses, Copia and Gypsy, are particularly abundant in plants [11]. These two types differ in gene order between RNaseH, reverse transcriptase (rt), and integrase. In Copia, integrase is located upstream of the rt, while in Gypsy, the order of these two genes is reversed [3]. One of the key enzymes responsible for the copy and paste replication mechanism is reverse transcriptase. Previously, studies have suggested that amplification of rt in Copia and Gypsy using degenerate primers that are complementary to the conserved domains of rt, is a feasible and efficient method to characterize these two types of retrotransposons in various plant species [13-18].

As a representative genus of the cosmopolitan family Moraceae (Rosales), Morus (mulberry), comprising more than 13 species (over 1, 000 cultivars), are widely distributed in Asia, Africa, Europe, and the United States [19,20]. Its delicious fruit is a rich source of medicines that are used to treat certain serious diseases, and has...
attracted considerable attention [21,22]. The relationship between mulberry and silkworm is one of the most suitable examples of “plant defense-insect adaptation” [23,24]. The *Morus notabilis* (mulberry) has a relatively small genome (357 Mb) with 14 chromosomes (2n = 14) [25]. Although recent studies of the mulberry genome have provided some insight into its repetitive genome fraction [25,26], detailed characterization of LTR retrotransposable elements has not yet been analyzed.

In the present work, our aims were to characterize the diversity of *rt* fragments of *Copia* and *Gypsy* retrotransposable elements found in the mulberry genome amplified and cloned using degenerate primers. The present work also seeks to characterize their heterogeneity and phylogenetic relationships. Additionally, fluorescence in situ hybridization (FISH) was utilized to better define the distribution of these elements on the chromosomes. Undoubtedly, these results will further improve our understanding of the role of retrotransposons in the structural, functional, and evolutionary dynamics of mulberry genomes.

2 Methods

2.1 Plant materials and isolation of DNA

Young leaves of *M. notabilis* C.K.Schn, which is composed of 14 chromosomes, were collected from Ya’an, Sichuan Province, China. Total genomic DNA was extracted from the young leaves using a standard CTAB protocol and was used as a template for polymerase chain reaction (PCR)[27].

2.2 Polymerase chain reaction (PCR) and cloning of PCR products

The *rt* domains of *Copia* and *Gypsy* retrotransposons were amplified from the genomic DNA using degenerate primers (Table S1). PCR amplification was carried out using the StepOnePlus system (Applied Biosystems, USA). The PCR reaction mixture contained 20 ng DNA, 10 pmol of each primer (BGI, China), 0.25 mM of each dNTP (Takara, Japan), 10× PCR buffer (including Mg²⁺, Takara, Japan), and 1 U of rTaq polymerase (Takara, Japan). The PCR program was: 94 °C for 3 min; followed by 35 cycles at 94 °C for 1 min, 50 °C for 1 min, and 72 °C for 1 min, with a final elongation step at 72 °C for 7 min. PCR products were purified from the 1.5% agarose gels using the Agarose Gel DNA Extraction Kit (TaKaRa, Japan) and cloned into the pMD19-T vector (TaKaRa, Japan) following the manufacturer’s instructions. Randomly chosen recombinant colonies were selected for further analysis. Two independent rounds of PCR amplification and cloning were carried out for each of the two elements. Positive clones were verified by PCR and sequenced in both directions using M13 universal primers (Sangon Biotech). Clones were named with an abbreviated taxon name (Mno), an abbreviation for the type of element (C for *Copia* and G for *Gypsy*) and the clone serial number.

2.3 Sequence data and phylogenetic analysis

Cloned sequences were checked against NCBI (http://blast.ncbi.nlm.nih.gov/) and GIRI (http://www.girinst.org/) databases for homology to previously characterized plant retroelement sequences using BLAST [28]. The nucleotide and protein sequences were aligned using MUSCLE (version 3.8.31) with default parameters [29]. BioEdit (version 7.2.5) was then used to calculate the sequence identities with the BLOSUM62 matrix method [30]. The locations of stop codons in *rt* sequences were identified using the Gene Structure Display Server (GSDS, http://gsds.cbi.pku.edu.cn/) [31]. Sequence logos were produced by WebLogo (http://weblogo.threeplusone.com/) [32]. To analyze the selective pressure of *rt* sequences, only intact sequences were retained for further analysis. A codon alignment of corresponding DNA sequences was converted from a multiple sequence alignment of proteins by PAN2NAL, with universal code [33]. Codeml, a part of the PAML package, was used to perform selection analyses [34]. Nonsynonymous (ka) and synonymous (ks) rates were used to estimate the selective pressure of *rt* fragments. The ratios ka/ks > 1, ka/ks =1, and ka/ks < 1 signify positive selection (adaptive molecular evolution), neutral mutations, and negative selection (purifying selection), respectively [35].

In the present research, all phylogenetic trees were constructed based on nucleotide sequences. Nucleotide sequences of the isolated *rt* fragments from mulberry and other previously characterized plant species (Table S2) were aligned using MUSCLE (version 3.8.31) with default parameters [29]. MrModeltest (version 2.3) was utilized to find best-fit substitution models [36]. According to these models, MrBayes (version 3.2) was used to perform Bayesian inference of phylogeny [37].
2.4 Chromosome preparation and fluorescence in situ hybridization

Young leaves from mulberry, which were treated with 2 mM 8-hydroxyquinoline in darkness for 3 h at room temperature (24 °C), were used to make chromosome spreads. Samples were fixed in methanol: glacial acetic acid solution (3:1) for 2 h at 4 °C, incubated in KCl (1/15 M) for 30 min and digested by a cellulose (5%) and pectinase (2.5%) mixture at 37 °C for 3 h. Samples were then washed and fixed in double-distilled H2O, and spread onto slides until cell walls were completely degraded. The clones containing rt inserts were labeled with digoxigenin-11-dUTP by PCR using universal M13 primers according to the protocols for the PCR DIG Probe Synthesis Kit (Roche). Fluorescence in situ hybridization was completed according to a standard procedure [38]. Briefly, the prepared chromosomes (on slides) were treated with RNase (100 μg/ml) at 37 °C for 15 min, and then digested with proteinase K (1 μg/ml) at 37 °C for 10 min. Samples were denatured with formamide (70%) for 10 min at 72 °C, and then immediately treated for 5 min each with 70%, 90%, and 100% anhydrous ethanol solutions precooled to -20 °C. The hybridization mixture (2× SSC, 0.25 μg salmon sperm DNA, 10% SDS, 50% DS, 50% formamide and 400 ng labeled DNA probe) was pre-denatured at 96 °C for 6 min. Slides and the DNA probe were then denatured at 80 °C for 10 min and then allowed to maintained at 37 °C for 16 h. After stringent washes (10% formamide for 10 s, 2× SSC at 37 °C for 5 min, and 0.2% Tween-20 at room temperature for 5 min), the immunodetection of a digoxigenated DNA probe was carried out with the FITC-conjugated antidigoxigenin antibody (Roche). Dehydrated preparations were counterstained and mounted with 4′,6-diamidino-2-phenylindole (DAPI). Slides were analyzed using a Leica DM2500 fluorescence microscope. Images were collected using the CV-M4+CL progressive scan CCD camera (DM2500, Leica) and CytoVision software (version 7.3.1).

Ethical approval: The conducted research is not related to either human or animals use.

3 Results

3.1 Isolation of rt fragments

Sequences of expected length (Copia 260 bp and Gypsy 430 bp) were obtained by degenerate primers, which were designed to amplify the conserved domains of rt gene of Copia and Gypsy retrotransposons. All obtained sequences were screened against sequences available in the NCBI and GIRI databases using BLAST. Two independent rounds of PCR amplification and cloning were carried out for both Copia and Gypsy elements. The first round of PCR and cloning yielded 52 sequences of Copia rt and 48 sequences of Gypsy rt. The second independent round of PCR and cloning yielded 54 sequences of Copia rt and 53 sequences of Gypsy rt. In total, 106 clones of Copia rt and 101 clones of Gypsy rt with homology to known retroelements were selected for further analysis. The clones were named with an abbreviated taxon name (Mno), an abbreviation for the type of element (C for Copia and G for Gypsy) and the clone serial number. All sequences were deposited in GenBank under accession numbers: Copia (KY490139 - KY490244), Gypsy (KY490245 - KY490345) (Table S3).

3.2 Characterization of rt fragments

The length of isolated Copia rt fragments ranged from 240 bp to 278 bp (first round, 255 bp to 275 bp; second round, 240 bp to 278 bp; average, 265 bp) (Fig. 1A). Isolated Gypsy rt fragments ranged from 408 bp to 437 bp in length (first round, 422 bp to 437 bp; second round, 408 to 435 bp; average, 430 bp) (Fig. 1B). As indicated in Fig. 2, both Copia and Gypsy rt sequences were rich in AT bases. The AT/GC ratio of Copia rt ranged from 1.16 to 1.58 (first round,

Figure 1. Length distribution of rt sequences of Copia and Gypsy retrotransposons. A length distribution of Copia rt sequences. B length distribution of Gypsy rt sequences.
1.6 to 1.58; second round, 1.43 to 1.55; mean, 1.45). As a comparison, the ratio of Gypsy rt ranged from 1.34 to 1.4 (first round, 1.34 to 1.4; second round, 1.34 to 1.39; mean, 1.35). Pairwise comparisons revealed similarity between Copia rt and Gypsy rt ranging from 0.419 to 0.992 and 0.533 to 0.997, respectively (Fig. 3). Considering these results, we predict that both Copia and Gypsy rt fragments are highly heterogeneous, with a higher sequence divergence of Copia rt relative to Gypsy rt fragments.

To further characterize detected rt fragments, these sequences were translated into amino acids and alignments corrected for frame shifts were necessary to maintain an open reading frame (ORF). Results revealed that fifty-seven out of 106 Copia rt sequences (53.8%) were intact sequences, which may have functional consequence, and the remaining 49 sequence fragments (45.3%) contained at least one premature stop codon and/or indels which disrupted the reading frame (Table 1). As a comparison, among the 101 Gypsy rt sequences, 49 clones (48.5%) possessed intact reading frames, while 52 clones (51.5%) contained at least one premature stop codons and/or frameshift mutation (Table 1).

Distribution patterns of premature stop codons and frameshift mutation sites found in rt fragments was also investigated. As shown in Fig. S1 and Fig. 4, there were no apparent distribution patterns in both Copia and Gypsy rt fragments. For example, the number of premature stop codons of Copia and Gypsy rt was ranged from 1 to 8 and 1 to 10, respectively (Table S3).

The selective pressure on intact rt sequences, was calculated from the difference between nonsynonymous (ka) and synonymous (ks) rates. The difference between ka and s substitution rates (ka/ks) for Copia rt sequences ranged from 0.01 to 0.297 (Fig. 5A). By contrast, the rates

---

**Figure 2.** AT/GC proportion distribution of rt sequence clones of Copia and Gypsy retrotransposons. A AT/GC ratio distribution of Copia rt sequences. B AT/GC ratio distribution of Gypsy rt sequences.

**Figure 3.** Identity distribution of rt sequence clones of Copia and Gypsy retrotransposons. A Identity distribution of Copia rt sequences. B Identity distribution of Gypsy rt sequences. The sequences were aligned using MUSCLE (version 3.8.31) with default parameters. BioEdit was used to calculate the sequence identities by BLOSUM62 matrix method.
related taxa according to the APG (The Angiosperm Phylogeny Group, http://www.theplantlist.org). So, the results suggested that the rt sequences and the host species trees showed incongruence.

3.4 Distribution of LTR retrotransposons in mulberry genome

We also performed fluorescent in situ hybridization (FISH) to study the distribution of these rt sequences along mulberry chromosomes. Chromosomal localization of Copia rt and Gypsy rt was performed using a heterogeneous probe cocktail containing all isolated clones, respectively. FISH analysis revealed that both hybridization signals of Copia and Gypsy are preferentially located in pericentromeric heterochromatin of mulberry chromosomes, with few signals were detected in the telomeric regions (Fig. 7).

3.3 Phylogenetic analysis of rt fragments

Phylogenetic analyses indicated that mulberry Copia rt and Gypsy rt sequences are homologous to rt sequences in other species (Fig. 6). Copia rt and Gypsy rt fragments of these species were categorized into five and six groups, respectively. Some interesting features of the results reveal that some mulberry rt sequences clustered with rt sequences of other plants. For example, MnoG_52 grouped more closely with GU954448 (Cucumis hystrix), instead of clustering with other Gypsy rt fragments in the same Group I in mulberry (Fig. 6B). The same phenomenon was also found in the phylogenetic tree of Copia rt sequences (Fig. 6A). In fact, Cucumis hystrix and mulberry are distantly related taxa according to the APG (The Angiosperm Phylogeny Group, http://www.theplantlist.org). So, the results suggested that the rt sequences and the host species trees showed incongruence.

Table 1. Statistics of different type clones of Copia and Gypsy retrotransposons.

<table>
<thead>
<tr>
<th>Type</th>
<th>No. in Copia clones</th>
<th>Proportion (%)</th>
<th>No. in Gypsy clones</th>
<th>Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>√</td>
<td>57</td>
<td>53.8</td>
<td>49</td>
<td>48.5</td>
</tr>
<tr>
<td>*</td>
<td>48</td>
<td>45.3</td>
<td>52</td>
<td>51.5</td>
</tr>
<tr>
<td>+</td>
<td>20</td>
<td>18.9</td>
<td>18</td>
<td>17.8</td>
</tr>
<tr>
<td>++</td>
<td>19</td>
<td>17.9</td>
<td>18</td>
<td>17.8</td>
</tr>
</tbody>
</table>

√ Intact sequences; * Premature translation termination codons; + Frameshift mutation.

Figure 4. Sequence logos of rt protein sequences alignment. A Sequence logos of Copia rt sequences; B Sequence logos of Gypsy rt sequences. Intact, intact sequences; Premature stop, premature stop codon contained sequences; Frameshift, frameshift mutation contained sequences. The sequences were aligned using MUSCLE (version 3.8.31) under default parameters. Sequence logos of these sequences were produced by WebLogo (http://weblogo.threeplusone.com/).

(ka/ks) of Gypsy rt sequences ranged from 0.035 to 0.916 (Fig. 5B).
4 Discussion

4.1 Characterization of LTR retrotransposons

Results obtained from the analysis of sequence length (Fig. 1), AT content (Fig. 2), and sequence similarity (Fig. 3), indicate that both Copia rt and Gypsy rt sequences that were amplified from the mulberry genome exhibited a high degree of heterogeneity, and a greater sequence divergence of Copia rt in comparison to Gypsy rt sequences. These conclusions are in agreement with the data reported previously in other angiosperm species, such as Sorghum [39] and Orobanche [40]. Two reasons likely account for this phenomenon. Firstly, as noticed from Fig. S1, Fig. 4 and Table 1, 45.3% and 51.5% rt sequences contained at least one premature stop codons and/or indels which disrupted the reading frame. There were also no well-defined distribution patterns of these elements. This is due to retrotransposable elements, which often experience deletions or fragmentation by illegitimate or unequal homologous recombination in the rt region during the process of transposition [41]. Secondly, our selective pressure analysis suggests that both Copia and Gypsy were under strong purifying selection pressure (Fig. 5), which drives the evolution of LTR retrotransposable elements by “selective silencing”, with random mutation and eventual deletion from the host genome [42].

4.2 Phylogenetic analysis of LTR retrotransposons

Interestingly, we observed a phenomenon where MnoG_52 grouped with GU954448 (Cucumis hystrix), instead of grouping with other Gypsy rt fragments in Group I in mulberry (Fig. 6B). While Cucumis hystrix and mulberry are distantly related taxa based on the APG (The Angiosperm Phylogeny Group, http://www.theplantlist.org) and phylogenetic analysis carried out by He et al [25]. The same phenomenon has also been found in the phylogenetic tree of Copia rt sequences (Fig. 6A). Previous studies suggest that HTT (horizontal TE transfer) is likely responsible for this phenomenon. As shown by El Baidouri et al., twenty-six genomes produced evidence for at least one case of HTT, and this is thought to be important in TE-driven genome evolution. These HTT events were observed for species as distantly related as palm and grapevine, tomato and bean and poplar and peach [43]. Hence, further analysis was performed in this work to prove whether or not HTT is the reason for this phenomenon. According to previous research, three criteria have been defined for the detection of HTTs: (a) patchy distribution of the TEs within a group of taxa, (b) high sequence similarity of the TE between distantly related taxa and (c) phylogenetic incongruence between the TE and host phylogeny [44-46]. In the present work, phylogenetic incongruence was observed between the Copia and Gypsy and their host phylogeny. However, the range of nucleotide sequence similarity of Copia rt and Gypsy rt is from 0.291 to 0.803, 0.509 to 0.753 (except the Morus bombycis), respectively (Table S4). These results suggest that the phenomenon, which is noted in the present study, did not result from HTTs. Further investigation will be necessary to make a reasonable interpretation of this phenomenon.

4.3 Chromosomal localization of LTR retrotransposons

Accumulating evidence has unambiguously shown that TEs play important roles in the structural, functional and evolutionary dynamics of genomes [2-4]. Distribution patterns are associated with their function. For example, some LTR retrotransposable elements preferentially insert within other LTR retrotransposons in centromere regions, and these elements are rapidly duplicated, inserted repeatedly, and play a role in the formation of centromeres [47]. In the present study, FISH results revealed that most of the hybridization signals are preferentially concentrated in pericentromeric and distal regions of chromosomes (Fig. 7),
Figure 6A. Phylogenetic analysis of rt sequences in mulberry and other plants. A Phylogenetic analysis of Copia rt sequences; B Phylogenetic analysis of Gypsy rt sequences. Red color, rt sequences of other plants downloaded from GenBank. The phylogenetic trees were constructed based on nucleotide sequences. MUSCLE (version 3.8.31) was used to align these sequences. MrModeltest (version 2.3) was used to find the best-fit substitution models. MrBayes (version 3.2) was used to perform Bayesian inference of phylogeny based on the best-fit substitution models.
Figure 6B. Phylogenetic analysis of rt sequences in mulberry and other plants. A Phylogenetic analysis of Copia rt sequences; B Phylogenetic analysis of Gypsy rt sequences. Red color, rt sequences of other plants downloaded from GenBank. The phylogenetic trees were constructed based on nucleotide sequences. MUSCLE (version 3.8.31) was used to align these sequences. MrModeltest (version 2.3) was used to find the best-fit substitution models. Mrbayes (version 3.2) was used to perform Bayesian inference of phylogeny based on the best-fit substitution models.
which is consistent with distribution patterns of these elements found in other species, such as Triticeae (Hordeum spontaneum and Aegilops speltoides) [48], Chenopodium quinoa [49], Arabidopsis thaliana [50], Solanum lycopersicum [51], Glycine max [52], and Asteraceae (Helianthus deserticola and Helianthus paradoxus) [53]. The distribution bias, where TEs are preferentially inserted into the heterochromatin regions, might be beneficial for the retrotransposable elements facilitating escape from negative selection that arises from insertion into genes rich regions [54,55]. Additionally, LTR retrotransposable elements found in centromeres of various plant species have also been shown to contribute to the formation and function of centromeres [47,55]. Hence, our future studies will focus on the potential roles of these elements in centromere function.

Acknowledgments: Bi Ma and Ningjia He conceived and designed the experiments. Bi Ma and Lulu Kuang performed the experiments. Bi Ma, Youchao Xin, and Fei Hou analyzed the data. Bi Ma wrote the manuscript. All authors reviewed drafts of this paper. The authors thank all the laboratory members who provided advice during this work. This project was funded by “Fundamental Research Funds for the Central Universities (XDJK2015C116)” and “China Postdoctoral Science Foundation (2016M592622)”.

Conflict of interest: Authors state no conflict of interest.

References

Detailed analysis of Copia and Gypsy rt in mulberry genome


[36] Nylander J.A.A., MrModeltest v2.3. Program distributed by the author., Evolutionary Biology Centre, Uppsala University, 2008.


Supplemental Material: The online version of this article (DOI: 10.1515/biol-2017-0031) offers supplementary material.