


Article

Evolutionary Divergence between *Toona ciliata* and *Toona sinensis* Assayed with Their Whole Genome Sequences

Xi Wang^{1,2}, Yu Xiao^{1,2}, Zi-Han He^{1,2}, Ling-Ling Li^{1,2}, Yan-Wen Lv^{1,2} and Xin-Sheng Hu^{1,2,*} 

¹ College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China

² Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, Guangzhou 510642, China

* Correspondence: xinsheng@scau.edu.cn

Abstract: *Toona ciliata* and *Toona sinensis* belong to the *Toona* genus of the Meliaceae family and are important timber species in China. *T. ciliata* is an endangered species at level II due to overcutting and a low rate of natural regeneration. *T. sinensis* was cultivated as an economic and nutritious tree for more than 2000 years. The two species differ in flower and leaf morphological traits, reproductive systems, and range size of natural distribution. To reveal the potential molecular basis of these divergences, we examined the similarities and differences in their whole genome sequences. Results indicate that *T. ciliata* had a higher number of expanded gene families than *T. sinensis*. The whole genome duplication (WGD) occurred before their speciation. The long-terminal repeats (LTRs) insertion was earlier in the *T. ciliata* genome (3.2985 ± 2.5007 Mya) than in the *T. sinensis* genome (3.1516 ± 2.2097 Mya). Twenty-five gene families in the *T. ciliata* genome were detected to be under positive selection compared with background branches of ten different land species. The *T. ciliata* genome was highly collinear with the *T. sinensis* genome, but had low collinearity with the genomes of more distant species. These genomic and evolutionary divergences are potentially associated with the differences between *T. ciliata* and *T. sinensis* in terms of their reproductive systems and ecological adaptation.

Keywords: Meliaceae; *Toona ciliata*; *Toona sinensis*; whole genome duplication; long-terminal repeats; genome sequence



Citation: Wang, X.; Xiao, Y.; He, Z.-H.; Li, L.-L.; Lv, Y.-W.; Hu, X.-S. Evolutionary Divergence between *Toona ciliata* and *Toona sinensis* Assayed with Their Whole Genome Sequences. *Genes* **2022**, *13*, 1799. <https://doi.org/10.3390/genes13101799>

Academic Editor: Vladimir Filkov

Received: 28 July 2022

Accepted: 4 October 2022

Published: 5 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

T. ciliata and *T. sinensis*—both also known as Chinese mahogany—are deciduous or semi-deciduous tree species in the *Toona* genus of the Meliaceae family [1]. Both species are important timber species in Southern China. *T. ciliata* is mainly distributed in southeast Asia, in countries such as India, Malaysia, and Indonesia. It is distributed in southern China at 200–1000 m above sea level. Its wood is tough and of straight texture, with beautiful patterning, dark reddish-brown heartwood, and light sapwood. It is an ideal material for art processes and antique furniture, and is a precious wood species in China, with a significant economic value [2]. *T. ciliata* is an endangered species at level II due to overcutting and a low rate of natural regeneration [3]. Inbreeding depression occurs in natural forests of *T. ciliata* [4].

T. sinensis is mainly distributed in subtropical and tropical regions in China and it grows below 1500 m above sea level [1]. It was cultivated as a precious economic tree in China for more than 2000 years [5]. It has multi-functional purposes as a high-quality wood material and as a medicinal plant [6,7]. The species contains flavonoids that are commonly used to treat bacterial dysentery, pulmonary cough, hematochezia, enteritis, and other diseases [8]. In general, *T. sinensis* has considerable value as a nutritious, medicinal plant and as a timber species [1,9].

The two species are identified in a taxonomy based on their leaf and flower morphological traits [1]. They differ substantially in natural distribution and ecological adaptation. *T. ciliata* has a much narrower range of natural distribution than *T. sinensis* and is embedded within the natural distribution of *T. sinensis* in China (Figure 1). One well-known difference is that the young leaves of *T. sinensis* are edible and nutritious [10,11]; however, the young leaves of *T. ciliata* are not edible, implying that distinct metabolic chemicals or pathways could be involved in the biosynthesis of some unknown nutritious components. In addition, the two species have distinct reproductive systems. *T. sinensis* is functionally sexually dioecious and possesses a complete outcrossing system; *T. ciliate*, in contrast, retains perfect flowers and possesses a predominant outcrossing system, with partial selfing and inbreeding [12]. These differences could be related to their underlying differences in genetic basis and are associated with their genomic divergences formed through long-term evolutionary processes.



Figure 1. Different ranges of natural distributions of *T. ciliata* (black circles) and *T. sinensis* (red circles) in China. This figure was drawn by synthesizing the distribution maps of specimens of *T. sinensis* and *T. ciliata* from the Chinese Virtual Herbarium website. Distribution of each species can be searched by inputting the species names on the website. Each point represents a record of specimens collected from the corresponding position. The red star represents the position of Beijing, the capital city of China.

Although the two species are sympatric, natural or artificial hybridization is rarely reported in the literature. Their different ranges of natural distribution in China imply that *T. sinensis* is adaptive to more diverse habitats than *T. ciliata*. Natural selection could be involved in forming their divergence in geographical distribution. Based on the sequences of nuclear ITS and cpDNA segments (*trnS-trnG*, *psbB*, *psbT*, and *psbN* genes), Mueller et al. [13] inferred that *T. sinensis* and *T. ciliata* were divergent at about 15.0 to 46.0 Mya

(95%CI), implying that potentially large genetic differences could exist between them. Molecular marker studies, including the use of sequence-related amplified polymorphisms (SRAPs) and simple sequence repeats (SSRs), indicated that *T. ciliata* had a high level of population genetic differentiation ($F_{st} = 0.7924$ for SRAPs and 0.35 for SSRs) and significant effects of isolation by distance (IBD) in its natural distribution in China [12,14,15]. However, these strong patterns did not occur in the population genetic structure of *T. sinensis* although it has a wider range of geographical distribution [16–18]. Nevertheless, these molecular studies only partially revealed genetic and evolutionary divergences between the two species.

It is well understood that comparative genomics offers a powerful approach to studying phylogeny, identifying conserved genes and gene families [19], and mapping gene collinearity among species. It also helps to characterize transposable elements, gene duplications, gene selection (positive/purifying selection), and evolution. Such analyses were widely applied to addressing many important issues concerned with evolutionary mechanisms and divergence among species [20]. Therefore, to reveal a broader molecular basis of evolutionary divergences between *T. ciliata* and *T. sinensis*, we compared the similarities and differences between the two species in their whole genome sequences.

We recently sequenced the whole genome of *T. ciliata* and inferred that *T. ciliata* and *T. sinensis* were divergent at about 6–25 Mya [21,22]. On the basis of this study, we further investigated their evolutionary divergences using comparative genomic analyses. This included analyses of the expansion and contraction of the gene family, genome collinearity, the whole genome duplication (WGD) inferred from the evolutionary rates at the synonymous site (K_s) and the four-fold degenerate synonymous site (4DTv), and the LTR insertion times. The genomic difference between the two species was tested in acquiring transposable elements (TEs). Insights into the evolutionary divergence between *T. ciliata* and *T. sinensis* were gained through this combination of analyses. Moreover, similarly to Wang et al. [22], we examined the four species analyzed by Ji et al. [21], including *Arabidopsis thaliana*, *Eucalyptus grandis*, *Salix purpurea*, and *Prunus persica*, and six other angiosperm plant species (*Citrus maxima*, *Citrus reticulata*, *Populus tremula*, *Glycine max*, *Amborella trichopoda*, and *T. sinensis*) to provide more contexts for genomic comparison. This also helped view the evolutionary divergence of *T. ciliata* from more distant species of land plant.

2. Materials and Methods

The sample for genome sequencing was one individual *T. ciliata* var. *ciliata* ($2n = 2x = 56$) collected from Pupiao, Baoshan city, Yunnan Province, China (25.04N, 99.06E). Genomic DNA (gDNA) was isolated from young and healthy leaves of the plant using the cetyltrimethylammonium bromide (CTAB) method [23]. A high-quality de novo assembly of the *T. ciliata* genome was determined by combining Nanopore and Hi-C sequencing analyses. The genome sequences were downloaded from the CNGB Nucleotide Sequence Archive (Accession number: CNP0001985; Table S1 for website). Wang et al. [22] detailed the methods for genome sequencing and assembly, including gene prediction and annotations, estimates of gene family, and repeats or duplications. The genomic sequences of 10 other species (*A. thaliana* [24], *A. trichopoda* [25], *C. maxima* [26], *C. reticulata* [27], *E. grandis* [28], *G. max* [29], *P. persica* [30], *P. tremula* [31], *S. purpurea* [32], and *T. sinensis* [21]) were downloaded from different websites (Table S1). These species were selected partly to compare to Ji et al. [21] and partly for the availability of whole genome sequences belonging to several angiosperm tree species. However, genome sequences of *Azadirachta indica* [NCBI SRA 053330] [33] and *Xylocarpus granatum* (GenBank accession: GCA_019650275.1) in the Meliaceae family were not included because gene annotations (gff files) were not provided for downstream genomic research.

Gene predictions were analyzed using three approaches (ab initio, homology-based, and transcriptome) [22]. The specific gene family of *T. ciliata* was analyzed with cluster-Profile v3.14.0 [34] for GO and KEGG enrichment. All the gene families for the eleven

species were derived using OrthoFinder V2.4 software [35] and annotated with the PANTHER V15 database [36]. GO and KEGG enrichment analyses were performed on the unique gene families of *T. ciliata*. Our previous study obtained about 37,030 gene families in the *T. ciliata* genome [22]. *A. trichopoda* was set as the outgroup, and MCMCTREE in PAML v4.9i software [37,38] was used to calculate the divergence times. Timetree (<http://www.timetree.org/>, accessed on 12 April 2021) was used to search for the fossil times of the eleven species.

The phylogenetic relationships among the eleven species were constructed using the maximum likelihood (ML) method and IQ-TREE v1.6.11 software [39]. Construction was based on 1276 single copy genes of the eleven species. MAFFT v7.205 [40] was used to compare the sequence of each single copy gene family, and then the PAL2NAL V14 [41] program was used to convert the compared protein sequence into codon alignment. Subsequently, gBlocks v0.91b [42] was used to remove regions with poor sequence alignment or large differences. Finally, all well-aligned gene family sequences of each species were catenated to obtain a super gene. The model detection tool ModelFinder [43] provided by IQ-TREE was used for model detection, and the best model was obtained as jtt + f + i + g4. The evolutionary tree was constructed using the maximum likelihood (ML) method, with 1000 bootstraps. The ML method was used to estimate divergence times using the correlated molecular clock and JC69 model.

Based on the above gene annotations and ML phylogeny, we further conducted the following analyses from different perspectives:

(1) Expansion and contraction of gene family: From the phylogeny with estimates of divergent times and the gene family clusters in 11 species, CAFE v4.2 [44] was used to estimate the expanded or contracted numbers of gene families for each species compared to its most recent ancestor.

(2) Natural selection detection (K_a/K_s (ω) test): Single-copy gene families were obtained from *P. tremula*, *S. purpurea*, *A. thaliana*, *C. reticulata*, *C. maxima*, *T. sinensis*, and *T. ciliata*. We aligned the protein sequences of each gene family with MAFFT v7.205 (parameters: -localpair—maxiterate 1000) [40], and used PAL2NAL [41] to reverse to the codon alignment sequences. We used Codeml (f3 \times 4 model of codon frequencies) based on the branch-site model in PAML v4.9 [38]. Using the “Chi2” program in PAML, we performed the likelihood ratio test (LRT) for two contrasting models, model A (the foreground branch ω was set with positive selection, i.e., $0 < \omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$) vs. null model $\omega (< 1)$. The branch of *T. ciliata* was set to be foreground and all other branches were background (each background branch ω was set as $0 < \omega_0 < 1$, $\omega_1 = 1$). When the LRT results were significant (p value < 0.01), the Bayesian empirical Bayes (BEB) method [45] was used to obtain the posterior probability of the positive selection site. Positive selection at the codon site was determined under a probability of greater than 0.95 ($p > 0.95$ for $\omega > 1$).

(3) Genome collinearity between species: DIAMOND v0.9.29.130 [46] was used to align gene sequences between *T. ciliata* and *T. sinensis* to determine orthologous gene pairs, with an E-value of less than 10^{-5} and a C score of greater than 0.5, where the C score was filtered by JCVI v0.9.13 [47]. Next, using gene sequences (gff3 file), we determined whether orthologous gene pairs were adjacent or not on chromosomes. This step was mainly carried out with MCScanX (parameter setting: -m 5) [48]. Finally, the genes in all collinear blocks were derived and drawn by JCVI. The same method was also used to analyze the collinearity of the gene sequences of *T. ciliata* with *T. sinensis*, *G. max*, and *A. thaliana* (E-value $< 10^{-5}$, C score > 0.5).

(4) Whole genome duplication (WGD) analysis: Mutation parameters 4DTv at the four-fold degenerate synonymous sites [49] and the K_s values at synonymous sites were estimated using WGD v1.1.1 [50]. The occurrences of the relative times of WGD within species and between species were inferred from the distributions of the 4DTv and K_s estimates.

(5) LTR insertion times: LTR_FINDER v1.07 [51] was used to search for LTR sequences and filter out redundant LTR sequences. The flanking sequences on both sides of each LTR were extracted and compared to MAFFT (parameters: -localpair—maxiterate 1000). The

distance (k) between two sides in terms of the Kimura model [52] in MBOSS V6.6.0 [53] was calculated, and $t = k/(2 \times r)$ was used to estimate the divergent time with the molecular clock ($r = 7 \times 10^{-9}$) [54]. The Kolmogorov–Smirnov test (non-parameter statistical test) with R was used to infer the difference between *T. ciliata* and *T. sinensis* in LTR insertion times.

3. Results

3.1. Expansion and Contraction of Gene Families

From the phylogeny of the eleven species (Figure S1) and the analysis of gene family clusters, the contraction and expansion of gene families were estimated for each species relative to its ancestor. Both the family-wide p -value and a Viterbi p -value of less than 5% were set to define the presence or absence of expansion or contraction of gene families. Compared with their common ancestors, *T. ciliata* had 470 expanded gene families and 1 contracted gene family (Figure 2). *T. sinensis* had 31 expanded gene families but 325 contracted gene families. PANTHER ver.14 [36] was used to annotate all gene families that were expanded or contracted in the 11 species (Table S2). All other investigated species except *A. thaliana* had more expanded than contracted gene families to different extents, compared with their corresponding common ancestors. The gene family in the *A. trichopoda* genome, the earliest divergent species in angiosperms, was slightly expanded.

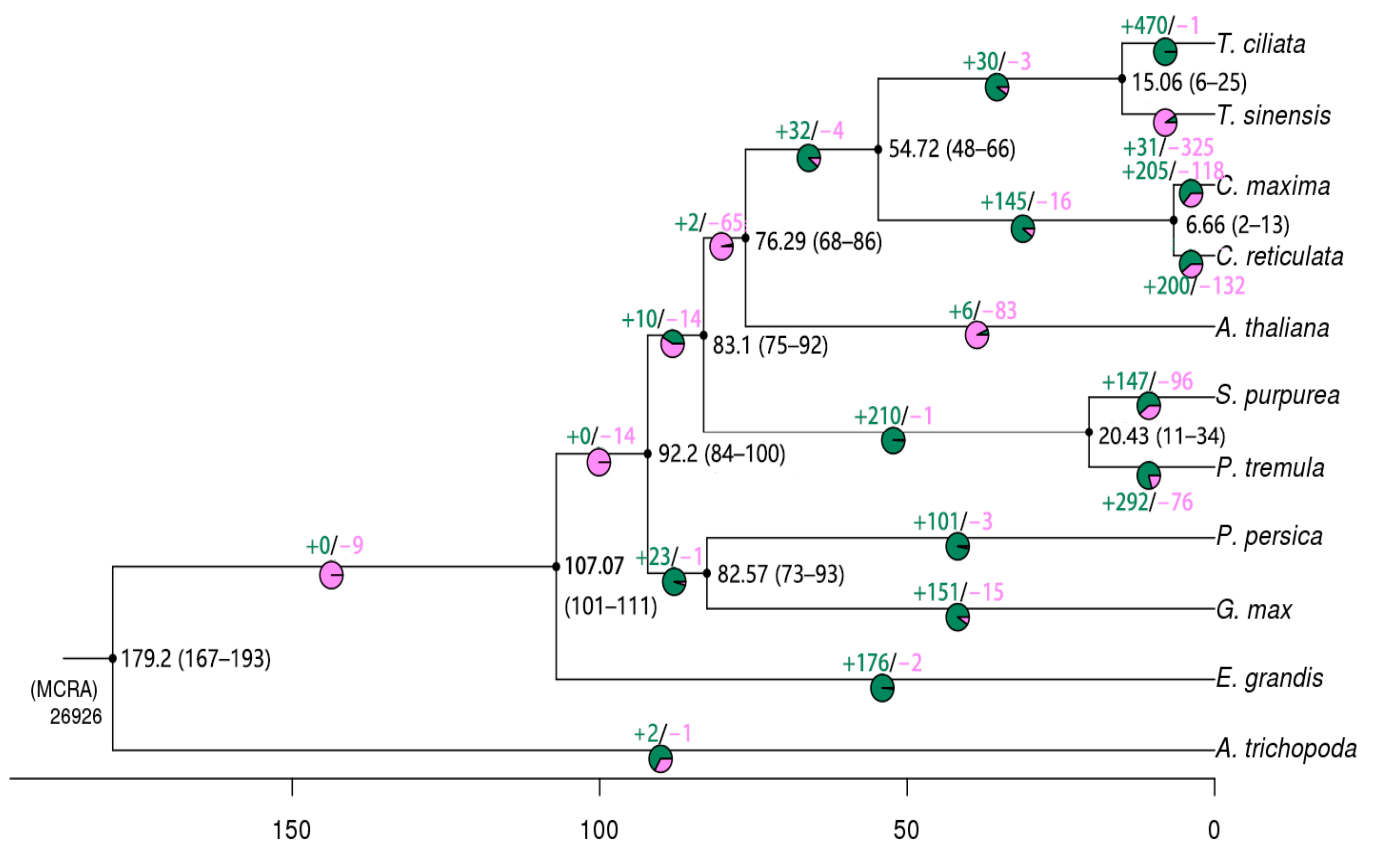


Figure 2. Expansion and contraction of gene families in eleven species. Pie chart sectors on each branch represent expansion of gene families in green and contraction of gene families in red. The values at each branch node are divergent times in Mya, with 95% confidence intervals in parentheses.

GO enrichment analyses show that the expanded gene families in *T. ciliata* were involved in many biological processes (Figure S2), including metabolic and cellular processes, biological regulation, responses to stimuli, signaling, reproduction and reproductive processes, multi-organism processes, and detoxification. In terms of cellular components, the expanded genes were involved in many components, such as cells and cell parts, mem-

branes and membrane parts, and organelles and organelle parts. The molecular functions of the expanded genes included the activities of binding, catalysis and transport, transcription regulation, determination of molecular structure, and molecular function regulation, etc. (Figure S2). The KEGG enrichment analyses show that the expanded genes were mostly associated with flavonoid biosynthesis, followed by the biosyntheses of sesquiterpenoid, triterpenoid, stilbenoid, diarylheptanoid, gingerol, and diterpenoid (Figure S2).

3.2. Detection of Positive Selection

Results show that twenty-five single-copy gene families were under positive selection (p -value of LRT test < 0.01); these are listed in Table S3. The codon sites with significant positive selection were specified in the single-copy gene family. The GO enrichment analyses indicate that these genes participated in diverse biological processes, including metabolic processes, cellular processes, responses to stimuli, organismal cellular component formation or biogenesis, and reproductive processes. In cellular components, these genes were involved in the formation of cells or cell parts, membranes or membrane parts, organelles, protein-containing complexes, membrane-enclosed lumens, cell junctions, and supramolecular complexes. In terms of molecular function, these genes were mainly involved in binding and catalytic functions (Figure S3). The KEGG analyses show that the genes under positive selection mostly participated in RNA transport, followed by the biosynthesis of glycosphingolipid, various types of N-glycan, ubiquinone and other terpenoid-quinone, and the degradation of glycosaminoglycan and other glycans (Figure S3).

3.3. Genome Collinearity Analysis

Figure 3a shows the collinearity relationship between twenty-eight shared chromosomes of the *T. ciliata* and *T. sinensis* genomes, indicating a close genetic relationship. Each chromosome pair in the two species has a strong collinear relationship. Figure 3b shows a dot matrix plot between *T. ciliata* and *T. sinensis*, exhibiting long diagonal lines across twenty-eight chromosomes but short broken lines perpendicular to the diagonals between neighbor chromosomes. However, *T. ciliata* had very weak collinearity relationships with *A. thaliana* and *G. max* (Figure S4), and most orthologous genes could have been lost or substantially divergent during the process of genome replications in each species.

3.4. Genome Duplications

The WGD event was identified by analyzing the distributions of 4DTv and KS estimates. Figure 4a shows the density distribution of the transversion rate, 4DTv, within *T. ciliata*, *T. sinensis*, *A. thaliana*, and *G. max*, or within the pairs of *T. ciliata* with one of the other three species. Different times of WGD events occurred within and between the species. Comparison of paralogous genes in duplicated collinear blocks revealed one peak of genetic distance at 0.0585 in the *T. ciliata* genome (11450 estimates) and a peak at 0.0530 in the *T. sinensis* genome (9524 estimates). One WGD event occurred in their genomes. Comparison of orthologous genes between *T. ciliata* and *T. sinensis* (15517 estimates) reveals one peak of genetic distance at 0.0151, indicating that the WGD event occurred before speciation. The difference in 4DTv between the two peaks, about 0.0406, is greater than 0.0151, indicating a long time interval from the WGD within genomes to speciation. Comparison of the orthologous genes of *T. ciliata* with *A. thaliana* or *G. max* show that they diverged a very long time ago.

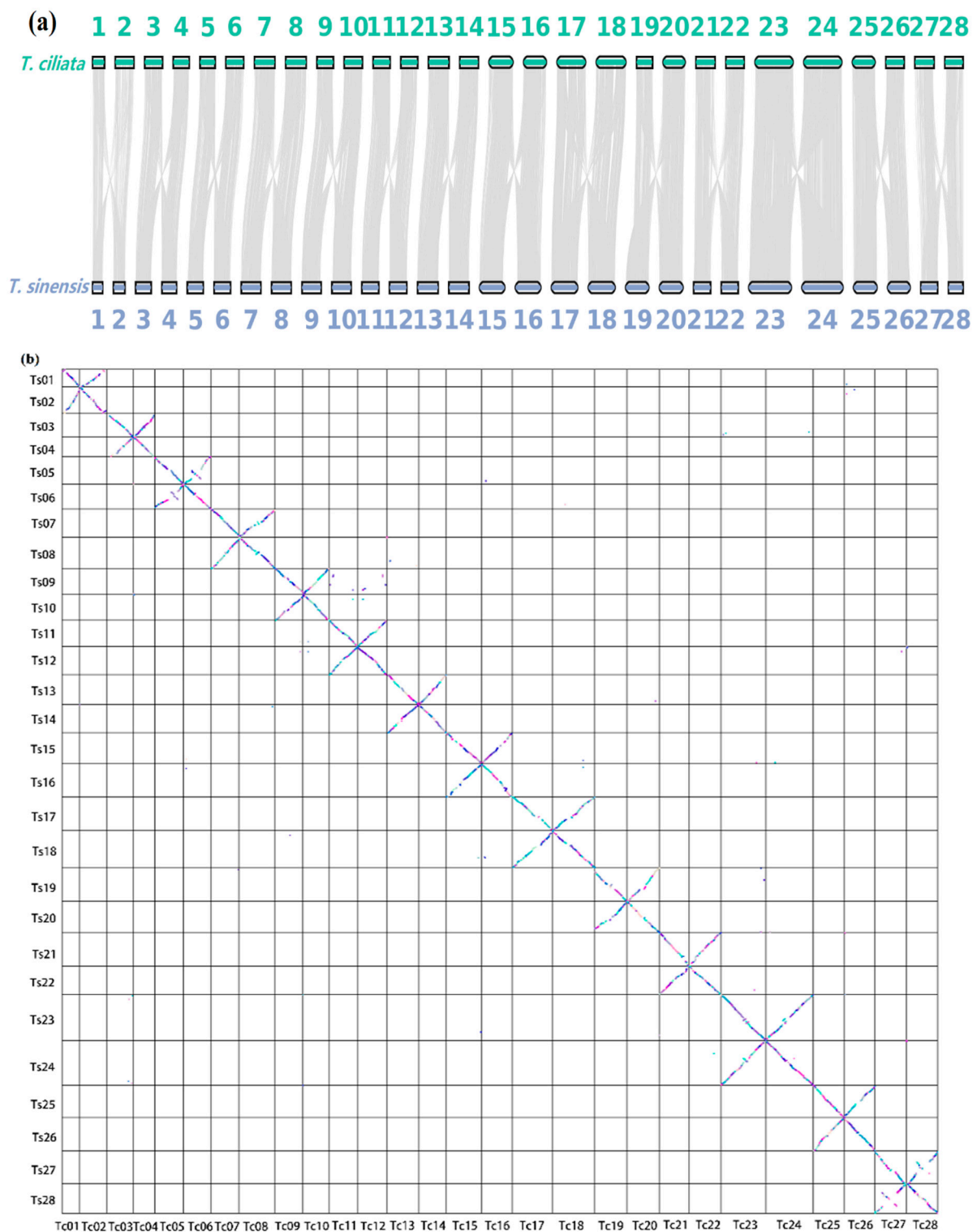


Figure 3. Collinearity map and dot plot of twenty-eight chromosomes shared by *T. ciliata* and *T. sinensis*: (a) Collinear map for each chromosome pair shared by the two species; (b) Dot plot of twenty-eight chromosomes shared by the two species. Symbols Tc01, Tc02, . . . , and Tc28 represent chromosomes 1, 2, . . . , and 28 of *T. ciliata* genome, respectively. Symbols Ts01, Ts02, . . . , and Ts28 represent chromosomes 1, 2, . . . , and 28 of *T. sinensis* genome, respectively.

Figure 4b shows the density distribution of the synonymous substitution rate (K_s) between paralogous genes within four species, or between orthologous genes within

the pairs of *T. ciliata* with one of the other three species. The results indicate one K_s peak at 0.1474 in the *T. ciliata* genome (11596 estimates) and one peak at 0.1365 in the *T. sinensis* genome (9703 estimates). One K_s peak also appears at 0.0463 when comparing the orthologous genes between *T. ciliata* and *T. sinensis* (16768 estimates), supporting the view that the WGD event occurred before the speciation of these two species. The difference in K_s between the peak for the WGD and the peak for speciation is about 0.0956, also indicating a long time interval between the WGD within the genomes and speciation. Comparisons of the orthologous genes of *T. ciliata* with *G. max* or *A. thaliana* show K_s peaks at 1.204 or larger, respectively, indicating that they were divergent a very long time ago.

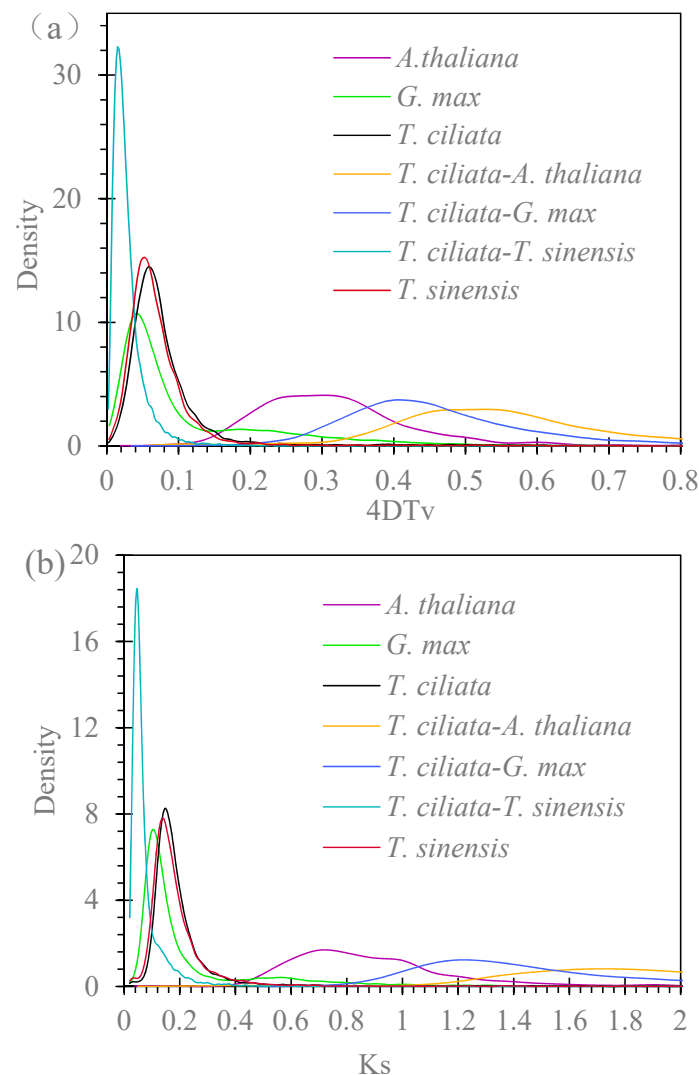


Figure 4. Evolutionary divergence in terms of whole genome duplication (WGD): (a) Distribution of 4DTv (four-fold degenerate transversion rate) within and between species genomes; (b) Distribution of K_s (synonymous substitution rate) within and between species genomes.

3.5. Divergence in LTR Insertion Time

Figure 5a shows the distributions of estimates of the LTR insertion times in five species, indicating large differences among species. Means of LTR insertion times were the longest in the genome of *A. trichopoda* (6.0478 ± 2.7766), followed by the genomes of *T. ciliata* (3.2985 ± 2.5007), *T. sinensis* (3.1516 ± 2.2097), *A. thaliana* (2.8794 ± 2.4481), and *A. max* (2.6980 ± 2.1762). LTR insertion generally occurred earlier into the *T. ciliata* genome than into the *T. sinensis* genome. A Kolmogorov–Smirnov test indicated that a significant

difference existed in the distribution of LTR insertion times between *T. ciliata* and *T. sinensis* (D statistics = 0.0553, p value = 1.506×10^{-8} , Figure 5b).

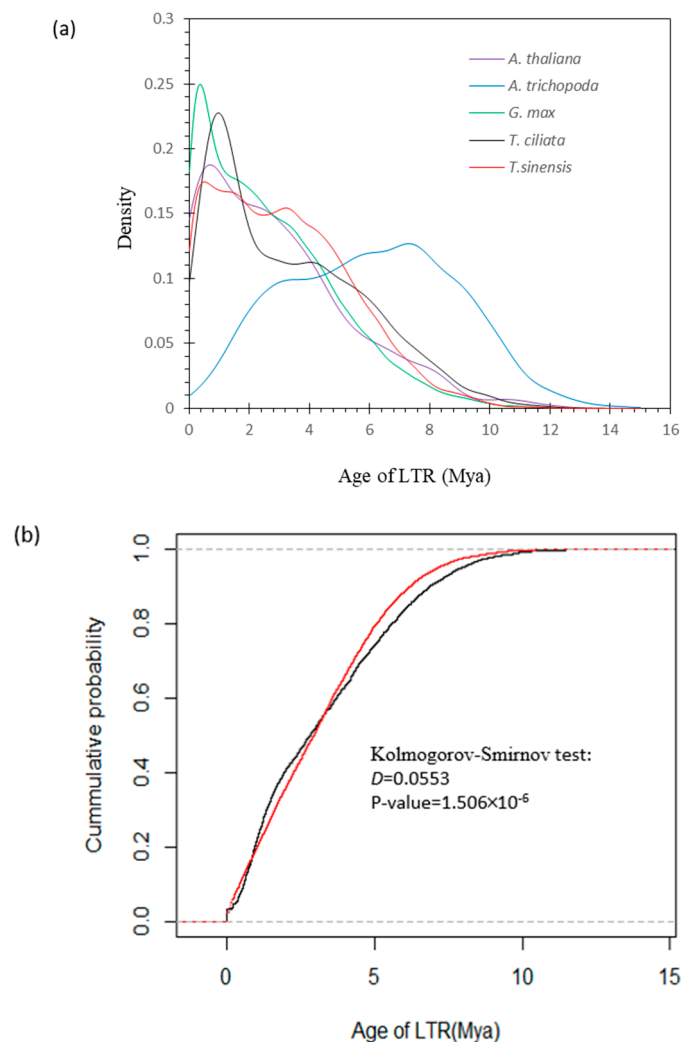


Figure 5. Evolutionary divergence in terms of LTR insertion times: (a) Distribution of the LTR insertion times in genomes of five species; (b) Cumulation probability of the distribution of the LTR insertion times in *T. ciliata* (black line; 4208 estimates) and *T. sinensis* (red line; 11,267 estimates) genomes.

4. Discussion

4.1. Genome Comparison

In this study, we compared recently published genome sequences of *T. ciliata* with the genomes of other land plant species, with an emphasis on the genomic divergence between *T. ciliata* and *T. sinensis*. Several divergences could be concluded between the *T. ciliata* and *T. sinensis* genomes. From the assemblies of the genome sequences of these two species [21,22], *T. ciliata* had a smaller genome size than *T. sinensis* (520 vs. 596 Mb), but harbored substantially more genes than *T. sinensis* (42159 vs. 34,345 genes). Except for the annotations with KOG or a different database (Interpro), both gene annotation percentages were comparable (97.92% for *T. ciliata* vs. 97.5% for *T. sinensis*), but the *T. ciliata* genome had more genes annotated than the *T. sinensis* genome in the Nr, Swissprot, KEGG, TrEMBL, and GO databases. The main reason for such different genome sizes could result from the larger abundant repeat sequences in the *T. sinensis* genome (48.62% for *T. ciliata* vs. 64.56% for *T. sinensis*). The *T. sinensis* genome had substantially more LTR sequences (219,375,967 bp for *T. ciliata* vs. 350,754,873 bp for *T. sinensis*) and a greater proportion of

LTR sequences (42.13% in the *T. ciliata* genome vs. 58.79% in the *T. sinensis* genome). The present study further indicates that the LTR sequences were more recently acquired in the *T. sinensis* genome than in the *T. ciliata* genome. This study also indicates that many gene families were lost in the *T. sinensis* genome during its speciation from their common ancestor.

Compared to the earlier analysis of *T. sinensis* genome [21], we detected only one peak in the distribution of either 4TDv or K_s estimates, each being comparable, respectively, to the first peak of the 4TDv or K_s estimates obtained by Ji et al. [21]. This consolidates the finding of one recent event of whole genome duplication. However, Ji et al.'s [21] second peak ($K_s = 1.48$) in Figure 5 was very small, and not detected in our analysis of the *T. sinensis* genome. Reasons for this are unknown.

4.2. Evolutionary Divergence and Its Implications

From TimeTree (<http://www.timetree.org/>, accessed on 12 April 2021), the divergent times between *T. ciliata* and *T. sinensis* covered a wide range of estimates in the literature, 7.4~48.4 Mya [55–57]. For instance, Muellner et al. [13] constructed a phylogeny using a few cpDNA segments and showed that the divergent time between the two species was 15.0 to 46.0 Mya. Based on 1276 single-copy gene sequences, the divergent times between *T. ciliata* and *T. sinensis* were at 15.06 (6–25 Mya) [22]. More precision estimates of divergent times between *T. ciliata* and *T. sinensis* need fossil record data of the species that are phylogenetically close to them, which could further confine the confidence intervals of divergent time estimates.

This study detected a large number of expanded gene families in the *T. ciliata* genome relative to the *T. sinensis* genome, which was likely formed in the long time process since speciation. Although these gene families were annotated (Table S3), their functions on phenotypes remain to be explored in the future. They could also be associated with the differences between the two species in phenotype or gene expression. Some of the expanded gene families and a few single-copy genes under positive selection were identified as being associated with reproduction and the reproductive process (Figures S2 and S3), implying that these genes could participate in species divergence in the reproductive system.

Generally, the genomic differences are likely associated with the differences in multiple traits, including the reproductive system and ecological adaptation. *T. ciliata* and *T. sinensis* adopt different strategies to avoid selfing and inbreeding. *T. ciliata* exhibits a predominantly outcrossing system, with selfing and inbreeding [12]. The functionally male and female flowers are disproportionally mixed and distributed in different positions on the same inflorescence, and most individuals are sexually monoecious rather than dioecious [58,59]. *T. sinensis* exhibits an outcrossing system by forming sexual diecious plants and potentially has self-incompatible genes. The more abundant LTRs and larger genome size in *T. sinensis* are likely related to the outcrossing system. This is consistent with the hypothesis that a selfing species reduces genome size and loses transposable elements [60], i.e., the characteristics of genomic selfing syndromes [61].

Difference in LTR abundance and insertion time could be related to the divergence in ecological adaptation. Although the historical cultivation could influence the present distribution of *T. sinensis* in China, a wider range distribution implies that *T. sinensis* is adaptive to more diverse habitats. Large abundant LTRs could be related to this broader adaptation of *T. sinensis*. Moreover, the genes that were present in the *T. ciliata* genome but lost in the *T. sinensis* genome are likely related to the adaptation of *T. ciliata*.

Evidence in the literature supports the significance of TEs in genome function and evolution [62,63]. The effects of LTRs on genome size variation were reported in wild barley *Hordeum spontaneum* where the copy number of BARE-1 LTR was significantly different between populations (8300~22100) and resulted in genome size variation [64]. The correlation among the BARE-1 LTR copy number, genome size, and local environmental conditions implied a linkage between the amplification of a specific TE and adaptive evolution. Wos et al. [65] also indicated that TEs including LTRs were adaptable to the

environment in *Arabidopsis arenosa*. Thus, abundant LTRs in the *T. sinensis* genome could be related to its larger range size of natural distribution in China. This awaits verification through appropriate data collection.

5. Conclusions

In this study, we examined the evolutionary divergence between *T. ciliata* and *T. sinensis* using their whole genome sequences. These two species are closely related within the *Toona* genus of the Meliaceae family. Their divergences are summarized as follows: (1) The *T. ciliata* genome had more expanded gene families than the *T. sinensis* genome following their speciation. Many gene families were lost in the *T. sinensis* genome. (2) Compared with *T. sinensis* and other land plant species investigated, *T. ciliata* had twenty-five gene families in its genome undergoing positive selection, and these genes were involved in diverse biological processes, cellular components, and molecular functions. (3) The *T. ciliata* genome was highly collinear with the *T. sinensis* genome, but had low collinearity with the genomes of distant species. (4) The whole genome duplication occurred in *T. ciliata* and *T. sinensis* before their speciation. A longer time interval existed between the main genome duplication and speciation times. (5) LTR insertion was earlier into the *T. ciliata* genome than into the *T. sinensis* genome. These genomic differences are likely associated with differences in their reproductive systems and ecological adaptation.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13101799/s1>, Figure S1: A phylogeny of 11 species based on 1276 single-copy genes, Figure S2: GO and KEGG analyses for expanded gene families in *T. ciliate*, Figure S3: GO enrichment and KEGG pathway analyses of the genes under positive selection, Figure S4: Chromosome collinearity analysis, Table S1: Websites and access numbers to download genome sequences of 11 plant species used in this study, Table S2: A list of expanded gene families in *T. ciliata* and its comparison with other 10 land plant species, Table S3: Summary of the genes and gene families of *T. ciliata* under positive selection detected using branch-site model.

Author Contributions: X.-S.H. and X.W. conceived and designed the study; X.W. analyzed data and drafted the manuscript; Y.X. analyzed data; Z.-H.H., L.-L.L. and Y.-W.L. participated in logistic assistance; X.-S.H. reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 32171819, and South China Agricultural University, grant number 4400-K16013.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Links to publicly archived datasets for all genome sequences of 11 species in this study are provided in Supplementary Materials.

Acknowledgments: We appreciate two anonymous referees for helpful comments on the earlier version of this article, and Yangqin Xie, Biomarker Biotechnology Company, Beijing, for useful discussions during the preparation of this paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Chen, S.-K.; Li, H.; Chen, B.-Y. Meliaceae. In *Flora of China* 43; Science Press: Beijing China, 1997.
2. Cheng, D.S.; Cui, T.L. Utilization value and cultivation techniques of *Toona sureni*. *For. By-Prod. Spec. China* **2010**, *4*, 39–40. [[CrossRef](#)]
3. Fu, L.G. *China Plant Red Data Book: Rare and Endangered Plants*; Science Press: Beijing, China, 1992; Volume 1.
4. Liang, R.L.; Liao, R.Y.; Dai, J. Endangered causes and protection strategy of *Toona ciliata*. *Guangxi For. Sci.* **2011**, *40*, 201–203. [[CrossRef](#)]

5. Jiao, Z.Y.; Zheng, J.W.; Yan, R.C.; Zhang, Y.; Tang, L.L. Overview of the main research on the multifunctional precious tree species *Toona sinensis*. *J. Jiangsu For. Sci. Technol.* **2019**, *46*, 51–56. [[CrossRef](#)]
6. Edmonds, J.M. The potential value of *Toona* species (Meliaceae) as multipurpose and plantation trees in Southeast Asia. *Commonw. For. Rev.* **1993**, *72*, 181–186.
7. Liao, J.W.; Yeh, J.Y.; Lin, Y.C.; Wei, M.M.; Chung, Y.C. Mutagenicity and safety evaluation of water extract of fermented *Toona sinensis* Roem or leaves. *J. Food Sci* **2009**, *74*, T7–T13. [[CrossRef](#)]
8. Hao, J.W. On the resource value and utilization of *Toona sinensis*. *Jiangsu Agric. Sci.* **2003**, *5*, 102–103. [[CrossRef](#)]
9. Zhang, J.F.; Zhang, Q.; Lu, G. Effects of *Toona sinensis* leaves extract on lipid metabolism and antioxidant activity of mice with hyperlipidemia. *J. Chin. Inst. Food Sci. Technol.* **2007**, *7*, 3–7. [[CrossRef](#)]
10. Edmonds, J.M.; Staniforth, M. *Toona sinensis* (Meliaceae). *Curtis's Bot. Mag.* **1998**, *15*, 186–193. [[CrossRef](#)]
11. Park, J.C.; Yu, Y.B.; Lee, J.H.; Choi, J.S.; Ok, K.D. Phenolic compounds from the rachis of *Cedrela sinensis*. *Korean J. Pharmacogn.* **1996**, *27*, 219–223.
12. Zhou, W.; Zhang, X.-X.; Ren, Y.; Li, P.; Chen, X.-Y.; Hu, X.-S. Mating system and population structure in the natural distribution of *Toona ciliata* (Meliaceae) in South China. *Sci. Rep.* **2020**, *10*, 16998. [[CrossRef](#)]
13. Mullner, A.N.; Pennington, T.D.; Koecke, A.V.; Renner, S.S. Biogeography of *Cedrela* (Meliaceae, Sapindales) in central and south America. *Am. J. Bot.* **2010**, *97*, 511–518. [[CrossRef](#)] [[PubMed](#)]
14. Li, P.; Zhan, X.; Que, Q.-M.; Qu, W.-T.; Liu, M.-Q.; OuYang, K.-X.; Li, J.-C.; Deng, X.-M.; Zhang, J.-J.; Liao, B.-Y.; et al. Genetic diversity and population structure of *Toona ciliata* Roem. based on sequence-related amplified polymorphism (SRAP) markers. *Forests* **2015**, *6*, 1094–1106. [[CrossRef](#)]
15. Zhan, X.; Li, P.; Hui, W.-K.; Deng, Y.-W.; Gan, S.-M.; Sun, Y.; Zhao, X.-H.; Chen, X.-Y.; Deng, X.-M. Genetic diversity and population structure of *Toona ciliata* revealed by simple sequence repeat markers. *Biotechnol. Biotechnol. Equip.* **2019**, *33*, 214–222. [[CrossRef](#)]
16. Xing, P.-Y.; Liu, T.; Song, Z.-Q.; Li, X.-F. Genetic diversity of *Toona sinensis* Roem in China revealed by ISSR and SRAP markers. *Genet. Mol. Res.* **2016**, *15*, 1–12. [[CrossRef](#)]
17. Chen, Q.; Rong, L.; Shao, Z.; Liu, T.; Wei, L.; Song, Z. Genetic diversity analysis of *Toona sinensis* germplasms based on SRAP and EST-SSR markers. *Acta Hort. Sin.* **2018**, *45*, 967–976. [[CrossRef](#)]
18. Zhou, W. Studies on Geographic Variation of Seeds and Seedling Traits and Genetic Diversity of *Toona sinensis* (Meliaceae). Ph.D. Thesis, South China Agricultural University, Guangzhou, China, 2020.
19. Norega, M.A.; Pennacchio, L.A. Comparative genomic analysis as a tool for biological discovery. *J. Physiol.* **2004**, *554*, 31–39. [[CrossRef](#)] [[PubMed](#)]
20. Baloch, A.A.; Kakar, K.U.; Nawaz, Z.; Mushtaq, M.; Abro, A.; Khan, S.; Latif, A. Comparative genomics and evolutionary analysis of plant CNGCs. *Biol. Methods Protoc.* **2022**, *7*, bpac018. [[CrossRef](#)] [[PubMed](#)]
21. Ji, Y.-T.; Xiu, Z.-H.; Chen, C.-H.; Wang, Y.; Yang, J.-X.; Sui, J.-J.; Jiang, S.-J.; Wang, P.; Yue, S.-Y.; Zhang, Q.-Q.; et al. Long read sequencing of *Toona sinensis* (A. Juss) Roem: A chromosome-level reference genome for the family Meliaceae. *Mol. Ecol. Resour.* **2021**, *21*, 1243–1255. [[CrossRef](#)] [[PubMed](#)]
22. Wang, X.; Xiao, Y.; He, Z.H.; Li, L.L.; Song, H.Y.; Zhang, J.J.; Cheng, X.; Chen, X.Y.; Li, P.; Hu, X.S. A chromosome-level genome assembly of *Toona ciliata* (Meliaceae). *Genome Biol. Evol.* **2022**, *14*, evac121. [[CrossRef](#)]
23. Doyle, J.J. A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
24. Sloan, D.B.; Wu, Z.; Sharbrough, J. Correction of persistent errors in *Arabidopsis* reference mitochondrial genomes. *Plant Cell* **2018**, *30*, 525–527. [[CrossRef](#)]
25. Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. *Science* **2013**, *342*, 1241089. [[CrossRef](#)]
26. Wang, X.; Xu, Y.; Zhang, S.; Cao, L.; Huang, Y.; Cheng, J.; Wu, G.; Tian, S.; Chen, C.; Liu, Y.; et al. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat. Genet.* **2017**, *49*, 765–772. [[CrossRef](#)]
27. Wang, L.; He, F.; Huang, Y.; He, J.; Yang, S.; Zeng, J.; Deng, C.; Jiang, X.; Fang, Y.; Wen, S.; et al. Genome of wild mandarin and domestication history of mandarin. *Mol. Plant* **2018**, *11*, 1024–1037. [[CrossRef](#)]
28. Bartholomé, J.; Mandrou, E.; Mabilia, A.; Jenkins, J.; Nabihoudine, I.; Klopp, C.; Schmutz, J.; Plomion, C.; Gion, J.-M. High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytol.* **2015**, *206*, 1283–1296. [[CrossRef](#)] [[PubMed](#)]
29. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J.; et al. Genome sequence of the palaeopolyploid soybean. *Nature* **2010**, *463*, 178–183. [[CrossRef](#)]
30. International Peach Genome Initiative; Verde, I.; Abbott, A.G.; Scalabrin, S.; Jung, S.; Shu, S.; Marroni, F.; Zhebentyayeva, T.; Dettori, M.T.; Grimwood, J.; et al. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat. Genet.* **2013**, *45*, 487–494. [[CrossRef](#)] [[PubMed](#)]
31. Lin, Y.C.; Wang, J.; Delhomme, N.; Schiffthaler, B.; Sundström, G.; Zuccolo, A.; Nystedt, B.; Hvidsten, T.R.; de la Torre, A.; Cossu, R.M.; et al. Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American and European aspen. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E10970–E10978. [[CrossRef](#)]
32. Zhou, R.; Macaya-Sanz, D.; Carlson, C.H.; Schmutz, J.; Jenkins, J.W.; Kudrna, D.; Sharma, A.; Sandor, L.; Shu, S.; Barry, K.; et al. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biol.* **2020**, *21*, 38. [[CrossRef](#)] [[PubMed](#)]

33. Krishnan, N.M.; Pattnaik, S.; Jain, P.; Gaur, P.; Choudhary, R.; Vaiyanathan, S.; Deepak, S.; Hariharan, A.K.; Krishan, P.B.; Nair, J.; et al. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *BMC Genom.* **2012**, *13*, 464. [[CrossRef](#)] [[PubMed](#)]
34. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omic* **2012**, *16*, 284–287. [[CrossRef](#)] [[PubMed](#)]
35. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [[CrossRef](#)]
36. Mi, H.; Muruganujan, A.; Ebert, D.; Huang, X.; Thomas, P.D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **2019**, *47*, D419–D426. [[CrossRef](#)]
37. Puttick, M.N. MCMCtreeR: Functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* **2019**, *35*, 5321–5322. [[CrossRef](#)] [[PubMed](#)]
38. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **1997**, *13*, 555–556. [[CrossRef](#)] [[PubMed](#)]
39. Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)] [[PubMed](#)]
40. Katoh, K.; Asimenos, G.; Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **2009**, *537*, 39–64. [[CrossRef](#)] [[PubMed](#)]
41. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34* (Suppl. S2), W609–W612. [[CrossRef](#)]
42. Talavera, G.; Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **2007**, *56*, 564–577. [[CrossRef](#)] [[PubMed](#)]
43. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.; Haeseler, V.A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587. [[CrossRef](#)] [[PubMed](#)]
44. Han, M.V.; Thomas, G.W.; Lugo-Martinez, J.; Hahn, M.W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **2013**, *30*, 1987–1997. [[CrossRef](#)] [[PubMed](#)]
45. Deely, J.J.; Lindley, D.V. Bayes empirical Bayes. *J. Am. Stat. Assoc.* **1981**, *76*, 833–841. [[CrossRef](#)]
46. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59. [[CrossRef](#)] [[PubMed](#)]
47. Tang, H.; Krishnakumar, V.; Li, J. jcv: JCVI utility libraries. *Zenodo* **2015**. [[CrossRef](#)]
48. Wang, Y.; Tang, H.; DeBarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)] [[PubMed](#)]
49. Comeron, J.M. A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. *J. Mol. Evol.* **1995**, *41*, 1152–1159. [[CrossRef](#)]
50. Zwaenepoel, A.; Van-de-Peer, Y. Wgd-simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* **2019**, *35*, 2153–2155. [[CrossRef](#)] [[PubMed](#)]
51. Xu, Z.; Wang, H. LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **2007**, *35*, W265–W268. [[CrossRef](#)] [[PubMed](#)]
52. Kimura, M. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 454–458. [[CrossRef](#)] [[PubMed](#)]
53. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **2000**, *16*, 276–277. [[CrossRef](#)]
54. Ossowski, S.; Schneeberger, K.; Lucas-Lledó, J.I.; Warthmann, N.; Clark, R.M.; Shaw, R.G.; Weigel, D.; Lynch, M. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **2010**, *327*, 92–94. [[CrossRef](#)] [[PubMed](#)]
55. Cavers, S.; Telford, A.; Arenal Cruz, F.; Perez Castaneda, A.J.; Valencia, R.; Navarro, C.; Buonamici, A.; Lowe, A.J.; Vendramin, G.G. Cryptic species and phylogeographical structure in the tree *Cedrela odorata* L. throughout the Neotropics. *J. Biogeogr.* **2013**, *40*, 732–746. [[CrossRef](#)]
56. Koecke, V.; Muellner-Riehl, A.N.; Pennington, T.D.; Schorr, G.; Schnitzler, J. Niche evolution through time and across continents: The story of Neotropical *Cedrela* (Meliaceae). *Am. J. Bot.* **2013**, *100*, 1800–1810. [[CrossRef](#)] [[PubMed](#)]
57. Koenen, E.J.; Clarkson, J.J.; Pennington, T.D.; Chatrou, L.W. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytol.* **2015**, *207*, 327–339. [[CrossRef](#)]
58. Styles, B.T. The flower biology of Meliaceae and its bearing on tree breeding. *Silvae Genet.* **1972**, *21*, 175–182.
59. Gouvea, C.F.; Dornelas, M.C.; Rodriguez, A.P.M. Floral development in the tribe Cedreleae (Meliaceae, sub-family Swietenioideae): *Cedrela* and *Toona*. *Ann. Bot.* **2008**, *101*, 39–48. [[CrossRef](#)]
60. Li, L.-L.; Wang, X.; Xiao, Y.; Chen, X.-Y.; Hu, X.-S. On the theories of plant mating system and molecular evolution and their applications. *Sci. Sin. Vitae* **2021**. [[CrossRef](#)]
61. Wright, S.; Ncoss, R.W.; Foxc, J.P.; Barrett, S.C.H. Genomic consequences of outcrossing and selfing in plants. *Int. J. Plant Sci.* **2008**, *169*, 105–118. [[CrossRef](#)]
62. Li, W.-H. *Molecular Evolution*; Cambridge University Press: Cambridge, UK, 1997.

63. Bucher, E.; Reinders, J.; Mirouze, M. Epigenetic control of transposon transcription and mobility in *Arabidopsis*. *Curr. Opin. Plant Biol.* **2012**, *15*, 503–510. [[CrossRef](#)]
64. Kalendar, R.; Tanskanen, J.; Immonen, S.; Nevo, E.; Schulman, A.H. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6603–6607. [[CrossRef](#)]
65. Wos, G.; Choudhury, R.R.; Kolar, F.; Parisod, C. Transcriptional activity of transposable elements along an elevational gradient in *Arabidopsis arenosa*. *Mob. DNA* **2021**, *12*, 7. [[CrossRef](#)] [[PubMed](#)]