


RESEARCH

Open Access



The *Tetracentron* genome provides insight into the early evolution of eudicots and the formation of vessel elements

Ping-Li Liu^{1,2†}, Xi Zhang^{1,2†}, Jian-Feng Mao^{1,2†}, Yan-Ming Hong^{1,2†}, Ren-Gang Zhang³, Yilan E^{1,2}, Shuai Nie^{1,2}, Kaihua Jia^{1,2}, Chen-Kun Jiang⁴, Jian He⁵, Weiwei Shen^{1,2}, Qizouhong He^{1,2}, Wenqing Zheng^{1,2}, Samar Abbas^{1,2}, Pawan Kumar Jewaria^{1,2}, Xuechan Tian^{1,2}, Chang-jun Liu⁶, Xiaomei Jiang^{7,8}, Yafang Yin^{7,8}, Bo Liu^{7,8}, Li Wang⁹, Biao Jin⁹, Yongpeng Ma¹⁰, Zongbo Qiu¹¹, František Baluška¹², Jozef Šamaj¹³, Xinqiang He¹⁴, Shihui Niu^{1,2}, Jianbo Xie^{1,2}, Lei Xie⁵, Huimin Xu¹⁵, Hongzhi Kong¹⁶, Song Ge¹⁶, Richard A. Dixon^{1,17}, Yuannian Jiao^{16*} and Jinxing Lin^{1,2*} 

* Correspondence: jiaoy@ibcas.ac.cn; linjx@ibcas.ac.cn

[†]Ping-Li Liu, Xi Zhang, Jian-Feng Mao and Yan-Ming Hong contributed equally to this work.

¹⁶Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

¹Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing 100083, China
Full list of author information is available at the end of the article

Abstract

Background: *Tetracentron sinense* is an endemic and endangered deciduous tree. It belongs to the Trochodendrales, one of four early diverging lineages of eudicots known for having vesselless secondary wood. Sequencing and resequencing of the *T. sinense* genome will help us understand eudicot evolution, the genetic basis of tracheary element development, and the genetic diversity of this relict species.

Results: Here, we report a chromosome-scale assembly of the *T. sinense* genome. We assemble the 1.07 Gb genome sequence into 24 chromosomes and annotate 32,690 protein-coding genes. Phylogenomic analyses verify that the Trochodendrales and core eudicots are sister lineages and showed that two whole-genome duplications occurred in the Trochodendrales approximately 82 and 59 million years ago. Synteny analyses suggest that the γ event, resulting in paleohexaploidy, may have only happened in core eudicots. Interestingly, we find that vessel elements are present in *T. sinense*, which has two orthologs of *AtVND7*, the master regulator of vessel formation. *T. sinense* also has several key genes regulated by or regulating *TsVND7.2* and their regulatory relationship resembles that in *Arabidopsis thaliana*. Resequencing and population genomics reveals high levels of genetic diversity of *T. sinense* and identifies four refugia in China.

Conclusions: The *T. sinense* genome provides a unique reference for inferring the early evolution of eudicots and the mechanisms underlying vessel element formation. Population genomics analysis of *T. sinense* reveals its genetic diversity and geographic structure with implications for conservation.

Keywords: *Tetracentron sinense*, Vessel, Phylogenomic, Whole genome duplication, VND7, Resequencing, Genetic diversity



Background

Eudicots are the largest and most diverse group of angiosperms, containing 75% of all angiosperm species [1]. Recent phylogenetic studies identified several early diverging eudicot branches and a strongly supported clade called the core eudicots [2, 3] comprising the majority of eudicot species. The rise and eventual dominance of the core eudicots may have benefited from the well-known γ hexaploidization event in their most recent common ancestor [4]. However, the timing and nature of the γ event are still controversial, and a recent study proposed a two-step model for the γ event, with the first whole-genome duplication occurring in the common ancestor of all eudicots [5]. Sequencing more early diverging eudicots can help us understand the nature of the γ event, as well as other genomic changes that occurred during the early diversification of eudicots.

Within eudicots, the four extant eudicot lineages that diverged the earliest are the Ranunculales, Proteales, Trochodendrales, and Buxales [6, 7]. The order Trochodendrales consists of one family (Trochodendraceae) and two species (*Tetracentron sinense* and *Trochodendron aralioides*). Although there has been controversy about the phylogenetic position of the Trochodendrales [8, 9], most phylogenetic analyses using organellar and nuclear genes strongly support the placement of Trochodendrales as early diverging eudicots [6, 7, 10, 11]. Whole-genome sequences from several Ranunculales and Proteales are available [5, 12, 13], genomes from Trochodendrales such as *T. sinense* will facilitate the comparative genomic analyses and shed deeper sight into the early evolution of eudicots.

Xylem, the water-conducting tissue in vascular plants, is composed of different cell types. The tracheids and vessel elements are conducting cells of the xylem [14]. Vessel elements are one of the most important features of angiosperms and differentiate them from other vascular plants such as gymnosperms (except gnetophytes [15]), which lack vessel elements. As they differentiate, vessel elements produce secondary cell walls and eventually undergo programmed cell death (PCD), forming a hollow tube that conducts water from the roots to the apical parts of the plant [14]. For almost 150 years, *T. sinense* was considered to produce secondary xylem that has tracheids but lacks vessel elements [16–18]. However, some studies suggest that vessel elements are present in the secondary xylem of *T. sinense* [19, 20]. Therefore, whether the secondary xylem of *T. sinense* contains vessel elements remains an outstanding question that can be addressed by morphological studies, and examination of the genetic network that specifies vessel elements.

VASCULAR RELATED NAC-DOMAIN PROTEIN 7 (VND7) functions as a master regulator of vessel element formation [21]. VND7 directly regulates many genes such as the transcription factor genes *MYB46* and *MYB83*, the cellulose synthase (CES) genes *CESA4* and *CESA8*, and the xylem-specific papain-like cysteine peptidase (XCP) genes *XCP1* and *XCP2* [22–26]. These genes mediate secondary cell wall formation and PCD, the two essential processes of vessel element differentiation. VND7 is positively regulated by several factors, including the LOB DOMAIN-CONTAINING PROTEINs (LBDs) LBD18 and LBD30, and VND1–VND5 [14, 27, 28]. Whether an ortholog of VND7 is present in *T. sinense* and whether the regulatory network of *T. sinense* VND7 parallels the *Arabidopsis thaliana* regulatory network remain unknown.

T. sinense ($2n = 48$ [29]) is a tertiary relict species that is mainly found in eastern Asia, including eastern Nepal and southwestern and central China [30–32]. The climate change during the Pleistocene ice age and human disturbances in recent years resulted in a sharp decline in *T. sinense* population size [32]. *T. sinense* has been documented as an endangered species in China and is currently listed in Appendix III of the Convention of International Trade of Endangered Species [31]. Although the evolutionary history of *T. sinense* has been investigated by examination of a few plastid regions and inter-simple sequence repeat analysis [31, 32], controversies remain, including the identification of potential refugia. In addition, little is known about the genome-scale genetic diversity and the population dynamics of this species in response to climate change, from the Pleistocene to the present. Genome-wide resequencing of *T. sinense* accessions from different populations would help us assess the levels and patterns of genetic variation and the evolutionary history of this endangered species, providing a practical strategy for its conservation.

Here we report the sequencing, assembly, and annotation of the *T. sinense* genome and its resequencing in key accessions. Our major objectives were (1) to explore the genomic changes that occurred during the early evolution of eudicots and core eudicots, (2) to investigate whether *T. sinense* has vessel elements in its secondary xylem and whether it has VND7 (and if so, whether the VND7 regulatory network is as in *A. thaliana*), and (3) to investigate the genetic diversity and evolutionary history of *T. sinense*. Our results reveal the role of polyploidization in the early evolution of eudicots and the nature of the known γ event, as well as exploring other genomic changes that occurred during eudicot diversification. Our genome-wide resequencing provided further insight into the evolutionary history of *T. sinense*. Moreover, light microscopy and the scanning electron microscopy (SEM) observations of slices of *T. sinense* wood and analyses of X-ray computed microtomography (microCT) support the idea that vessel elements are present in the secondary xylem of *T. sinense*. We also investigated the genetic basis of the development of vessel elements by identifying *T. sinense* orthologs of VND7 and exploring their regulation. Therefore, our results provide genomic insight on the evolution of this intriguing species and on the evolution of xylem vessels.

Results

SMRT sequence and assembly of the *T. sinense* genome

To analyze the genome of *T. sinense*, we first generated 82 G Illumina paired-end reads (150 bp). The genome size of the sequenced individual was estimated to be ~ 1.12 Gb, and its heterozygosity rates were estimated to be 0.22%, based on k -mer analysis (k -mer frequency distributions) (Additional file 1: Fig. S1). We then used a combination of single-molecule real-time (SMRT) sequencing technology from Pacific Biosciences (PacBio) and high-throughput chromosome conformation capture (Hi-C) to produce the final sequenced and assembled genome of *T. sinense*. In total, we generated 14.8 M PacBio single-molecule long reads (average read length 8 kb, longest read 84 kb), for 118 Gb total sequence, giving $\sim 100\times$ coverage of the assembled genome. De novo assembly yielded 3389 contigs, with a contig N50 length of 1.99 Mb, which was further improved to 2.8 Mb with Gapcloser software [33]. The total assembly size was 1.17 Gb,

which is consistent with the genome size estimated by *k*-mer analysis (Additional file 1: Fig. S1).

To further refine the *T. sinense* assembly, Hi-C libraries were constructed and sequenced. The Hi-C read pairs were mapped onto the draft assembly (Additional file 1: Fig. S2), yielding an assembly with a scaffold N50 of 45 Mb. The final reference assembly comprised 24 chromosome-scale pseudomolecules hereafter referred to as chromosomes (Additional file 1: Fig. S2b), with maximum and minimum lengths of 65 Mb and 33 Mb, respectively. The total length of the chromosomes accounted for 92.24% (1.07 Gb) of the assembled genome size of 1.17 Gb. The assembled chromosome number is equal to previous estimates of the haploid chromosome number of *T. sinense* ($2n = 48$). BUSCO [34] assessment showed that 93.4% complete genes were captured.

Repeat and gene annotation

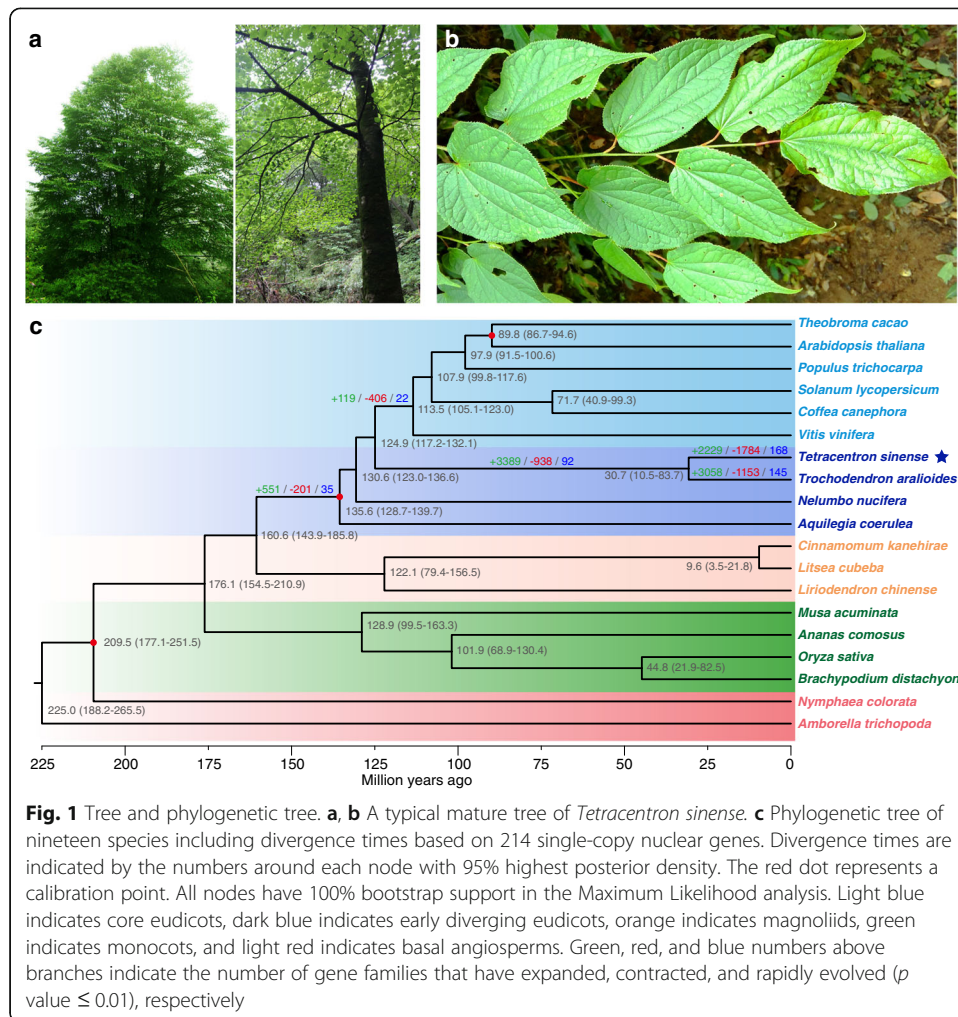
We identified 787 Mb of repetitive sequences in the genome of *T. sinense*, accounting for 67.8% of the genomic assembly. Transposable elements were the predominant component (65.9%), with the long terminal repeat (LTR) family being the largest part (32.2%) of these transposons. Within the LTR family, the *Gypsy* subfamily was the most abundant, making up 20.5% of the genome, followed by the *Copia* subfamily (11%) (Additional file 1: Table S1).

Based on the remaining repeat-masked *T. sinense* genome, we annotated 32,690 protein-coding gene models by combining ab initio gene predictions and evidence-based gene annotations using the MAKER pipeline. These gene models had an average transcript length of 1492 bp and an average coding-sequence length of 1304 bp. The gene models contained an average of 5.4 exons per gene with average exon and intron lengths of 153 bp and 283 bp, respectively. In addition, we identified 166 rRNAs, 811 tRNAs, and 1062 non-coding RNAs (ncRNAs). A BUSCO assessment based on conserved plant gene models identified 93.8% complete gene models.

Phylogenomic and gene family evolution analyses

To ascertain the phylogenetic position of *T. sinense* and the Trochodendrales, we performed phylogenomic analysis using 214 single-copy gene families (orthogroups) from 19 sequenced angiosperm genomes identified by OrthoFinder v2.3.1 [35]. These orthologous groups with single-copy genes were concatenated to construct a maximum likelihood (ML) phylogenetic tree using IQ-TREE. As shown in the ML tree (Fig. 1), the major lineages of angiosperms (monocots, magnoliids, and eudicots) were all recovered as monophyletic groups with 100% support. In eudicots, the core eudicot species formed a strongly supported monophyletic clade (100% bootstrap support). The Ranunculales, Proteales, and Trochodendrales formed successive sister clades to the core eudicots with 100% support, consistent with previous analyses [6, 7, 36].

We further estimated the divergence times of eudicot lineages using the MCMCTree [37] and BEAST 2 [38] programs; these two methods yielded consistent divergence times with only small differences (Fig. 1; Additional file 1: Fig. S3). According to the MCMCTree analysis (Fig. 1), crown-group eudicots arose 136 million years ago (MYA; 129–140 MYA, 95% highest posterior density). Apart from the Ranunculales, the earliest divergences within the eudicots based on our analyses were the split of the Proteales



and Trochodendrales from the rest of the eudicots at ~ 131 (123–137) MYA and ~ 125 (117–132) MYA, respectively. In our analysis, clades corresponding to the core eudicots originated at ~ 113 (105–123) MYA. Within the Trochodendrales, the divergence of its two species was estimated to have occurred at 31 (11–84) MYA.

We also investigated gene family evolution during the early diversification of eudicots. According to a CAFE analysis [39], five times as many gene families expanded in the lineage leading to all eudicots (551), as compared to core eudicots (110). However, twice as many gene families contracted in the lineage leading to core eudicots compared with the lineage leading to all eudicots (Fig. 1; Additional file 1: Fig. S4), consistent with a number of characteristics shared by eudicots. We found that in the lineages leading to eudicots, core eudicots, and the most recent ancestor of Trochodendrales, some defense gene families evolved rapidly. In the lineage leading to core eudicots, the TM3 MAD-box gene family that includes the important flowering time gene *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* has expanded (Additional file 1: Fig. S5a).

In the lineage leading to the common ancestor of the Trochodendrales, there was much more gene family expansion (3524) compared to gene family contraction (938).

To the genome of *T. sinense*, 2229 and 1784 gene families had undergone expansion and contraction, respectively. Among these families, 168 evolved rapidly. In contrast, more gene families expanded (3058) and fewer gene families (1153) contracted in *T. trochodendron*. The rapidly evolving gene families of *T. sinense* include many genes enriched in functions related to defense, toxin catabolic process, response to salicylic acid, and oxidative stress (Table S2). These genes mainly belong to gene families, such as the defense-related *NBS-LRR* (nucleotide-binding site leucine-rich repeat), lectin *RLK* (receptor-like kinase), and *GST* (glutathione transferase) gene families (Additional file 1: Fig. S5b). The gene families involved in terpenoid biosynthesis also evolved rapidly (Additional file 1: Table S2). Furthermore, the *ALG24* MAD-box genes expanded rapidly in *T. sinense* (Additional file 1: Fig. S5c). In agreement with the enrichment of GO terms, the KEGG enrichment suggested that genes involved in plant–pathogen interaction, glutathione metabolism, and terpenoid biosynthesis evolved rapidly in the *T. sinense* genome (Additional file 1: Fig. S6).

Whole-genome duplications

Whole-genome duplications (WGDs, polyploidization) have played a major role in the evolutionary history of the angiosperms. Intragenomic syntenic analysis of *T. sinense* showed strong collinearity between large segments of chromosomes, such as 2, 5, 6, and 24; 4, 9, 15, and 19; 4, 10, 18, and 23; 7, 11, 12, and 13; and 8, 14, 16, and 20 (Fig. 2a). The widespread occurrence of one-versus-three syntenic blocks suggested that two WGD events (named α and β) have occurred in the genome of *T. sinense*.

T. aralioides is a sister of *T. sinense*. To investigate whether *T. aralioides* also experienced WGDs, we performed an intragenomic analysis of this species. Indeed, two WGDs also occurred in *T. aralioides* since intragenome comparisons showed a 1:3 syntenic depth ratio (Additional file 1: Fig. S7a). To explore if the WGDs are common or lineage-specific, we performed intergenomic syntenic analyses between *T. sinense* and *T. aralioides*. The results showed a 1:1 syntenic relationship between *T. sinense* and *T. aralioides* (Additional file 1: Fig. S7b), suggesting that the two WGDs are common to the Trochodendrales.

Furthermore, we assessed the synteny between *T. sinense* or *T. aralioides* and *Amborella trichopoda*, which is the only living representative of the sister lineage to all other extant flowering plants and did not experience lineage-specific WGD events [40]. We identified a 1:4 syntenic depth ratio in the *A. trichopoda*–*T. sinense* (Fig. 2b; Additional file 1: Fig. S8a) and *A. trichopoda*–*T. aralioides* (Additional file 1: Fig. S8b) comparisons. We also performed intergenomic comparisons between *T. sinense*/*T. aralioides* and several species including the basal angiosperm *Nymphaea colorata*, the magnoliid *Cinnamomum kanehirae*, two early diverging eudicots (*Aquilegia coerulea* and *Nelumbo nucifera*), and the core eudicot *Vitis vinifera*. Genomic comparisons of *T. sinense*/*T. aralioides* and *N. colorata* showed a 4:2 syntenic pattern, and comparisons of *T. sinense*/*T. aralioides* and *C. kanehirae* showed a 4:4 syntenic pattern (Fig. 2b, c; Additional file 1: Fig. S8c–f). Given the known lineage-specific WGD in *N. colorata* and two lineage-specific WGDs in *C. kanehirae* [41, 42], our results suggested that the two WGDs in Trochodendrales occurred after the divergence of eudicots and magnoliids.

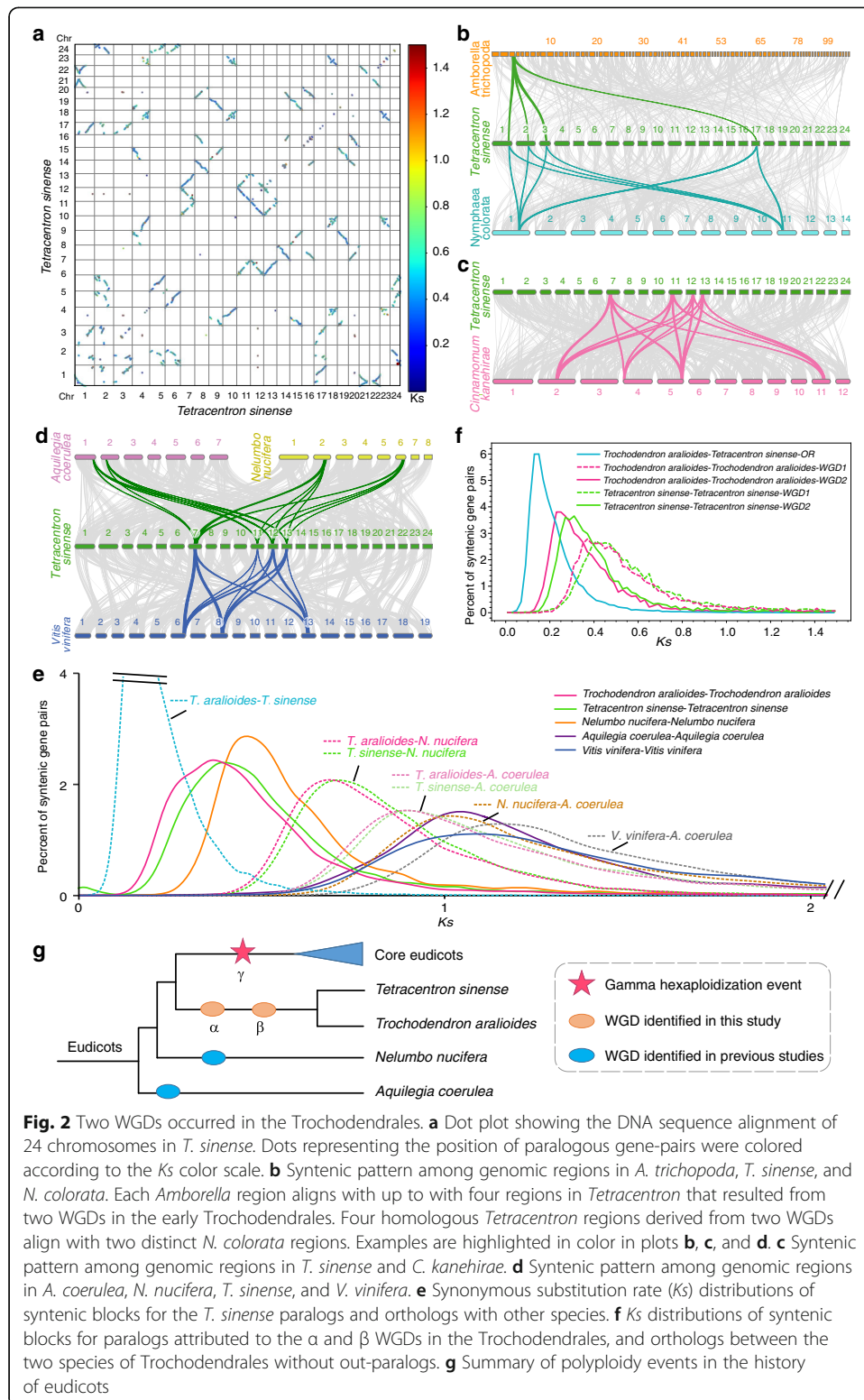


Fig. 2 Two WGDs occurred in the Trochodendrales. **a** Dot plot showing the DNA sequence alignment of 24 chromosomes in *T. sinense*. Dots representing the position of paralogous gene-pairs were colored according to the *Ks* color scale. **b** Syntenic pattern among genomic regions in *A. trichopoda*, *T. sinense*, and *N. colorata*. Each *Amborella* region aligns with up to with four regions in *Tetracentron* that resulted from two WGDs in the early Trochodendrales. Four homologous *Tetracentron* regions derived from two WGDs align with two distinct *N. colorata* regions. Examples are highlighted in color in plots **b**, **c**, and **d**. **c** Syntenic pattern among genomic regions in *T. sinense* and *C. kanehirae*. **d** Syntenic pattern among genomic regions in *A. coerulea*, *N. nucifera*, *T. sinense*, and *V. vinifera*. **e** Synonymous substitution rate (*Ks*) distributions of syntenic blocks for the *T. sinense* paralogs and orthologs with other species. **f** *Ks* distributions of syntenic blocks for paralogs attributed to the α and β WGDs in the Trochodendrales, and orthologs between the two species of Trochodendrales without out-paralogs. **g** Summary of polyploidy events in the history of eudicots

In addition, the synteny relationship between *N. nucifera* and *N. colorata* and between *N. nucifera* and *C. kanehirae* is 2:2 and 2:4, respectively (Additional file 1: Fig. S8g, h), indicating one WGD in *N. nucifera* after the divergence from *N. colorata* and

C. kanehirae. Several earlier studies showed that the WGD in *N. nucifera* is lineage-specific rather than eudicot-wide, rejecting the possibility that any WGD occurred in the common ancestor of all extant eudicots [13, 43, 44]. The collinear relationships between *T. sinense*/*T. aralioides* and *A. coerulea*, *N. nucifera*, and *V. vinifera* are 4: 2, 4:2, and 4:3 (Fig. 2d; Additional file 1: Fig. S9), respectively, consistent with independent WGDs in the respective lineages of the Trochodendrales, *A. coerulea*, *N. nucifera*, and *V. vinifera*.

Distributions of synonymous substitutions per synonymous site (K_s) were further analyzed to confirm the WGDs in Trochodendrales and other early diverging lineages of eudicots. The K_s distributions of *T. sinense* paralogs and *T. aralioides* paralogs in syntenic regions showed two overlapping peaks supporting two WGDs in a common ancestor of *T. sinense* and *T. aralioides* (Fig. 2e). As shown in Fig. 2e, the peaks of K_s distribution of *T. sinense*/*T. aralioides* paralogs (retained from two WGD events) are much smaller than those of orthologs comparing *T. sinense*/*T. aralioides* to *A. coerulea* and to *N. nucifera*, strongly suggesting the two WGDs in *T. sinense*/*T. aralioides* occurred after their divergence from *A. coerulea* and *N. nucifera*. In addition, the *N. nucifera* paralogs displayed lower peak K_s value than orthologs comparing *N. nucifera* to *T. sinense*/*T. aralioides* and to *A. coerulea*, further confirming that the WGD in *N. nucifera* is also lineage-specific.

According to the synteny analysis, we investigated the K_s distributions for paralogous gene pairs that originated from the two WGD events (Fig. 2f). Using divergence time and median K_s values of syntenic blocks between *T. sinense* and *T. aralioides*, we estimated the synonymous substitutions per site per year as $3.25e-9$ for the Trochodendrales (Additional file 1). Using this rate, we estimated that the α and β WGD events of Trochodendrales occurred at around 82 MYA and 59 MYA according to the median K_s of *T. sinense* paralogs, and 78 MYA and 54 MYA according to the median K_s of *T. aralioides* paralogs. This dating again suggests that the two WGDs occurred after the divergence of Trochodendrales from Proteales (*N. nucifera*).

Lastly, we performed integrated syntenic and phylogenomic analyses to confirm the timing of these WGDs in early diverging eudicots. We obtained anchor genes from the inter-genomic synteny blocks with ratio of 1:2:4 in *A. trichopoda*, *N. nucifera*, and *T. sinense*. All genes in the same sets of syntenic block were concatenated and used to construct phylogenetic trees. In total, we obtained 31 groups of concatenated genes (a total of 94 genes) and 87% of the trees based on the concatenated genes (Additional file 1: Fig. S10) supported the hypothesis that the WGDs in *N. nucifera* and *T. sinense* are lineage-specific.

Occurrence of vessel elements in secondary xylem

Under a light microscope, we found there was a distinct growth ring delineated by a band of latewood in the cross-section. Earlywood cells were larger and thin-walled, while those in latewood were smaller and thick-walled as reported by the previous study [17]. Occasionally, we noticed some peculiar cells occurred through the growth ring boundaries with much larger diameter ($56.0 \pm 0.82 \mu\text{m}$, $n = 35$) than normal cells ($44.7 \pm 0.98 \mu\text{m}$, $n = 35$) (Fig. 3a). In the radial section, the peculiar cells had oblique end-walls (Fig. 3b). In the tangential section, we can clearly observe there are two types

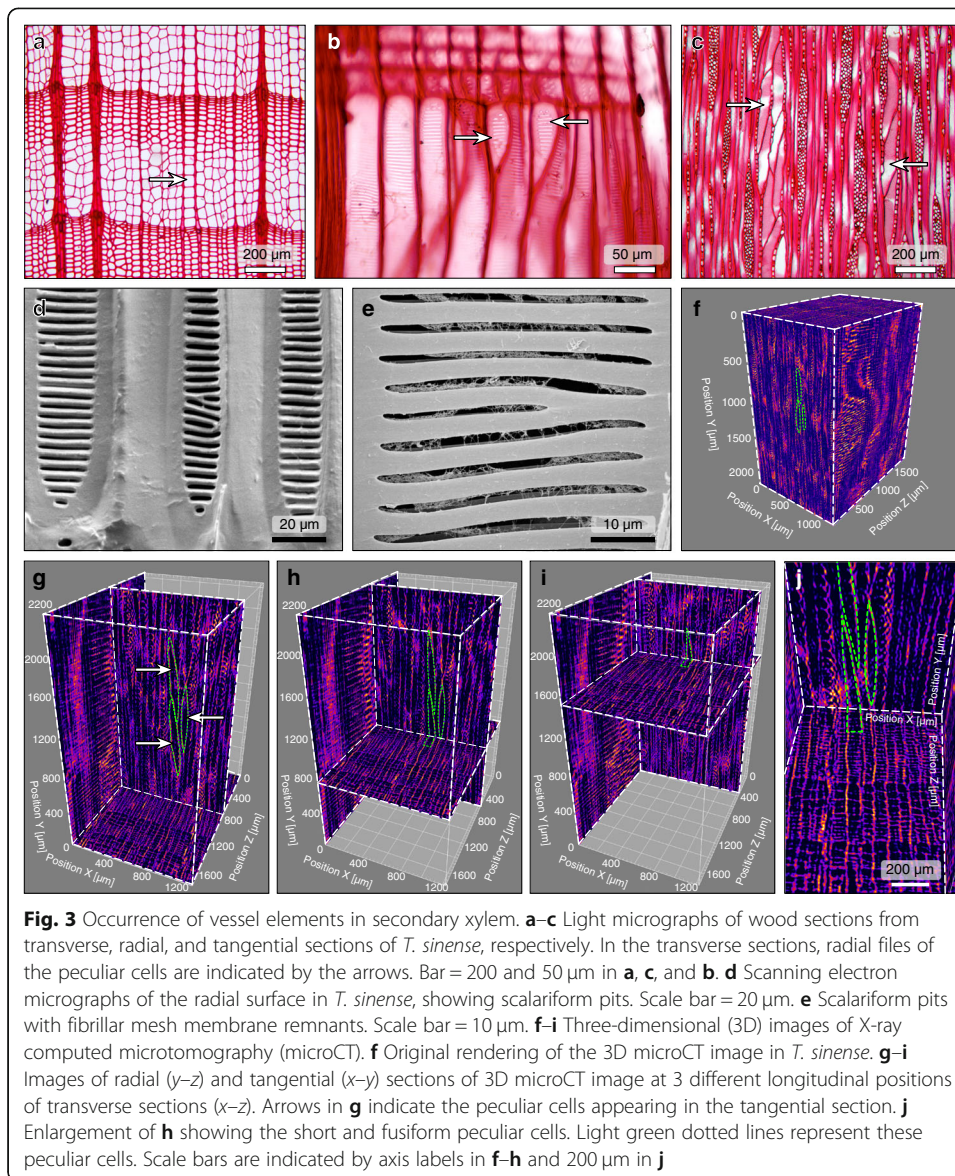
of tracheary elements. Most of the cells appeared as fibrous with 3.0–4.8 mm in length while the peculiar cells were much shorter ranging from 308 to 597 μm and present as fusiform cells with many pits (Fig. 3c). Under scanning electron microscope (SEM), we observed the normal cells often have scalariform bordered pits on radial walls, sometimes with pit membrane remnants that are like a fibrillar mesh as reported previously (Fig. 3d, e). Using X-ray computed microtomography (microCT), we made a 3-dimensional reconstruction of different cells based on 1032 serial slice images and found the short and fusiform peculiar cells are sporadically present in *T. sinense* wood (Fig. 3f–j). Therefore, we considered the larger and shorter peculiar cells as vessel elements, supporting the idea that vessel elements are present in *T. sinense*, although its vessel elements are primitive [19, 20].

The evolution of *VNS*, *LBD*, *R2R3-MYB*, *CesA*, and *PLCP* gene families which are involved in the formation of vessel elements

AtVND7, *AtLBD18/AtLBD30*, *AtMYB46/AtMYB83*, *AtCESA4* and *AtCESA8*, and *AtXCP1/AtXCP2*, belonging to the *VNS*, *LBD*, *R2R3-MYB*, *CesA*, and *PLCP* gene families [25, 45–48], respectively, are key genes involved in the development of vessel elements in *A. thaliana* [21, 49]. To identify the *T. sinense* genes orthologous to these genes, we explored the evolutionary history of *VNS*, *LBD*, *R2R3-MYB*, *CesA*, and *PLCP* gene families.

The *VNS* gene family contains *VND*, *NST*, and *SMB* subfamilies [45]. As shown in the phylogenetic tree, genes homologous to *VND*, *NST*, and *SMB* each fell into different clades with high bootstrap values and each clade did not contain sequences from *Physcomitrella patens* (also known as *Physcomitrium patens*) and *Selaginella moellendorffii*, suggesting that the origin of these subfamilies occurred after the divergence of lycophytes. In the phylogenetic tree (Fig. 4a; Additional file 1: Fig. S11a), genes from the *VND* group clustered into two large clades: one including *A. thaliana* *VND4–VND6* and the other including *VND1–VND3* and *VND7*. Each of these clades contains sequences from gymnosperms (in black in the tree, Fig. 4a) and angiosperms, suggesting that the duplication leading to the ancestors of *VND4–VND6*, and *VND1–VND3* and *VND7* occurred before the divergence of gymnosperms and angiosperms. In the clade containing *VND1–VND3* and *VND7*, genes homologous to *VND1–VND3* and *VND7* each formed a subclade and each subclade only contained sequences from angiosperms, suggesting that the duplication leading to the origin of the *VND7* ancestor occurred after the divergence of angiosperms and gymnosperms. Sequences from the *VND7* clade were aligned and two diagnostic domains of *VND7* subfamily were found in all these sequences (Additional file 1: Fig. S11b), suggesting that they are bona fide *VND7* genes.

The transcription factors *ASL19/LBD30* and *ASL20/LBD18* are involved in a positive feedback loop to regulate *VND7* expression [27]. We identified 27 *LBD18/LBD30* genes in angiosperms. However, no *LBD18/LBD30* genes were identified in genomes of three gymnosperm species, one lycophyte, and one moss, suggesting that the origin of the *LBD18* subfamily may have occurred after the divergence of angiosperms and gymnosperms. In the ML tree (Additional file 1: Fig. S12a), the two genes (*AtLBD18* and *AtLBD30*) from *A. thaliana* formed one clade, suggesting that they originated from



lineage-specific duplications. The two copies in *T. sinense* (Tessin01G0033700 and Tessin21G0059700) are putative orthologs to *A. thaliana* *AtLBD18* and *AtLBD30*.

In total, 87 *R2R3-MYB* genes were identified in the *T. sinense* genome. These sequences together with sequences from *A. thaliana* and *O. sativa* were subjected to phylogenetic analysis. As shown in the ML tree (Additional file 1: Fig. S12b), sequences from the three species fell into different clades. All subgroups (except for 2 subgroups only including sequences from *A. thaliana*) contain sequences from *A. thaliana*/*T. sinense* and *O. sativa*, suggesting these subgroups originated before the splitting of monocots and dicots. In subgroup MYB46 and MYB83, there are three *T. sinense* sequences (Tessin13G0114100, Tessin11G0137400, Tessin07G144300).

In the ML tree (Additional file 1: Fig. S12c), the *CesA* sequences of *P. patens* formed their own clade. *CesAs* in seed plants fell into six major clades, which were designated as six subfamilies: *CesA1/10*, *CesA3*, *CesA6*, *CesA4*, *CesA7*, and *CesA8*. Each subfamily

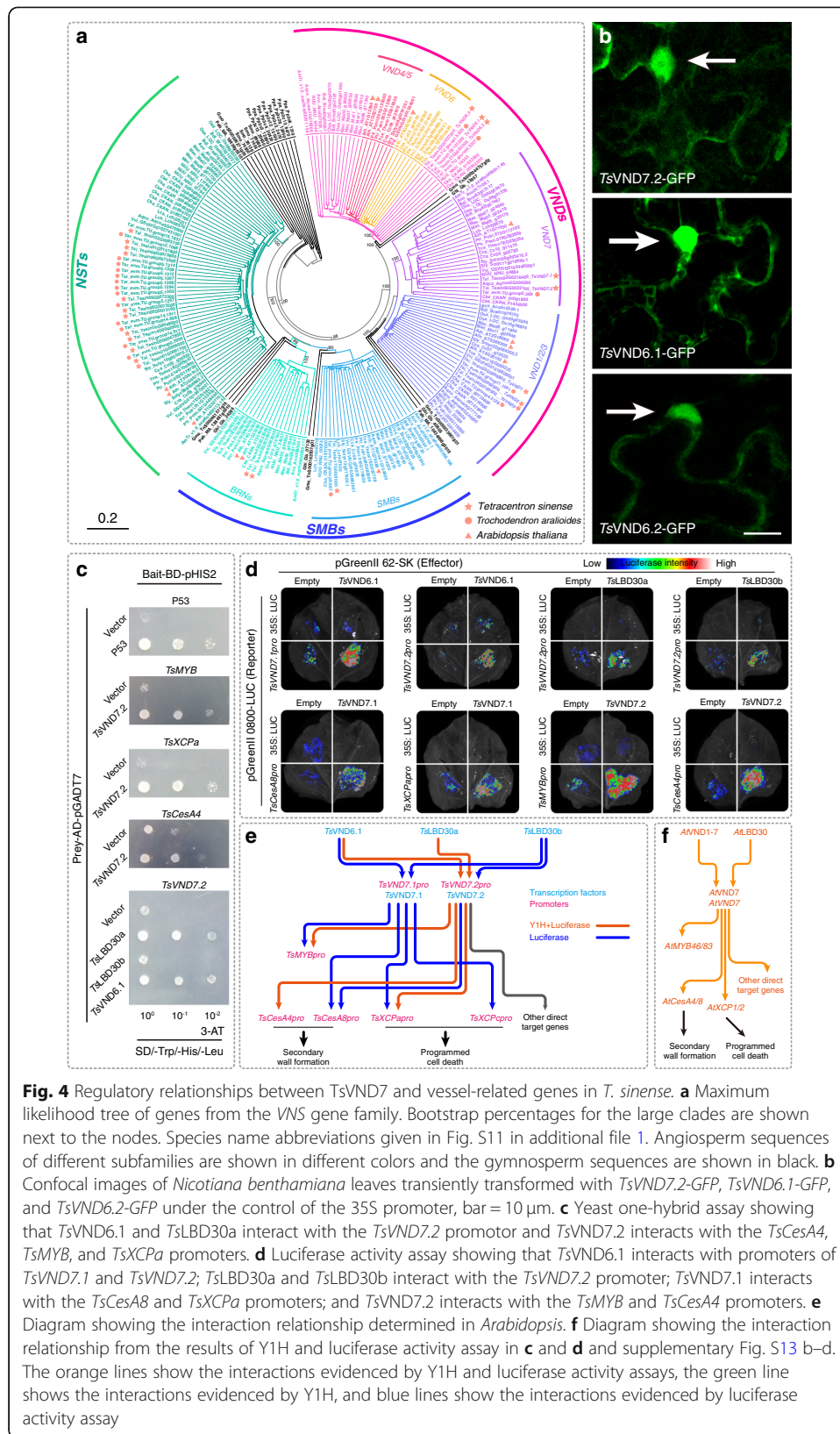


Fig. 4 Regulatory relationships between *TsVND7* and vessel-related genes in *T. sinense*. **a** Maximum likelihood tree of genes from the VNS gene family. Bootstrap percentages for the large clades are shown next to the nodes. Species name abbreviations given in Fig. S11 in additional file 1. Angiosperm sequences of different subfamilies are shown in different colors and the gymnosperm sequences are shown in black. **b** Confocal images of *Nicotiana benthamiana* leaves transiently transformed with *TsVND7.2-GFP*, *TsVND6.1-GFP*, and *TsVND6.2-GFP* under the control of the 35S promoter, bar = 10 μm. **c** Yeast one-hybrid assay showing that *TsVND6.1* and *TsLBD30a* interact with the *TsVND7.2* promoter and *TsVND7.2* interacts with the *TsCesA4*, *TsMYB*, and *TsXCPa* promoters. **d** Luciferase activity assay showing that *TsVND6.1* interacts with promoters of *TsVND7.1* and *TsVND7.2*; *TsLBD30a* and *TsLBD30b* interact with the *TsVND7.2* promoter; *TsVND7.1* interacts with the *TsCesA8* and *TsXCPa* promoters; and *TsVND7.2* interacts with the *TsMYB* and *TsCesA4* promoters. **e** Diagram showing the interaction relationship determined in *Arabidopsis*. **f** Diagram showing the interaction relationship from the results of Y1H and luciferase activity assay in **c** and **d** and supplementary Fig. S13 b–d. The orange lines show the interactions evidenced by Y1H and luciferase activity assays, the green line shows the interactions evidenced by Y1H, and blue lines show the interactions evidenced by luciferase activity assay

included sequences from angiosperms and gymnosperms, indicating the occurrence of ancient gene duplication events before the divergence of angiosperms and gymnosperms. In the phylogenetic tree, subfamilies *CesA4*, *CesA7*, and *CesA8* clustered together and the remaining subfamilies clustered together. *CesA* sequences of *S. moellendorffii* fell into each of these two clades. All subfamilies, including subfamilies *CesA4* and *CesA8*, also contained only one *T. sinense* gene except for subfamily *CesA1/10*, which contained three *T. sinense* genes.

Congruent with previous studies [48], most *PLCP* genes fell into nine subfamilies (*SAG12*, *THI*, *CEP*, *XCP*, *RD21*, *XBCP3*, *RD19*, *ALP*, and *CTB*). *SAG12*, *THI*, *CEP*, *XCP*, *RD21*, and *XBCP3* have been classified as L-like cathepsins, and *RD19*, *ALP*, and *CTB* have been classified as F, H, and B-like cathepsins [48], respectively. In the phylogenetic tree (Additional file 1: Fig. S12d, e), all subfamilies of L-like cathepsins clustered together while F, H, and B-like cathepsins were relatively distant from them. L, F, H, and B-like proteins appeared in the moss and lycophyte, suggesting that ancient duplications led to the origin of these subfamilies. As shown in the ML tree, *XCP* genes originated before the divergence of gymnosperms from *S. moellendorffii*.

Nuclear localization and transcriptional regulatory network of *VND* genes

To investigate the intrinsic function of *VND*s in *T. sinense*, we first analyzed their subcellular localization by transient expression. Like *Arabidopsis VND6* and *VND7*, *TsVND7.2*-GFP, *TsVND6.1*-GFP, and *TsVND6.2*-GFP predominantly localize to the nucleus, as seen by co-localization assays with GFP fluorescence merged with a bright-field image (Fig. 4b; Additional file 1: Fig. S13a).

Yeast one-hybrid (Y1H) analysis was subsequently performed to determine the interaction between *TsVND7.1* and *TsVND7.2* and other genes. The yeast-carrying Bait-BD-pHIS2 vectors expressing the promoter of *TsVND7.1*, *TsVND7.2*, *TsMYB*, *TsCesA4*, *TsCesA8*, *TsXCPa*, or *TsXCPc* grew normally on selective medium without 3-AT but failed to grow with 3-AT, showing that no self-activation occurred (Additional file 1: Fig. S13b). The following tests were carried out by promoters fused to Bait-BD-pHIS2 co-transformed with transcription factors fused to Prey-AD-pGADT7. The results showed that *TsVND7.2* interacted with the promoters of *TsMYB*, *TsCesA4* and *TsXCPa*, and *TsVND7.2* can be regulated by *TsVND6.1* and *TsLBD30a* (Fig. 4c). Conversely, *TsVND7.1* did not interact with the *TsMYB*, *TsCesA4*, *TsCesA8*, *TsXCPa*, *TsXCPb*, or *TsXCPc* promoters and was not regulated by *TsLBD30a*. *TsVND7.2* also did not bind to the promoters of *TsCesA8*, *TsXCPb*, and *TsXCPc* (Additional file 1: Fig. S13c).

The interaction between *TsVND7.1/TsVND7.2* and other genes was further assessed using luciferase activity assays (Fig. 4d; Additional file 1: Fig. S13d). In these assays, *TsVND7.2* interacted with promoters of *TsMYB*, *TsCesA4*, *TsCesA8*, *TsXCPa* and *TsXCPc*, and *TsVND7.2* can be regulated by *TsVND6.1*, *TsLBD30a*, and *TsLBD30a*. *TsVND7.1* also showed similar interactions with other genes except *TsLBD30a*, *TsCesA4*, and *TsXCPc*.

Demographic history analysis of *T. sinense*

To explore the historical demographic fluctuations and present-day genetic diversity within this endemic species, we resequenced 55 individuals of *T. sinense* from a range-

wide sampling from China (Fig. 5a). ADMIXTURE analysis [50] of Chinese samples revealed that the deepest splits occurred among the southwestern population (population YN), the populations near the Sichuan Basin in central China (represented by populations CQ from the Jinfo Mountains and SC from the Emei Mountains), and the populations from the North and East parts of central China (represented by populations SNJ from the Shennong Mountains and populations SX and TS from the Qinling Mountains) (Fig. 5b; Additional file 1: Fig. S14). When $K = 4$, the TS population from the Qinling Mountains in north-central China diverged from the third clusters and four ancestries of *T. sinense* occurred in China (Fig. 5b). Principal component analysis (PCA) and phylogenetic analysis also revealed four major clusters (Fig. 5c, d; Additional file 1: Fig. S14b, c).

T. sinense exhibited higher within-species genetic diversity than within-population level genetic diversity. The within-population genetic diversity of *T. sinense* ($6.30\text{--}11.58\text{e-}3$ for Watterson's estimator (θ_w) and $6.48\text{--}11.25\text{e-}3$ for nucleotide diversity (π), Additional file 1: Table S3) was obviously low compared to the species level diversity. The genetic differentiation statistics (fixation index; F_{st}) further showed that genetic differentiation between YN population and all other populations ($0.293\text{--}0.382$) was higher than those between others (most < 0.293), and the genetic differentiation between SX and SNJ population (0.102) was the lowest (Additional file 1: Table S4).

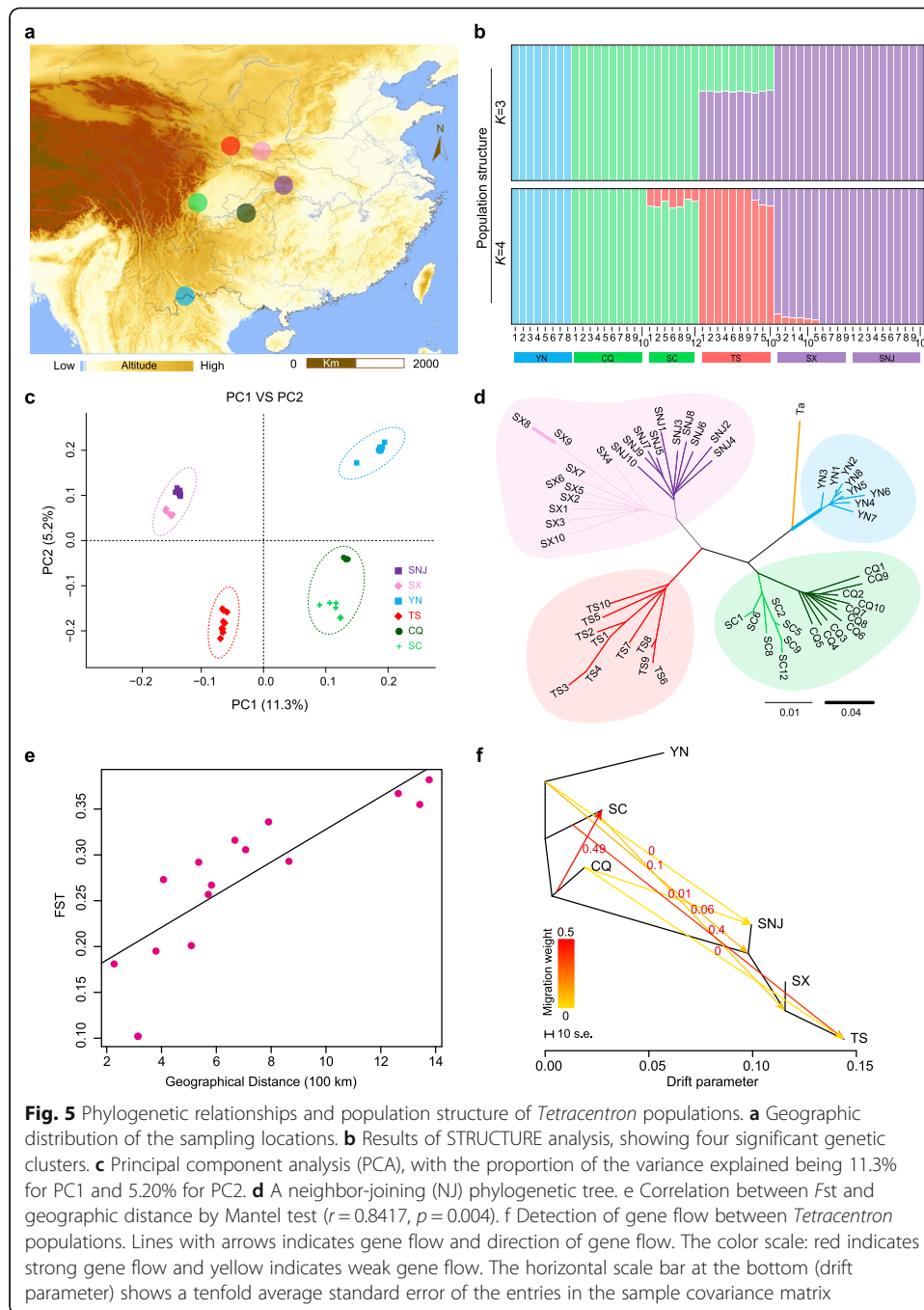
The Mantel test revealed a significant positive correlation between the genetic and geographic distance ($r = 0.8417$, $p = 0.004$, Fig. 5e). To estimate gene flow between populations, we performed Treemix v1.13. analysis [51], which suggested that the TS population from the Qinling Mountains in north-central China is the result of admixture between the SC population the near Sichuan Basin and the SX population from the Qinling Mountains in the north and east parts of central China (Fig. 5f). Gene flow is also observed in population SC and CQ. It is consistent with that they were in the same group according to phylogenetic, PCA, and structure analysis.

An analysis of demographic history using stairway plots revealed bottlenecks the populations may have undergone (Fig. 6), including one ranging $0.3\text{--}0.67$ MYA, and a recent bottleneck around 0.03 MYA. For three populations (SC, SX, and YN), their effective population size is declining since they were affected by the Last Glacial Maximum and their extinction is underway.

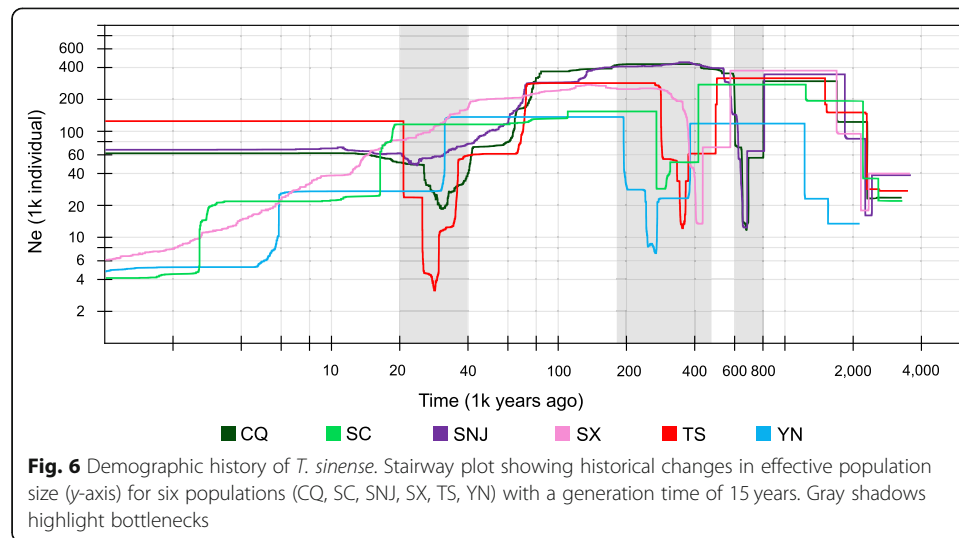
Discussion

Polyploidization in the early evolution of eudicots

Previous studies suggested that Ranunculales, Proteales, Trochodendrales, and Buxales formed the grade of early diverging eudicots [6, 10, 11]. However, these analyses mainly relied on chloroplast and mitochondrial genes as well as a few nuclear ribosomal genes. It is particularly necessary to compare organellar gene-based phylogenetic estimates with topologies derived from analyses of low-copy nuclear genes [1]. The *T. sinense* genome provides an opportunity to resolve the relationships between lineages based on genome-wide low-copy nuclear genes. Our analysis based on 214 single-copy nuclear gene families (orthogroups) provides strong support for the conclusion that the Ranunculales, Proteales, and Trochodendrales are as successive sisters to the core eudicots. Within eudicots, the Ranunculales diverged at ~ 136 ($129\text{--}140$) MYA, followed by the



Proteales ~ 131 (123–137) MYA, the Trochodendrales ~ 125 (117–132) MYA, and the core eudicots ~ 113 (105–123) MYA (Fig. 1). This estimate is generally consistent with several estimates from other relaxed molecular clock analyses based on DNA and/or genomic data [36, 52]. The narrow range of ages suggests a rapid diversification of the early eudicots. In Trochodendrales, the divergence time of *Tetracentron*–*Trochodendron* is approximately 31 MYA (Fig. 1; Additional file 1: Fig. S3). This is generally congruent with the age of the earliest reliable fossils of *Tetracentron* and *Trochodendron*, which date to the early middle Eocene epoch [30, 53].



Prevalent polyploidization events have been identified during the evolutionary history of angiosperms [54]. Previous studies demonstrated that the rise and eventual dominance of core eudicots might have benefited from the γ hexaploidization event in the most recent common ancestor of core eudicots [4]. Based on genomic investigations of *V. vinifera* [55] and *N. nucifera*, Ming et al. [13] proposed that the triplication event involved an initial tetraploidization event after the divergence of early diverging eudicots, then involved hybridization with a now-extinct species that branched off around the same time as, or earlier than *N. nucifera*. By contrast, a recent study using the *A. coerulea* genome [5] also proposed a hybrid origin of core eudicots, involving a tetraploidization event that occurred in the ancestor of all eudicots, then a hexaploidization in the ancestor of core eudicots. In this study, we performed comprehensive analyses including synteny, K_s , and phylogenomic analysis to clarify the nature of the γ event. Our results strongly suggested that the WGDs that occurred in *A. coerulea*, *N. nucifera*, and the two species of Trochodendrales are lineage-specific, and no WGD occurred in the common ancestor of all eudicots (Fig. 2g). The ancient WGD (α) in Trochodendrales dates to about 59–54 MYA, which coincides with the Cretaceous–Tertiary (KT) mass extinction event [56]. In fact, many independent WGDs have been identified in many lineages and have been proposed to have contributed to survival during this period of dramatic environmental changes [54, 56].

It is also worth noting that after the two successive WGDs in *T. sinense*, the genome structure remained well conserved. K_s analyses for orthologs between *T. sinense*/*T. aralioides* and *V. vinifera* showed the Trochodendrales lineage has the lowest nucleotide substitution rate among the studied species (Fig. 2e). The lineage nucleotide substitution rate in *N. nucifera* is about 30% lower than that of *V. vinifera* [13]. We found that the nucleotide substitution rate of *T. sinense* is much lower than *N. nucifera*, since the peak of the K_s distribution of orthologs between *T. sinense* and *A. coerulea* is smaller than that of *N. nucifera* and *A. coerulea* (representing the same divergence event) (Fig. 2e). Together, these results indicate that the genome of *T. sinense* could serve as an exceptional comparative reference genome for investigating early evolution of eudicots.

Genomic analysis of gene families can shed light on the morphological characteristics shared by different clades. The early diverging lineages of eudicots tend to show some ancestral characteristics that differentiate them from the core eudicots. For example, the flowers of core eudicots are organized in a predictable manner with a stable number of parts (e.g., flowers with parts in five or multiples of five, a clear differentiation of sepals and petals), but in the Trochodendrales (like in some early angiosperms), the perianth is not differentiated into typical sepals and petals [8]. MADS-box (ABCE) genes are among the most important regulators of flower development. Analysis of the *Amborella* genome revealed that duplication and diversification of floral MADS-box genes likely occurred before the origin of extant angiosperms [40]. Our results showed that A-, B-, C-, E-function genes did not significantly expand or contract in the lineages leading to eudicots and core eudicots. The differences in flower structure between eudicots and core eudicots may be due to the differences between broadly overlapping ABCE gene expression patterns in early diverging eudicots and angiosperms, and strict spatially restricted ABCE gene expression in core eudicots, as suggested by some studies in Nymphaeales, Magnoliids, and Ranunculales [41, 57, 58]. Notably, in the lineage leading to core eudicots, the TM3 MADS-box gene family that includes the important flowering time gene *SOCI* [59] has expanded. In the genome of *T. sinense*, the *SVP* gene family also expanded rapidly. This family includes the MAD-box gene *AGL24*, which is not only involved in the flowering transition through activation of *SOCI* [59, 60] but also in flower development together with class A and E genes [61, 62]. The expanded *SVP* clades could be involved in the complex networks regulating flowering in *T. sinense*.

Vessel elements in secondary xylem of *T. sinense* and the transcriptional regulation of VND7-related genes

The presence of vessels in the xylem of angiosperms and their absence in the xylem of gymnosperms (except for gnetophytes [15]) is one of the striking differences between these two large groups of plants. The Tetracentraceae (also as the Trochodendraceae), however, is one of several angiosperm families that have been described as lacking vessels. Carlquist [63] found porous pit membranes on the end walls of early wood tracheary elements and considered that tracheary elements in *T. sinense* are transitional. According to the study of Ren et al., vessel elements exist in *T. sinense*, but they are of a primitive type [19, 20]. Our light microscopy and SEM observations of slices of *T. sinense* wood and analyses of X-ray computed microtomography (microCT) (Fig. 3) provide strong evidence that there are two types of tracheary elements and some of the peculiar fusiform tracheary elements with obvious end-wall perforation plates in *T. sinense* are vessels.

VND7 belongs to the *VND* subfamily of the *VNS* gene family and is thought to be the master regulator of vessel differentiation [21], as *VND7* can induce various cells to trans-differentiate into vessel elements in *Arabidopsis* and other species. Gene family analyses showed that *VND7* occurred in angiosperms including in *T. sinense* and *T. aralioides*, but not in gymnosperms, indicating that the duplication event leading to the origin of the *VND7* ancestor occurred after the divergence of angiosperms and gymnosperms and before the split of angiosperm lineages (Fig. 4a; Additional file 1: Fig. S11a).

This is consistent with a previous study that found no orthologs of *VND7* or *VND1–3* in gymnosperms [15]. The gnetophyte *Gnetum montanum* contains vessel-like water-conducting cells. However, the lack of *VND7* in *G. montanum* suggested that different molecular mechanisms underpin the origin and development of vessels in *Gnetum* and angiosperms, and the vessels in *Gnetum* and angiosperms are convergent characters [15]. By contrast, based on the phylogenetic analysis of the *VNS* gene family, we identified *TsVND7.1* and *TsVND7.2* of *T. sinense* as orthologs of *A. thaliana VND7*. Multiple sequence alignments showed that *TsVND7.1* and *TsVND7.2* have the same seven conserved subdomains as other VNDs and the conserved sites important for the functions of VNDs (Additional file 1: Fig. S11b). A transient expression assay showed that VND-GFP fluorescence localized to the nuclei of plant cells (Fig. 4b; Additional file 1: Fig. S13a) implying that VNDs in *T. sinense* act as nuclear transcription factors, as in *A. thaliana*.

Xylem vessel formation is regulated by a complex transcriptional network. *VND7* transcription factors directly regulate many genes including *MYB46/MYB83*, *CESA4* and *CESA8*, and *XCP1/XCP2* [21–26]. These genes are involved in vessel element differentiation, specifically in secondary cell wall formation and PCD. *VND7* is also positively regulated by several genes including *LBD18/LBD30* and *VND1–VND5* [14, 27, 28]. Gene family analyses suggested that all subfamilies that these genes belong to originated before the divergence of the angiosperms. We found that two genes, *TsLBD30a* (Tessin01G0033700) and *TsLBD3b* (Tessin21G0059700), are orthologs of the *AtLBD18/30* gene pair; three *TsMYBs* (Tessin13G0114100, Tessin11G0137400 and Tessin07G0144300) were found to be orthologous to *MYB46/MYB83*; *TsCesA4* (Tessin16G0055000) and *TsCesA8* (Tessin01G0177100) are orthologs of *AtCesA4* and *AtCesA8*, respectively; and *TsXCPa/b/c* (Tessin02G0006000, Tessin24G0110100, and Tessin06G0166500) are orthologous to *AtXCP1* and *AtXCP2* (Additional file 1: Supplementary Notes and Fig. S12). To ascertain whether these genes can regulate or are regulated by *TsVND7.1* and *TsVND7.2*, we performed yeast one-hybrid analysis and luciferase activity assay. Yeast one-hybrid analysis showed that *TsVND6.1* and *TsLBD30a* proteins can bind to the *TsVND7.2* promoter and *TsVND7.2* can interact with the promoters of *TsMYB*, *TsCESA4*, and *TsXCPa*. Luciferase activity assays not only confirmed the interactions between *TsVND7.2* and relevant genes as shown by yeast one-hybrid, but also identified the interaction between *TsVND7.2* and *TsLAB30b/TsCesA8* (Fig. 4e). Although the two techniques showed some differences, the main results of both techniques demonstrated that the regulatory relationships between *TsVND7* and vessel-related genes are similar to those in *A. thaliana* (Fig. 4f), supporting the anatomical observations that vessel elements are present in *T. sinense*.

Demographic history of *T. sinense*

T. sinense was widespread in many parts of the Northern Hemisphere until the Pliocene epoch of the Quaternary period [30, 64–67], indicating that the present-day endemism of *T. sinense* in east Asian is refugial. Previous studies identified five or three refugia in China [31, 32]. Our ADMIXTURE analysis of whole genomic SNPs of Chinese samples revealed four major groups (Fig. 5b; Additional file 1: Fig. S14). In good agreement with the ADMIXTURE analysis, principal component analysis (PCA) and

phylogenetic analysis also revealed the same four major clusters (Fig. 5c, d; Additional file 1: Fig. S14). Hence, our results suggested that in addition to the Southwestern China refuge [31], there were three refugia in central China: one in the Sichuan Basin corresponding to the area covered by the Jinfo and Emei Mountains, one in the east part including the Shennong Mountains, and one in the Qinling Mountains. These areas have long been regarded as important global Pleistocene glacial refugia and the areas around the Sichuan Basin of central China and the Shennong Mountains in the east of central China were also likely refugia for many other relict tree species such as *Cathaya*, *Cercidiphyllum*, *Ginkgo*, and *Taxus* [68–70].

The genetic diversity of species is the result of historical evolution and also the basis of evolution; moreover, sufficient genetic variability is required for endangered species to survive and adapt to changing environments. Nucleotide diversity analysis of *T. sinense* showed high within-species genetic diversity (Additional file 1: Table S3) and its genetic diversity is higher than those reported for the relict species *A. trichopoda* [40] and *Ginkgo biloba* [69]. The high genetic diversity detected in this tertiary relict species might reflect the accumulation of mutations over long evolutionary time scales and in populations experiencing geographical isolation and heterogeneous habitats. This high level of genetic diversity indicated that this endangered species maintains a relatively high evolutionary potential. Congruent with the previous study [32], we observed a low genetic diversity within populations and high genetic differentiation among populations of *T. sinense* (Additional file 1: Table S4). The small population effect and geographical isolation were proposed to the main factors [32]. A population graph inferred by Treemix indicated similar relationships between populations to those revealed in the NJ tree and low levels of gene flow except between population SC and TS (0.4). Gene flow was fewer between distant populations, as confirmed by the Mantel test between the genetic and geographic distance ($r = 0.8417$, $p = 0.004$).

An analysis of demographic history using the stairway plot method revealed that all populations have undergone multiple cycles of expansion and contraction. During the early Quaternary, all populations underwent contraction, followed by expansion. For all populations, two significant bottlenecks were identified (Fig. 6). The time of the more ancient bottleneck identified for two populations (CQ and SC) corresponds to the Nanyangxungla glaciation (0.5–0.72 MYA, the most extensive glaciation in China [71]) and the time of the more ancient bottleneck of the other four population corresponds to the Guixiang glaciation (0.13–0.3 MYA [71]). The most recent bottleneck of all six populations occurred around 0.03 MYA, corresponding to the Last Glacial Maximum (0.001–0.007 MYA). For three populations (SC, SX, and YN), their effective population size is declining since they were affected by the Last Glacial Maximum and their extinction is likely underway. As ongoing effects of an expanding human population threaten *T. sinense* populations, conservation efforts should focus on preserving the genetic diversity of endemic *T. sinense* trees.

Conclusions

Tetracentron sinense belongs to the angiosperm order Trochodendrales, which consists of only one family (Trochodendraceae) and two species (*Tetracentron sinense* and *Trochodendron aralioides*). Based on the *Tetracentron sinense* reference genome and the previously published genomes from other species, our phylogenomic analyses support

that the Trochodendrales is the sister lineage to the core eudicots. The genomes of the two species of Trochodendrales also revealed two WGD events (α and β) in Trochodendrales at approximately 82 and 59 MYA. Based on our comprehensive analysis including syntenic, *Ks* distribution and phylogenomic analysis, we conclude that the Trochodendrales did not experience the γ event and the event may have only happened in core eudicots. Through light microscopy and SEM of slices of *T. sinense* wood and microCT, we found that vessel elements are present in *T. sinense*. Consistent with the anatomical results, the *T. sinense* genome contains two putative orthologs of *VND7*, a master regulator of angiosperm vessel formation, which appears to function in a similar transcriptional network in *T. sinense* and *Arabidopsis*. Population genomics analysis of individuals across the native range of *T. sinense* in China revealed that all the populations have undergone two bottlenecks and there is a high level of genetic diversity in the species. As *Tetracentron sinense* is a basal eudicot with unique tracheary elements, its genome provides a unique reference for inferring the early evolution of the eudicots and the mechanisms underlying the formation of vessel elements. The population genomics analysis facilitates the interpretation of their genetic diversity and geographic structure, thus providing genetic implications for future conservation.

Methods

Illumina short-read sequencing

The sequenced individual of *Tetracentron sinense* is growing in the Kunming Botanical Garden, Kunming, China. Genomic DNA was extracted from fresh leaves and purified using the Tiangen Isolation/Extraction/Purification Kit (Tiangen Biotech (Beijing) Co., Ltd.). Then, it was fragmented with a Covaris M220 Focused-ultrasonicator Instrument. Illumina PCR-free libraries with insert sizes of 300–500 bp were constructed using the NEBNext Ultra DNA library Pre Kit for Illumina sequencing. Then, 150-bp paired-end sequencing was performed using the Illumina HiSeq X Ten platform.

Genome survey

Using Illumina short reads, *k*-mer distribution was estimated using JELLYFISH v2 [72]. The overall characteristics of the genome such as genome size, repeat contents, and heterozygosity rate were estimated using GCE software [73].

PacBio library construction and sequencing

SMRTbell libraries were constructed according to the manufacturer's protocol of SMRTbell Express Template Preparation Kit (Pacific Biosciences). Briefly, high-quality and high-molecular-weight genomic DNA was purified using the Mobio PowerClean Pro DNA Clean-Up Kit, and DNA quality was assessed by agarose gel electrophoresis and Thermo Fisher Scientific Qubit Fluorometry. Genomic DNA was further sheared to a size range of 15–50 kb using AMPure beads (Beckman Coulter) or g-TUBE (Covaris) and enzymatically repaired and converted into SMRTbell template libraries. Then, hairpin adapters were ligated after exonuclease digestion. The resulting SMRTbell templates were size-selected by Blue Pippin electrophoresis (Sage Sciences). The size-selected SMRTbell DNA fragments ranging from 15 to 50 kb were sequenced on a

PacBio Sequel instrument using S/P2-C2 sequencing chemistry (21 cells, 14.7 million reads, ~100×).

Hi-C library construction and sequencing

Hi-C was performed according to the following procedure. Young leaves were fixed with formaldehyde and lysed, and then the cross-linked DNA was digested with the HindIII restriction enzyme and the 5' overhangs were biotinylated. After labeling with biotin-14-dCTP, the resulting free blunt ends were ligated again. Then, the purified DNA was treated to remove biotin from non-ligated DNA ends. For the fragmentation, DNA was then sheared with a Covaris M220 Focused-ultrasonicator Instrument. The sheared DNA was then repaired and biotin-containing fragments were isolated using streptavidin beads. A-tailing and adapters were ligated subsequently, and sequencing libraries were generated and sequenced on an Illumina HiSeq X Ten instrument with 150-bp paired-end reads.

Genome assembly and pseudomolecule building

The primary assembly was performed with PacBio long reads using six different approaches, CANU (v0.1) [74], SMARTdenovo (v 0.2) (<https://github.com/ruanjue/smartdenovo>), WTDBG (v0.3) (<https://github.com/ruanjue/wtdbg>), SMARTdenovo after CANU correction (v0.4), and WTDBG2 after CANU correction (v0.6). Then, based on the assembled genome size, number of contigs, average contig size, and N50, assemblies from SMARTdenovo (v 0.2), and WTDBG2 after CANU correction (v0.6) were selected for merging using quickmerge [75] to improve the contiguity. Finally, the draft assembly was polished using PacBio long reads with Arrow software (<https://github.com/PacificBiosciences/GenomicConsensus>) and corrected using Illumina paired-end reads with Pilon software [76].

Hi-C read pairs were aligned to the draft assembly v1.0 using Juicer software [77]. The resulting contact matrices and draft assembly v1.0 were used for further Hi-C scaffolding with the 3D (3d-dna) pipeline [78]. This procedure includes correcting the mis-joins, scaffolding the input contigs, polishing the megascaffold, splitting the megascaffold into raw chromosomal scaffolds, and then sealing raw chromosomal scaffolds to yield chromosome-length scaffolds. After the first round of 3d-dna, we manually adjusted the assembly with Juicebox⁵⁴. Furthermore, to improve the chromosome-scale assembly, we performed 3d-dna within each chromosome [78]. The chromosome-length assembly was further improved by manual adjustment with Juicebox [79], including correcting mis-joins, changing the order and orientation of draft scaffolds with translocation and inversion errors, and adjusting chromosome boundaries (Fig. S2). Finally, we performed LR-Gapcloser [33] with PacBio long reads to close gaps (in the assembly) two times, and pilon [76] with Illumina paired-end short reads five times to improve the assembly. Redundans was used to remove redundancy in the unplaced contigs [80].

To evaluate the quality of the assembled genome, we mapped the Illumina paired-end reads, PacBio long reads, and RNA-seq reads to the genome using BWA-MEM [81], Minimap2 [82], and hisat2, respectively. We also used BUSCO to examine the

completeness of gene content with the Embryophyta odb9 database and default parameters [34]. The LAI program was used to assess the quality of the assembly [83].

RNA library construction and sequencing

We performed RNA sequencing (RNA-seq) to improve the prediction of gene models. Different plant tissues from *T. sinense*, including bud, stem, leaf, fruit, phloem, and xylem, were collected and frozen in liquid nitrogen. Total RNA was extracted with the RNAPrep Pure Plant Kit (Tiangen Biotech Co., Ltd.) according to the manufacturer's instructions. Transcriptome libraries were constructed using NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs Inc.) following the manufacturer's procedure. Then, 150-bp paired-end sequencing was performed using the Illumina HiSeq X Ten platform.

Transcriptome assembly

Three approaches were used to assemble the transcriptome of *T. sinense*. First, RNA-seq reads from six selected tissues were de novo assembled into transcripts with Trinity [84]. In the second and third approaches, RNA-seq reads were first mapped to the assembled genome of *T. sinense* using hisat2 [85]. Then, they were assembled into transcripts using the genome-guided strategy by Trinity and StringTie [86]. Transcripts obtained from different approaches were then merged and redundant transcripts were removed with a cutoff of 95% in both identity and coverage with CD-HIT [87].

Genome annotation

The library of repeat families was generated from our assembly using RepeatModeler [88]. Then with this repeat library, RepeatMasker was used to identify repetitive elements in the genomic sequences. Protein-coding gene prediction of the *T. sinense* genome was conducted by the MAKER genome annotation pipeline [89], integrating ab initio prediction with de novo assembled transcripts and protein homology data from *Nelumbo nucifera* and *Arabidopsis thaliana*. Gene functional annotation was performed using BLAT by searching against the Swiss-Prot, TrEMBL, NR, Pfam and eggNOG databases [90], and using InterProScan [91] by searching against protein databases of InterPro. The gene ontology and KEGG information for each gene model was extracted from Swiss-Prot, TrEMBL.

The tRNA genes were identified by tRNAscan-SE [92], rRNA genes were identified by RNAMMER [93], and non-coding RNAs (ncRNAs) genes were identified by searching against the Rfam database [94].

Phylogenetic analysis and estimation of the divergence time of species

OrthoFinder v2.3.1 [35] was used to generate gene family (orthogroups) classifications from 19 species including two basal angiosperms (*Amborella trichopoda* and *Nymphaea colorata*), three magnoliids (*Liriodendron chinense*, *Cinnamomum kanehirae*, and *Litsea cubeba*), four monocots (*Musa acuminata*, *Ananas comosus*, *Oryza sativa*, and *Brachypodium distachyon*), four basal eudicots (*Aquilegia coerulea*, *Nelumbo nucifera*, *T. sinense*, and *Trochodendron aralioides*), and six eudicots (*Arabidopsis thaliana*, *Populus trichocarpa*, *Vitis vinifera*, *Theobroma cacao*, *Solanum lycopersicum*, and

Coffea canephora). Orthologous groups (OGs) with single-copy genes were used in the phylogenetic analysis. Amino acid sequences from each OG were first aligned with MAFFT [95]; the alignments were then adjusted manually and ambiguous positions were deleted with TrimAI [96]. Then, the sequence alignments were concatenated. Finally, the concatenated alignment was used to construct a maximum likelihood (ML) phylogenetic tree using IQ-TREE software [97].

We estimated the divergence time of species using MCMCTree [37] and BEAST2 [38]. For the age estimation using MCMCTree, the tree was calibrated with angiosperm crown age (209 million years ago, MYA) [36], eudicot crown age range (127.9–139.4 MYA), and *Theobroma cacao*–*Arabidopsis thaliana* divergence range (87–95 MYA). The latter two calibrated times were obtained from TreeTime [98]. For the age estimation using BEAST [38], the earliest tricolpate pollen (~ 125 MYA) fossil associated with eudicots and the fossil (100.5 MYA) assigned to the Pentapetalae were used as calibration points. The angiosperm crown age of 209 MYA [36] was also used as a calibration point.

Expansion and contractions of gene families were determined using CAFE 2.2 (Computational Analysis of Gene Family Evolution) [39]. The program uses a birth and death process to model gene gain and loss over a phylogeny. Enrichment of KEGG and Gene Ontology terms for *T. sinense* or *T. aralioides* gene families that rapidly evolved (expanded or contracted rapidly) were performed using goseq [99].

Whole-genome duplication analysis

We performed intragenomic comparisons for the two species of Trochodendraceae, *Aquilegia coerulea*, *Nelumbo nucifera*, and *Vitis vinifera*. We also performed intergenomic comparisons between *T. sinense* or *T. aralioides* and several species including two basal angiosperms (*Amborella trichopoda* and *Nymphaea colorata*), one magnoliid (*Cinnamomum kanehirae*), two early diverging eudicots (*A. coerulea*, *N. nucifera*), and one core eudicot (*Vitis vinifera*). OrthoFinder v2.3.1 [35] was used to identify paralogous gene pairs within species and orthologs between species. Based on these paralogous and orthologous gene pairs, collinear blocks were identified using MCsax [100]. Protein sequences of paralogous and orthologous gene pairs in collinear blocks were first aligned by MUSCLE [101]. The protein alignments were then converted to codon alignments by PAL2NAL [102]. Finally, *Ks* values were calculated using the Nei-Gojobori algorithm implemented in the PAML package [103].

To confirm the timing of these WGDs in early diverging eudicots, we further performed integrated syntenic and phylogenomic analyses. We obtained anchor genes from the inter-genomic synteny blocks present in ratios of 1:2:4 in *A. trichopoda*, *N. nucifera*, and *T. sinense*. All genes in the same sets of syntenic blocks were concatenated. These concatenated datasets were used to perform phylogenetic analysis. We then calculated the percentage of trees that supported the idea that the WGDs in *N. nucifera* and *T. sinense* are lineage-specific.

Timing of WGD events

We first calculated the *Ks* value of the orthologs in syntenic regions between *T. sinense* and *T. aralioides* using KaKs_Calculator [104] based on the YN model. Given the

median K_s value (0.2) of *T. sinense* and *T. aralioides* and their divergence date T (30.7 MYA), we calculated the synonymous substitutions per site per year (r) for Trochodendraceae as equaling $3.25e-9$ ($r = K_s/2T$).

We then applied the r value ($3.25e-9$) to estimate the time of the Trochodendraceae WGD events ($T = K_s/2r$). Since the median K_s values of paralogous pairs of *T. sinense* corresponding to the two WGDs were 0.532 and 0.382, we estimated that the two WGDs occurred around 82 and 59 MYA. Based on the median K_s of paralogous pairs of *T. aralioides* corresponding the two WGDs (0.507 and 0.348), we dated the two WGDs to around 78 and 54 MYA. The small difference in time of the two WGDs in the Trochodendraceae inferred from these two species may be due to a difference in their evolution rates.

Light microscopy

Sectioning blocks of *T. sinense* [1 cm (L) \times 1 cm (R) \times 1 cm (T)] were cut with razor blades and then softened in 2% ethylenediamine at 60 °C for 2 days. Fifteen-micrometer-thick transverse, radial, and tangential sections were cut with a sliding microtome. Thereafter, sections were stained sequentially with 1% aqueous safranin solution, rinsed, and finally mounted beneath cover slides. Images were captured by a light microscope (Olympus BX50, Japan).

Examination of secondary xylem with SEM

For scanning electron microscope (SEM) analysis, the stem materials from *T. sinense* were collected from the upper slope of Fenshuiling, Yunnan province, China. Stem samples were boiled in water and sectioned on a sliding microtome, then cut into pieces about 1 cm in length and fixed overnight in 2.5% glutaraldehyde, followed by washing with 0.1 M PBS (pH 7.2) three times. After sequential dehydration in 30%, 60%, 80%, 90%, 95%, and 100% ethyl alcohol (15 min each), the samples were incubated in liquid carbon dioxide at 0 °C and 20 °C sequentially (20 min each), then subjected to critical point drying at 35 °C for 15 min using a Tousimis autosamdri-825 critical point dryer. Vacuum coating was performed using a HITACHI MC 10003 Sputter Coater under 10 Pa and 1000 V for 2 min. The images were taken with an accelerating voltage of 5–10 kV using a Hitachi SU80202 SEM.

MicroCT imaging

Wood blocks of *T. sinense* were imaged by X-ray microtomography using a Bruker SkyScan microCT system 1172 (Skyscan, Belgium) with a 20–100 kV, 10 W X-ray source. The microCT device was set to operate at a voltage of 33 kV and power of 7 W. The sample was scanned at a voxel size resolution of 1.792 μm . The software Skyscan's NRecon program (Skyscan, Belgium) was used to perform the reconstruction of the projection images, resulting in a volume of 3304 \times 3304 \times 1806 μm from 1032 slice images with a resolution of 1888 \times 1888 \times 1086 pixels each. Imaris volume rendering and analysis software (Imaris \times 64 9.0.1, BITPLANE, Oxford Instruments company) was used to render images from the three-dimensional data set.

Gene family analysis

AtVND7, *AtLBD18/AtLBD30*, *AtMYB46/AtMYB83*, *AtCESA4*, *AtCESA8*, and *AtXCP1/AtXCP2* are key genes involved in the development of vessel elements in *A. thaliana* [21, 49]. These genes belong to the *VNS*, *LBD*, *R2R3-MYB*, *CesA*, and *PLCP* gene families [45–48], respectively. To investigate the evolutionary history of these gene families and identify the *T. sinense* genes orthologs, genes of *VNS*, *LBD*, *CesA*, and *PLCP* gene families from 22 representatives of major plant lineages, including 19 species of angiosperms used in the phylogenetic analysis of species, 3 gymnosperms (*Ginkgo biloba*, *Gnetum montanum*, and *Picea abies*), 1 lycophyte (*Selaginella moellendorffii*), and 1 moss (*Physcomitrella patens/ Physcomitrium patens*), were obtained using the *hmmsearch* command in the HMMER package [105] with their specific Pfam models (Extended Data Information). *R2R3-MYB* transcription factor homologs were obtained by a BLASTp search with e-value $1e-5$ (Extended Data Information). Then, identical and defective sequences were identified and manually eliminated in BioEdit. Protein sequences or nucleotide sequences were aligned using MAFFT [95] and the alignment was manually adjusted in BioEdit. Maximum likelihood trees were constructed using IQ-TREE software [97].

Plasmid construction and transient infiltration

To investigate the subcellular localization of the VNDs, we created transgenic plants expressing *AtVND6*, *AtVND7*, *TsVND6.1*, *TsVND6.2*, and *TsVND7.2* fusions with GFP. To this end, the *AtVND6*, *AtVND7*, *TsVND6.1*, *TsVND6.2*, and *TsVND7.2* coding region fragments were amplified from wild-type (WT) complementary DNA (cDNA) of *A. thaliana* and *T. sinense* by using PCR with gene-specific primer sets (Supplementary Table 4). After that, the cDNAs were each subcloned into a modified pCAMBIA 2300-GFP vector under the control of the 35S promoter to generate the *35S::AtVND6-GFP*, *35S::AtVND7-GFP*, *35S::TsVND6.1-GFP*, *35S::TsVND6.2-GFP*, and *35S::TsVND7.2-GFP* constructs. Transient infiltration of *Nicotiana benthamiana* was carried out using an *Agrobacterium tumefaciens* (GV3101). Images were captured using a Zeiss LSM 880 Airyscan (Carl Zeiss, Germany) for confocal microscopy. An argon laser (488-nm line) was used for GFP excitation and the emission was captured at 510–550 nm (green).

Yeast one-hybrid analysis

Yeast one-hybrid analysis was performed to investigate the interactions between *TsVND6.1*, *TsVND7.1*, *TsVND7.2* or *TsLBD30a* transcription factors and the promoters of *TsVND7.2*, *TsMYB*, *TsCesA4*, *TsCesA8*, *TsXCPa*, *TsXCPb*, and *TsXCPC* (about 500 bp in length from the translation start site). All the promoters and transcription factor genes were amplified by PCR using TransStart FastPfu DNA polymerase (TransGenBiotech, Beijing, China). Using BamHI/EcoRI or EcoRI/SacI digestion and T4 ligation reactions, the transcription factor genes were inserted into the AD vector (pGADT7, Takara Clontech Co., Ltd.) and the promoters were inserted into the BD vector (pHIS2, Takara Clontech Co., Ltd.). Then, the vectors were transformed into yeast strain Y187 and the transformed yeast was cultured for 2–3 days on SD/-Trp/-His double deficiency medium. Single clones harboring different vector combinations were transferred to SD/-THL medium containing 10, 20, 30, 40, 50, and 60 mM 3-AT and

incubated at 30 °C for 0.5–8 h, and the color changes in the yeast cells were observed. All primers are shown in Table S5 in Additional file 1.

Luciferase activity assay

To assay the activating activity of *TsVND6.1*, *TsVND7.1*, *TsVND7.2*, *TsLBD30a*, and *TsLBD30b* to the *TsVND7.1*, *TsVND7.2*, *TsMYB*, *TsCesA4*, *TsCesA8*, *TsXCPa*, *TsXCPb*, and *TsXCPC* promoters, the pGreenII 0800-LUC reporter vector and pGreenII 62-SK effector vector (Biovector Science Lab) were used. The *TsVND7.1*, *TsVND7.2*, *TsMYB*, *TsCesA4*, *TsCesA8*, *TsXCPa*, *TsXCPb*, and *TsXCPC* promoters were cloned into pGreenII 0800 at the HindIII and BamHI sites to fuse them with the *LUC* reporter gene (*TsVND7.1*, *TsVND7.2*, *TsMYB*, *TsCesA4*, *TsCesA8*, *TsXCPa*, *TsXCPb*, or *TsXCPC* pro-LUC). The coding sequences for *TsVND6.1*, *TsVND7.1*, *TsVND7.2*, *TsLBD30a*, and *TsLBD30b* were cloned into the pGreenII 62-SK vector at the EcoRI and HindIII sites. The recombinant vectors were co-transformed into *A. tumefaciens* GV3101. After centrifugation, the bacteria were collected and resuspended at $A_{600} = 0.6$ in infection solution (10 mM MES, 10 mM MgCl₂, and 200 μM acetosyringone) for infiltration. The prepared suspensions were infiltrated into *N. benthamiana* leaves and grown in the dark for 48 h. The leaves were then sprayed with luciferin and kept in the dark for 10 min to quench the fluorescence. A CCD imaging apparatus (NightOWL LB 983 NC100, Berthold Technologies GmbH & Co. KG, Bad Wildbad) was used for image capture.

Resequencing

DNA of 55 individuals from 6 populations, which represent most of the known distribution of *T. sinense* trees, was extracted and sequenced using Illumina technology. Raw data were filtered using Fastq [106]. Clean reads were aligned to the *T. sinense* genome using BWA-MEM [81]. Duplicate reads were marked with sambamba-markdup [107]. FreeBayes [108] was used to conduct SNP calling.

Ancestry estimation of individuals was inferred using ADMIXTURE [50] with different values of K ($K = 1$ to 10). We performed 5-fold cross-validation to determine the most likely number of clusters. The GCTA tool [109] was used to generate genetic relationship matrix (GRM) files and the GRM file was used to compute principal components (PCs). We constructed a neighbor-joining (NJ) phylogenetic tree using MEGA7 [110] based on the whole-genome SNPs of the 55 individuals. Branch support was estimated from 1000 bootstrap replicates. *T. aralioides* was used as an outgroup.

Population genetic parameters, including nucleotide diversity (π) [111] and the Watterson estimator (θ_w) [112], were used to measure the degree of variability within a population or species and F_{st} (fixation index) was used to measure the degree of genetic differentiation, which was calculated using the thetaStat command of ANGSD software [113] over non-overlapping 20-kb windows. To estimate gene flow between populations, we performed Treemix v1.13 on the SNPs with the settings `-se -bootstrap -k 500 -m`, and values for `(-m)` ranging from 1 to 10. We applied the stairway plot [114] to the whole-genome sequences of nine populations to infer the historical effective population size. The estimated generation time and mutation rate were set to 15 and $2.745e-9$, respectively.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-020-02198-7>.

Additional file 1. Supplementary Notes, Supplementary Figs. S1–S14, and Supplementary Tables S1–S5.

Additional file 2. Review history.

Acknowledgements

Research reported in this publication was performed in the frame of National Natural Science Foundation of China and was supported by Beijing Forestry University, and by Institute of Botany, Chinese Academy of Sciences. We thank Kunming Institute of Botany, Chinese Academy of Sciences, and Shennongjia National Park for providing environmental DNA samples. We thank Dr. Zhao Liang, Yu Zhi-Yong, Wu Wei, Hu Jie, Wei Denglang, Liao Rong-Li, and Peng Lin-Peng for help with DNA sample collections. We thank Quan-zheng Yun from Beijing Ori-Gen Science and Technology Co., Ltd. for conducting genome sequencing and providing bioinformatics support. The wood testing work was supported by Beijing Zhongkebaice Technology Service Co., Ltd. The work was conducted by Beijing Advanced Innovation Center for Tree Breeding by Molecular Design.

Peer review information

Andrew Cosgrove was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 2.

Authors' contributions

JXL and YNJ conceived and planned the project. P-LL, XZ, J-FM, and Y-MH are the lead investigators and contributed equally to the work. P-LL, R-G Z, J-FM, CKJ, SHN, JBX, and LX managed and coordinated all the bioinformatics activities surrounding assembly, quality control, set-up, and maintenance of *Tetracentron sinense*. YLE, QZHH, and WQZ constructed the vector for overexpression and GFP fusion protein transgenic plants. XZ carried out the yeast one-hybrid assay. XZ and Y-MH performed the microCT experiment. WWS, Y-MH, and HMX performed the wood maceration and contributed to the scanning electron microscope imaging. SN, KHJ, JH, and XCT participated in bioinformatics data analysis. XMJ, YFY, and BL prepared and produced most of the wood sections. YPM contributed to material preparation and collection. SA, PKJ, LW, BJ, ZBQ, FB, JS, and XQH helped in correcting the manuscript. C-JL, HZK, SG, and RAD revised the manuscript and provided suggestions for experimental design and project direction. All authors contributed to the writing of the paper and commented on the manuscript. P-LL, XZ, J-FM, and Y-MH contributed equally. Correspondence and requests for materials should be addressed to JXL. The authors read and approved the final manuscript.

Funding

The research reported in this publication was supported by the National Natural Science Foundation of China (31530084), the Fundamental Research Funds for the Central Universities (2019ZY28), the Program of Introducing Talents of Discipline to Universities (111 project, B13007), and the National Science Foundation for Young Scientists of China (Grant No. 32000558).

Availability of data and materials

All raw sequencing reads generated in this study have been deposited to the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) with Bioproject accession number PRJNA625382 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA625382>) [115]. The genome assembly sequences and gene annotations have been deposited at DDBJ/ENA/GenBank under the accession JABCRI010000000. Original images of *T. sinense* captured from field and original figures captured from microCT used for cell parameters analysis have been deposited to the figshare online database (<https://doi.org/10.6084/m9.figshare.13159991.v2>) [116].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing financial interests.

Author details

¹Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing 100083, China. ²College of Biological Sciences & Biotechnology, Beijing Forestry University, Beijing 100083, China. ³Beijing Ori-Gen Science and Technology Co., Ltd, Beijing 102206, China. ⁴College of Life Sciences, Peking University, Beijing 100871, China. ⁵School of Ecology and Nature conservation, Beijing Forestry University, Beijing 100083, China. ⁶Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA. ⁷Wood Collections (WOODPEDIA), Chinese Academy of Forestry, Beijing 100091, China. ⁸Department of Wood Anatomy and Utilization, Research Institute of Wood Industry, Chinese Academy of Forestry, Beijing 100091, China. ⁹College of Horticulture and Plant Protection, Yangzhou University, Yangzhou 225009, China. ¹⁰Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China.

¹¹College of Life Science, Henan Normal University, Xixiang 453007, China. ¹²Institute of Cellular and Molecular Botany (IZMB), University of Bonn, Kirschallee 1, 53115 Bonn, Germany. ¹³Centre of the Region Haná for Biotechnological and Agricultural Research, Faculty of Science, Palacký University, 78301 Olomouc, Czech Republic. ¹⁴State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, China. ¹⁵College of Biological Sciences, China Agricultural University, Beijing 100193, China. ¹⁶Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. ¹⁷BioDiscovery Institute and Department of Biological Sciences, University of North Texas, Denton, TX 76203, USA.

Received: 18 March 2020 Accepted: 6 November 2020

Published online: 02 December 2020

References

- Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat Commun*. 2014;5:4956.
- Bremer B, Bremer K, Chase MW, Fay MF, Reveal JL, Soltis DE, et al. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc*. 2009;161:105–21.
- Byng JW, Chase MW, Christenhusz MJM, Fay MF, Judd WS, Mabberley DJ, et al. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc*. 2016;181:1–20.
- Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, Mckain MR, McNeal J, et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biol*. 2012;13:R3.
- Akoev G, Nordborg M. The *Aquilegia* genome reveals a hybrid origin of core eudicots. *Genome Biol*. 2019;20:256.
- Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, et al. Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot*. 2011;98:704–30.
- Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. From algae to angiosperms—inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evol Biol*. 2014;14:23.
- Endress PK. Floral structure, systematics, and phylogeny in *Trochodendrales*. *Ann Mo Bot Gard*. 1986;73:297–324.
- Doweld AB. Carpology, seed anatomy and taxonomic relationships of *Tetracentron* (Tetracentraceae) and *Trochodendron* (Trochodendraceae). *Ann Bot-London*. 1998;82:413–43.
- Chase MW, Soltis DE, Olmstead RG, Morgan D. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcl*. *Ann Mo Bot Gard*. 1993;80:528–80.
- Soltis DE, Soltis PS, Nickrent DL, Johnson LA, Hahn WJ, Hoot SB, et al. Angiosperm phylogeny inferred from 18s ribosomal DNA sequences. *Ann Mo Bot Gard*. 1997;84:1–49.
- Guo LR, Winzer T, Yang X, Li Y, Ning Z, He Z, et al. The opium poppy genome and morphinan production. *Science*. 2018;362:343–7.
- Ming R, Leebens-Mack J, Ayyampalayam S, Bowers JE, Mckain MR, McNeal J, et al. Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol*. 2013;14:R41.
- Ohashi-Ito K, Iwamoto K, Fukuda H. LOB domain-containing protein 15 positively regulates expression of VND7, a master regulator of tracheary elements. *Plant Cell Physiol*. 2018;59:989–96.
- Wan T, Liu Z-M, Li L-F, Leitch AR, Leitch IJ, Lohaus R, et al. A genome for gnetophytes and early evolution of seed plants. *Nat Plants*. 2018;4(2):82–9.
- Bailey IW, Thompson WP. Additional notes upon the angiosperms *Tetracentron*, *Trochodendron*, and *Drimys*, in which vessels are absent from the wood. *Ann Bot*. 1918;32:503–12.
- Suzuki M, Joshi L, Fujii T, Noshiro S. The anatomy of unusual tracheids in *Tetracentron* wood. *IAWA Bulletin*. 1991;12:23–33.
- Carlquist S. Pit membrane remnants in perforation plates of primitive dicotyledons and their significance. *Am J Bot*. 1992;79:660–70.
- Ren Y, Chen L, Tian XH, Zhang XH, Lu AM. Discovery of vessels in *Tetracentron* (Trochodendraceae) and its systematic significance. *Plant Syst Evol*. 2007;267:155–61.
- Li H-F, Chaw S-M, Du C-M, Ren Y. Vessel elements present in the secondary xylem of *Trochodendron* and *Tetracentron* (Trochodendraceae). *Flora*. 2011;206(6):595–600.
- Yamaguchi M, Mitsuda N, Ohtani M, Ohme-Takagi M, Kato K, Demura T. VASCULAR-RELATED NAC-DOMAIN 7 directly regulates the expression of a broad range of genes for xylem vessel formation. *Plant J*. 2011;66:579–90.
- Zhong R, Richardson EA, Ye ZH. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. *Plant Cell*. 2007;19:2776–92.
- McCarthy RL, Zhong R, Ye ZH. MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *Plant Cell Physiol*. 2009;50:1950–64.
- Kim WC, Ko JH, Kim JY, Kim J, Bae HJ, Han KH. MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in *Arabidopsis*. *Plant J*. 2013;73:26–36.
- Somerville C. Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol*. 2006;22:53–78.
- Funk V, Kositsup B, Zhao C, Beers EP. The *Arabidopsis* xylem peptidase XCP1 is a tracheary element vacuolar protein that may be a papain ortholog. *Plant Physiol*. 2002;128:84–94.
- Soyano T, Thitamadee S, Machida Y, Chua NH. *ASYMMETRIC LEAVES2-LIKE19/LATERAL ORGAN BOUNDARIES DOMAIN30* and *ASL20/LBD18* regulate tracheary element differentiation in *Arabidopsis*. *Plant Cell*. 2008;20:3359–73.
- Endo H, Yamaguchi M, Tamura T, Nakano Y, Nishikubo N, Yoneda A, et al. Multiple classes of transcription factors regulate the expression of *VASCULAR-RELATED NAC-DOMAIN7*, a master switch of xylem vessel differentiation. *Plant Cell Physiol*. 2015;56:242–54.
- Ratter JA, Milne C. Chromosome numbers of some primitive angiosperms. Notes from the Royal Botanic Garden, Edinburgh. 1973;32:423–8.
- Pigg KB, Wehr WC, Ickert-Bond SM. *Trochodendron* and *nordenskioldia* (Trochodendraceae) from the middle eocene of Washington State, USA. *Int J Plant Sci*. 2011;162:1187–98.

31. Sun Y, Moore MJ, Yue L, Feng T, Chu H, Chen S, et al. Chloroplast phylogeography of the East Asian Arcto-Tertiary relict *Tetracentron sinense* (Trochodendraceae). *J Biogeogr.* 2014;41:1721–32.
32. Li S, Gan X, Han H, Zhang X, Tian Z. Low within-population genetic diversity and high genetic differentiation among populations of the endangered plant *Tetracentron sinense* Oliver revealed by inter-simple sequence repeat analysis. *Ann For Sci.* 2018;75:74.
33. Xu GC, Xu TJ, Zhu R, Zhang Y, Li SQ, Wang HW, et al. LR_GapCloser: A tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience.* 2019;8: <https://doi.org/10.1093/gigascience/giy157>.
34. Simao FR, Waterhouse M, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31:3210–2.
35. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 2015;16:157. <https://doi.org/10.1186/s13059-015-0721-2>.
36. Li HT, Yi TS, Gao ML, Ma PF, Zhang T, Yang JB, et al. Origin of angiosperms and the puzzle of the jurassic gap. *Nat Plants.* 2019;5:461–70.
37. Yang ZH, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol.* 2006;23:212–26.
38. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, et al. BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput Biol.* 2014;10:e1003537.
39. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics.* 2006;22:1269–71.
40. Albert VA, Barbazuk WB, de Pamphilis CW, Der JP L-MJ, Ma H, et al. The amborella genome and the evolution of flowering plants. *Science.* 2013;342:1241089.
41. Zhang L, Chen F, Zhang X, Li Z, Zhao Y, Lohaus R, et al. The water lily genome and the early evolution of flowering plants. *Nature.* 2020;577:79–84.
42. Chaw S-M, Liu Y-C, Wu Y-W, Wang H-Y, Lin C-YI WC-S, et al. Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. *Nat Plants.* 2019;5:63–73.
43. Gui S, Peng J, Wang X, Wu Z, Cao R, Salse J, et al. Improving *Nelumbo nucifera* genome assemblies using high-resolution genetic maps and BioNano genome mapping reveals ancient chromosome rearrangements. *Plant J.* 2018;94:721–34.
44. Shi T, Rahmani RS, Gugger PF, Wang M, Li H, Zhang Y, et al. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. *Mol Biol Evol.* 2020;37:2394–413.
45. Ohtani M, Nishikubo N, Xu B, Yamaguchi M, Mitsuda N, Goue N, et al. A NAC domain protein family contributing to the regulation of wood formation in poplar. *Plant J.* 2011;67:499–512.
46. Matsumura Y, Iwakawa H, Machida Y, Machida C. Characterization of genes in the *ASYMMETRIC LEAVES2/LATERAL ORGAN BOUNDARIES (AS2/LOB)* family in *Arabidopsis thaliana*, and functional and molecular comparisons between *AS2* and other family members. *Plant J.* 2009;58:525–37.
47. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. *MYB* transcription factors in *Arabidopsis*. *Trends Plant Sci.* 2010;15:573–81.
48. Richau KH, Kaschani F, Verdoes M, Pansuriya TC, Niessen S, Stueber K, et al. Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. *Plant Physiol.* 2012;158:1583–99.
49. Zhong R, Ye Z-H. Complexity of the transcriptional network controlling secondary wall biosynthesis. *Plant Sci.* 2014;229:193–207.
50. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
51. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 2012;8:e1002967.
52. Strijk JS, Hinsinger DD, Zhang F, Cao K. *Trochodendron aralioides*, the first chromosome-level draft genome in *Trochodendrales* and a valuable resource for basal eudicot research. *Gigascience.* 2019;8: <https://doi.org/10.1093/gigascience/giz136>.
53. Pigg KB, Dillhoff RM, DeVore ML, Wehr WC. New diversity among the Trochodendraceae from the early/middle eocene Okanogan highlands of British Columbia, Canada, and northeastern Washington State, United States. *Int J Plant Sci.* 2007;168:521–32.
54. Wu S, Han B, Jiao Y. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol Plant.* 2020;13:59–71.
55. Lyons E, Pedersen B, Kane J, Freeling M. The value of nonmodel genomes and an example using synmap within CoGe to dissect the hexaploidy that predates the rosids. *Tropical Plant Biol.* 2008;1:181–90.
56. Fawcett JA, Maere S, Van de Peer Y. Plants with double genomes might have had a better chance to survive the cretaceous-tertiary extinction event. *Proc Natl Acad Sci.* 2009;106:5737–42.
57. Chanderbali AS, Yoo M-J, Zahn LM, Brockington SF, Wall PK, Gitzendanner MA, et al. Conservation and canalization of gene expression during angiosperm diversification accompany the origin and evolution of the flower. *Proc Natl Acad Sci.* 2010;107:22570–22575.
58. Sharma B, Kramer EM. *Aquilegia* B gene homologs promote petaloidy of the sepals and maintenance of the C domain boundary. *Evodevo.* 2017;8:22.
59. Lee H, Suh SS, Park E, Cho E, Ahn JH, Kim SG, et al. The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. *Gen dev.* 2000;14:2366–76.
60. Liu C, Chen H, Er HL, Soo HM, Kumar PP, Han J-H, et al. Direct interaction of *AGL24* and *SOC1* integrates flowering signals in *Arabidopsis*. *Development.* 2008;135:1481–91.
61. Gregis V, Sessa A, Colombo L, Kater MM. *AGAMOUS-LIKE24* and *SHORT VEGETATIVE PHASE* determine floral meristem identity in *Arabidopsis*. *Plant J.* 2008;56:891–902.
62. de Folter S, Immink RGH, Kieffer M, Parenicova L, Henz SR, Weigel D, et al. Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. *Plant Cell.* 2005;17:1424–33.

63. Carlquist S, Schneider EL. The tracheid-vessel element transition in angiosperms involves multiple independent features: cladistic consequences. *Am J Bot*. 2002;89:185–95.
64. Ozaki K. *Tetracentron* leaves from the neogene of Japan. *Trans Proc Palaeont Soc Japan N. S.* 1987;146:77–87.
65. Suzuki M, Joshi L, Noshiro S. *Tetracentron* wood from the Miocene of Noto Peninsula, Central Japan, with a short revision of homoxyllic fossil woods. *Bot Mag Tokyo*. 1991;104:37–48.
66. Manchester SR, Crane PR, Dilcher DL. *Nordenskiöldia* and *Trochodendron* (Trochodendraceae) from the Miocene of Northwestern North America. *Botanical Gaz*. 1991;152:357–68.
67. Grimsson F, Denk T, Zetter R. Pollen, fruits, and leaves of *Tetracentron* (Trochodendraceae) from the Cainozoic of Iceland and Western North America and their palaeobiogeographic implications. *Grana*. 2008;47:1–14.
68. Tang CQ, Yang Y, Ohsawa M, Yi SR, Momohara A, Su WH, et al. Evidence for the persistence of wild *Ginkgo biloba* (*ginkgoaceae*) populations in the DALOU Moutains, southwestern China. *Am J Bot*. 2012;99:1408–14.
69. Zhao YP, Fan G, Yin PP, Sun S, Li N, Hong X, Hu G, et al. Resequencing 545 *Ginkgo* genomes across the world reveals the evolutionary history of the living fossil. *Nat Commun*. 2019;10:4201.
70. Qi XS, Chen C, Comes HP, Sakaguchi S, Liu YH, Tanaka N, et al. Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (*Cercidiphyllaceae*). *New Phytol*. 2012; 196:617–30.
71. Zheng B, Xu Q, Shen Y. The relationship between climate change and quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation. *Quat Int*. 2002;97:93–101.
72. Marçais G, Kingsford CA. Fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*. 2011;27:764–70.
73. Liu B, Shi Y, Yuan Y, Hu X, Zhang H, Li N, et al. Estimation of genomic characteristics by analyzing *k*-mer frequency in *de novo* genome projects. *arXiv*. 2012:arXiv:1308.2012v2.
74. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res*. 2017;27:722–36.
75. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res*. 2016;44:e147.
76. Walker B, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9:e112963.
77. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016;3:95–8.
78. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. *De novo* assembly of the aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356:92–5.
79. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 2016;3:99–101.
80. Pyszczyk LP, Gabaldon T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res*. 2016;44: e113.
81. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013:arXiv:1303.3997v2.
82. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–100.
83. Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Res*. 2018; 46:e126.
84. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
85. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12: 357–60.
86. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33:290–5.
87. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2.
88. Price AL, Jones NC, Pevzner PA. *De novo* identification of repeat families in large genomes. *Bioinformatics*. 2005;21:i351–8.
89. Holt C, Yandell M. Maker2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 2011;12:491.
90. Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res*. 2002;12:656–64.
91. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–40.
92. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25:955–64.
93. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. Rnammer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res*. 2007;35:3100–8.
94. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res*. 2018;46:D335–42.
95. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res*. 2002;30:3059–66.
96. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3.
97. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74.
98. Kumar S, Stecher G, Suleski M, Hedges SB. Timetree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 2017;34:1812–9.
99. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. 2010;11:R14.

100. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40:e49.
101. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
102. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 2006;34:W609–12.
103. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
104. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom Proteom Bioinf.* 2010;8:77–80.
105. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39:W29–37.
106. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:884–90.
107. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P, et al. Sambamba: fast processing of NGS alignment formats. *Bioinformatics.* 2015;31:2032–4.
108. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv.* 2012;arXiv:1207.3907v2.
109. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88:76–82.
110. Kumar S, Stecher G, Tamura K. Mega7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
111. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci.* 1979;76:5269–73.
112. Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 1975;7:256–76.
113. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics.* 2014;15:336.
114. Liu X, Fu YX. Exploring population size changes using SNP frequency spectra. *Nat Genet.* 2015;47:555–9.
115. Liu P-L, Zhang X, Mao J-F, Hong Y-M, Zhang R-G, E YL, et al. *Tetracentron sinense* Genome sequencing, assembly, resequencing and RNA-sequencing. *NCBI Sequence Read Archive*, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA625382> (2020).
116. Lin JX; Zhang X; Liu P-L: Original images of *Tetracentron sinense*. *Figshare.* Figure. 2020: <https://doi.org/10.6084/m9.figshare.13159991.v2>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

