Integrative Conservation

WILEY

INTEGRATIVE CONSERVATION XTBG

RESEARCH ARTICLE OPEN ACCESS

The Genome of *Phoebe zhennan* (Lauraceae) and Comparative Genomics of *Phoebe*

Mao Yameng[1,2] | Zou Dianyang[1,2] | Li Lang[1] | Li Jie[1]

[1]Plant Phylogenetics and Conservation Group, Center for Integrative Conservation & Yunnan Key Laboratory for the Conservation of Tropical Rainforests and Asian Elephants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming, China | [2]University of Chinese Academy of Sciences, Beijing, China

Correspondence: Li Lang (lilang@xtbg.ac.cn) | Li Jie (jieli@xtbg.ac.cn)

Received: 10 February 2025 | **Revised:** 16 October 2025 | **Accepted:** 20 October 2025

Editor-in-Chief: Binbin Li | **Handling Editor:** Peng Liu

Funding: This study was supported by National Natural Science Foundation of China (31970222) and The 14th 5-Year Plan of the Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (XTBG-1450101).

Keywords: chromosome-level genome | Lauraceae | lignin biosynthesis | MYB gene family | *Phoebe zhennan*

ABSTRACT

Phoebe zhennan is a Chinese endemic and valuable timber species of *Phoebe* (Lauraceae). It is known for its fragrant, high-value wood and is the primary source of the renowned "golden-thread nanmu." In this study, we present a chromosome-level genome assembly of *P. zhennan*, consisting of 12 chromosomes (979.28 Mb; scaffold N50: 87.69 Mb) with an estimated 26,262 protein-coding genes. Whole-genome duplication (WGD) analysis revealed that *P. zhennan* has undergone two WGD events: an older one shared by Magnoliales and Laurales, and a younger one shared by Lauraceae and Calycanthaceae species. Phylogenetic analysis showed that magnoliids (together with Chloranthales) form a sister clade to monocots. We identified 197 members of the MYB gene family and further selected 44 MYB genes that may be associated with phenylpropanoid lignin biosynthesis, including 28 genes acting as positive regulatory transcription factors and 16 as negative regulators. Gene expression analysis indicated that *PzMYB6* is highly expressed in the cambium, suggesting it may play a critical regulatory role in the lignin biosynthesis pathway of *P. zhennan*. Overall, this chromosome-level reference genome provides new insights into the genome evolution of *Phoebe* and offers essential genetic resources for the rational utilization and conservation of this valuable timber species.

1 | Introduction

Phoebe Nees, a genus within the tribe Perseeae, holds significant economic and ecological importance within the family Lauraceae (Li et al. 1982, 2008; Li et al. 2025). According to POWO (https://powo.science.kew.org), *Phoebe* comprises 69 accepted species native to tropical and subtropical Asia. In China, the genus includes 33 species and 3 varieties, of which 28 (including varieties) are endemic. These are mainly distributed across Yunnan, Sichuan, Hubei, Guizhou, Guangxi, and Guangdong (Li et al. 1982, 2008; Li et al. 2011; Yang and Liu 2015; Huang et al. 2011; Rohde et al. 2017; Liu et al. 2020; Yang et al. 2021).

The precious wood of *Phoebe* species, commonly known as "golden-thread nanmu," is celebrated for its fragrance, exceptional quality, and a unique "golden thread" figure. Highly valued for centuries in fine craftsmanship and architecture—

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Integrative Conservation* published by John Wiley & Sons Australia, Ltd on behalf of Xishuangbanna Tropical Botanical Garden (XTBG).

Summary

Phoebe zhennan is a treasured and endangered timber species native to China, prized for its aromatic fragrance, durability, and striking wood grain. However, the lack of genetic resources has long impeded both its conservation and a deeper understanding of the biological basis for its superior wood properties. In this study, we successfully sequenced and assembled the complete genome of *P. zhennan*. Our analyses uncovered two ancient whole-genome duplication events in its evolutionary history, which likely facilitated species diversification and the emergence of its distinctive traits. Phylogenetic reconstruction further resolved the phylogenetic position of magnoliids, revealing a closer relationship to monocots than to eudicots. Most importantly, we identified key genes involved in lignin biosynthesis—the primary structural component of wood. Among these, the transcription factor gene *PzMYB6* stood out as a master regulator highly expressed in the trunk cambium, which coordinates the developmental processes that lead to the formation of its robust and highly valued timber. This high-quality genome assembly serves as a crucial resource for informing conservation strategies, monitoring genetic diversity in wild populations, and advancing the breeding of superior cultivars for sustainable utilization.

Practitioner Points

- This study provides the first chromosome-level genome of *Phoebe zhennan*, serving as an essential foundation for its genetic conservation and breeding programs.
- We identified key genes—particularly a master regulator, PzMYB6—that control lignin synthesis, the process that makes wood strong and durable. This discovery opens possibilities for molecular breeding to improve wood quality.
- The genomic data can be used to assess genetic diversity and wild population health, design effective nature reserves, and develop genetic markers to assist in selecting superior trees for reforestation and sustainable cultivation.

and historically regarded as the "Emperor's Wood" in imperial China—it has long symbolized luxury and exclusivity (Cheng 1983; Li et al. 2004; Chen 2013). The main sources of golden-thread nanmu are *Phoebe* species from subtropical China, among which *P. bournei*, *P. chekiangensis*, *P. hui*, and *P. zhennan* are the most valuable (Li et al. 1982, 2008). Of these, *P. bournei* and *P. zhennan* are considered the finest sources, producing tall, thick trunks and exceptionally high-quality timber (Li et al. 1982; Lin 1988; Ding et al. 2019). However, the delimitation of these two *Phoebe* species has long been controversial. Their morphological characteristics overlap considerably, and their geographic ranges intersect (Lee et al. 1979; Li et al. 2008; Zou and Wu 1997).

Recently, Ding et al. (2019) utilized an integrative approach combining morphological analysis and molecular data to clarify the boundaries between *P. bournei* and *P. zhennan*. Their findings identified two distinct lineages corresponding to these species. Morphologically, *P. bournei* exhibits more prominent higher-order leaf venation, thicker petioles, and a narrower leaf apex angle than *P. zhennan*. Geographically, *P. bournei* is primarily distributed in central Fujian, southern Jiangxi, and the Nanling Mountains, while *P. zhennan* is mainly found in the eastern Qionglai Mountains and southern Sichuan hills. The two species are largely distinct, with overlap occurring only in the Wuling Mountains.

Subsequently, the genome of *P. bournei* was sequenced and studied. Chen, Sun, et al. (2020) identified two whole-genome duplication (WGD) events in the *P. bournei* genome, supporting the phylogenetic placement of magnoliids at the base of angiosperms as a sister clade to monocots and eudicots. Their study also identified several important gene families in the *P. bournei* genome, including 63 MADS-box genes (with four AGL12-like and four ANR1-like genes potentially involved in root regulation), 17 GH3 and 77 SAUR genes (linked to growth and light-mediated development), and five FIN219-like and two SAUR50-like genes (associated with photomorphogenesis and floral or stem development). Moreover, four MYB46 and three EPSP homologous genes, which promote lignin biosynthesis, were identified and hypothesized to contribute to the straight trunk characteristic of *P. bournei*. Han et al. (2022) later produced a chromosome-level genome assembly of *P. bournei* and identified 72 full-length terpene synthase (TPS) genes. Comparative genomic analysis highlighted extensive lineage-specific duplications and contractions within TPS subfamilies, which have contributed to the remarkable terpenoid diversity observed in Lauraceae.

Historically, *P. zhennan* forests were once widespread in subtropical evergreen broad-leaved forests of China, with their northern distribution boundary extending up to 35° N (Li et al. 1982, 2008; Ding et al. 2015). However, excessive logging, particularly during the Ming and Qing dynasties, has nearly exhausted natural populations (Lin 1988). In China, the high economic value of *P. zhennan* has historically led to severe overexploitation, pushing it toward endangerment. The IUCN Red List of Threatened Species (https://www.iucnredlist.org/), the China Species Red List (Wang and Xie 2004), and the List of National Key Protected Wild Plants (Forestry et al. 2021) all classify *P. zhennan* as threatened. Currently, it is classified as a National Class II key protected plant in China and is listed as Vulnerable in terms of conservation status.

As one of the most important transcription factor families, the myeloblastosis transcription factor (MYB TF) family is widely distributed across eukaryotes (Stracke et al. 2001; Dubos et al. 2010). Members of the MYB family play crucial roles in regulating various biological processes in plants, such as secondary metabolism, stress responses, and growth and development. Secondary metabolism includes pathways such as those of phenylpropanoids, flavonoids, terpenoids, and glucosinolates. The phenylpropanoid lignin biosynthesis pathway is the major route for lignin synthesis in plants and is regulated by multiple members of the MYB gene family (Xiao et al. 2021; Sun, Wang, et al. 2022). Chen, Sun, et al. (2020) identified four homologous genes of *MYB46* and three homologous genes of *PtrEPSP* in the *P. bournei* genome and proposed that these genes may be associated with the

formation of straight trunks in *P. bournei*. However, the functional mechanisms of the MYB gene family and the lignin biosynthesis pathway in *P. zhennan* remain unclear.

Here, we present a chromosome-level genome assembly of *P. zhennan*. Our research aims to provide insights into the polyploidization history of *P. zhennan*, as well as the expression and evolution of the MYB gene family involved in phenylpropanoid lignin biosynthesis. The findings are expected to supply essential genomic data for this prized species. Furthermore, our study offers new perspectives for comparative genomics and genome evolution in *Phoebe*, providing scientific evidence to support the sustainable utilization, effective conservation, and rational management of these valuable plant resources in China.

2 | Results

2.1 | Genome Assembly and Main Features

Genomic survey analysis using DNBSEQ platform data (300–400 bp short-insert library) generated approximately 177.6 Gb of raw sequences, with 176.43 Gb retained after filtering. K-mer-based analysis estimated the *P. zhennan* genome size to be 0.89 Gb, with a heterozygosity of 3.19% and a repetitive content of 51.71%.

PacBio HiFi sequencing libraries were constructed following the single-molecule real-time (SMRT) sequencing protocol. A 20 kb library was generated, and sequencing was performed in CCS mode using the PacBio Sequel II platform, completing sequencing for one cell. A total of 33.16 million raw reads were obtained, generating 525.24 Gb of raw data. After CCS filtering, 1.98 million HiFi reads were obtained, totaling 35.27 Gb. Based on an approximate genome size of 1 Gb, the sequencing depth was ~35X. The draft genome was assembled using Hifiasm, resulting in a total contig length of 1073 Mb with an N50 of 10.46 Mb. The completeness of the gene space was assessed as 99% based on BUSCO analysis against the embryophyta_odb10 database.

A Hi-C library was constructed and sequenced on the DNBseq platform across eight lanes, yielding a total of 128.24 Gb of raw sequencing data. After filtering, 125.59 Gb of clean data were obtained. Hi-C scaffolding produced a 1002 Mb genome assembly with a contig-level N50 of 10.52 Mb and a scaffold-level N50 of 87.69 Mb. The genome was successfully anchored to chromosomes, achieving a mapping rate of 97.74% and a total length of 979.28 Mb (Figure 1B,C). BUSCO assessment (Simão et al. 2015) against the embryophyta_odb10 database showed a completeness rate of 98.9%, with a repeat alignment rate of 6.1%. The LTR Assembly Index (LAI) value was 14.76.

2.2 | Genome Annotation

Through repeat sequence annotation, a total of 541.0 Mb of repetitive sequences were identified in the *P. zhennan* genome, accounting for 53.99% of the genome size. Among these, Copia elements accounted for 105.8 Mb (10.56% of the genome), while Gypsy elements totaled 244.6 Mb (24.41% of the genome). The final annotation, through an integrated structural annotation pipeline (homology-based prediction, de novo prediction, and RNA-Seq evidence), yielded 28,279 protein-coding genes with 98.2% BUSCO completeness (single-copy: 1,512 [93.68%]; duplicated: 73 [4.52%]). Functional annotation identified 26,262 functional genes, accounting for 92.87% of the total gene set (Table 1). The overlap of annotations across five databases is depicted in Table 2, where 17,562 genes were annotated in all databases.

2.3 | Synteny Analysis and Whole-Genome Duplications (WGD)

To investigate WGD events in *P. zhennan*, we first identified syntenic blocks within the genome based on gene sequence similarity (Figure 2A). The syntenic dot plot of the *P. zhennan* genome revealed two distinct clusters of Ks values (0.3–0.9 and 0.9–1.4), corresponding to two ancient WGD events. Major syntenic blocks between chromosomes 4–5 and 5–12, as well as within chromosome 10. We calculated the median synonymous substitution rates (Ks) for homologous gene pairs within each block and fitted a Ks distribution curve. The resulting graph revealed two distinct peaks at Ks values of 0.6 and 1.0 (Figure 2B). The first Ks peak (0.6) corresponds to the most recent WGD event, while the second Ks peak (1.0) indicates an older WGD event. This suggests that the *P. zhennan* genome has undergone at least two WGD events in its evolutionary history. The synteny comparison between *Amborella trichopoda* and *P. zhennan* shows a depth ratio of 1:4 for the syntenic blocks (Figure 2D), providing additional evidence for these two ancient WGD events in the *P. zhennan* genome.

To further determine the nature of the WGD events in *P. zhennan*, we performed both intraspecific and interspecific synteny analyses between the *P. zhennan* genome and those of seven other magnoliid species (*Chimonanthus salicifolius*, *Magnolia sinica*, *Saururus chinensis*, *P. bournei*, *Persea americana, Lindera megaphylla*, and *Cinnamomum camphora*), and calculated Ks values to compare their distributions. From the Ks distribution graph, we observed a shared peak at $Ks \approx 0.9$ among several magnoliid species. At $Ks \approx 0.6$, *P. zhennan*, *P. bournei*, *Persea americana, Lindera megaphylla, Cinnamomum camphora*, and *Chimonanthus salicifolius* exhibited a common peak, whereas other species such as *Magnolia sinica* and *Saururus chinensis* did not (Figure 2C). In summary, these results suggest that *P. zhennan* experienced two WGD events: the older one is shared by Magnoliales and Laurales, and the younger one is shared by all species of Lauraceae and Calycanthaceae analyzed here.

2.4 | Comparative Genomic Analysis

We conducted a comparative genomic analysis using the whole genomes of 21 angiosperms. A total of 692,624 genes were clustered into 41,813 orthogroups, representing 91.5% of all genes. Fifty percent of the genes were in contained orthogroups with 34 or more genes (G50 = 34) and within the largest 6095 orthogroups (O50 = 6095). There were 5,749 orthogroups shared by all species analyzed. Using four fossil calibration points to estimate divergence times, we calculated divergence with 95% confidence intervals. The results indicated that magnoliids (along with Chloranthales) and monocots form a sister group, while *Phoebe* and *Persea* cluster together (Figure 3A). The divergence between *P. zhennan* and *P. bournei* occurred approximately 11.14

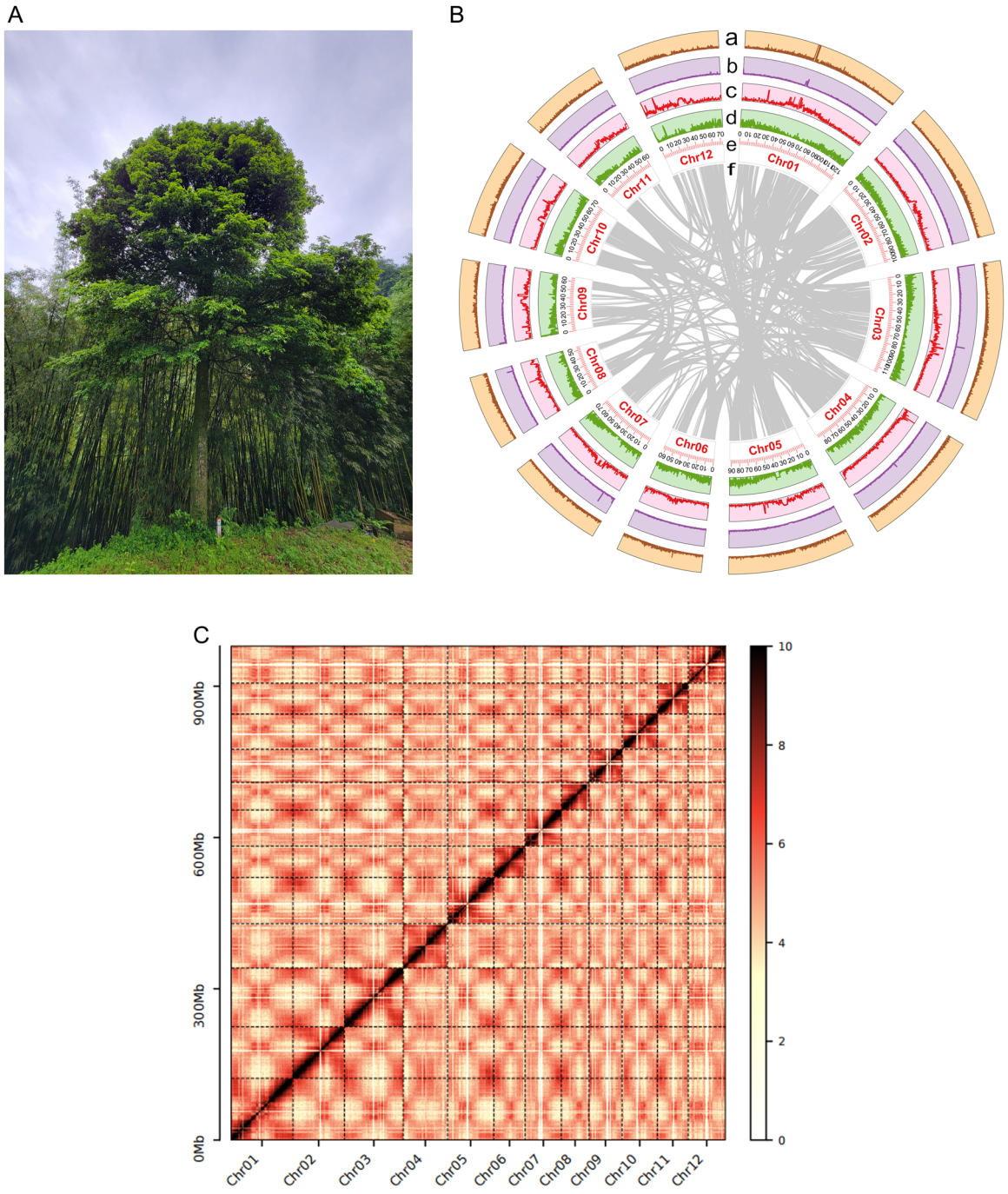


FIGURE 1 | *Phoebe zhennan* and landscape of its genome. (A) Morphological image of *P. zhennan*. (B) Circos plot of the *P. zhennan* genome, with tracks representing: (a) LTR density in a 50 kb window; (b) Repeat gene density in a 50 kb window; (c) GC content in a 100 kb window; (d) Gene density in a 100 kb window; (e) Chromosome structure; (f) Genomic collinearity. (C) Hi-C heatmap of chromosome interaction strength in the *Phoebe zhennan* genome (bin size: 500 kb).

million years ago (mya), and the divergence between Lauraceae and other Laurales occurred around 101.17 mya.

Gene family analysis using CAFE5 revealed that in *P. zhennan*, 658 gene families experienced significant expansion, while 1561 gene families underwent significant contraction. In comparison, the number of expanded gene families in *Persea americana*, *Phoebe bournei*, and *Cinnamomum camphora* was 1313, 1695, and 1057, respectively, while the number of contracted gene families was 932, 1,149, and 454, respectively. Even between the closely related species *P. zhennan* and *P. bournei*, marked differences in the number of expanded and contracted gene families were evident.

We further performed GO and KEGG enrichment analyses on the significantly expanded gene families in *P. zhennan*. The GO terms were mainly enriched in pathways associated with secondary metabolic processes (GO:0019748) and UDP-glycosyltransferase activity (GO:0035251) (Figure 3B). In the KEGG analysis, enriched pathways included monoterpenoid biosynthesis (KO:00902), phenylpropanoid biosynthesis (KO:00940), and the MAPK signaling

TABLE 1 | Comparison of gene structure features in ***Phoebe zhennan*** predicted using different methods.

	Gene set	Number	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
De novo	Augustus	118,521	4549.44	1452.7	3.77	384.9	1116.25
homolog	*Aquilegia coerulea*	25,479	11,855.28	1358.54	5.42	250.44	2372.37
	Arabidopsis thaliana	21,626	12,265.3	1372.68	5.72	239.84	2306.1
	Cinnamomum micranthum	33,236	11,828.15	1382.0	5.32	259.75	2417.83
	Daucus carota	23,932	11,956.41	1322.69	5.31	248.9	2464.8
	Ginkgo biloba	20,916	11,779.37	1233.64	4.74	260.13	2817.87
	Liriodendron chinense	25,752	12,365.59	1356.75	5.16	262.72	2643.62
	Mimulus guttatus	24,329	11,819.69	1330.89	5.49	242.58	2337.88
	Phoebe bournei	30,616	19,061.54	1213.72	4.56	266.43	5019.74
RNA-seq	Transcript	54,448	19,121.99	1163.36	5.19	224.03	4283.16
Final	—	28,279	9861.1	1421.34	5.64	252.08	1819.55

TABLE 2 | Functional annotation statistics of the final predicted protein-coding genes in the *Phoebe zhennan* genome.

Database	Number	Percent
Nr	26,189	92.61%
Swissprot	20,735	73.32%
KEGG	20,670	73.09%
KOG	20,474	72.40%
TrEMBL	26,187	92.60%
Interpro	24,676	87.26%
GO	16,314	57.69%
Overall	26,262	92.87%

pathway (KO:04010) (Figure 3C). These enriched pathways are mostly related to the synthesis of secondary metabolites and processes involved in wood secondary growth. UDP-glycosyltransferase activity contributes to the synthesis of cell wall components, particularly in the cross-linking of lignin and cellulose. The phenylpropanoid biosynthesis pathway is directly linked to lignin biosynthesis, a key structural element of the plant cell wall that supports xylem function. The MAPK signaling pathway plays a crucial role in cell division, differentiation, and stress responses in plants, and is involved in stem development and vascular tissue formation.

2.5 | MYB Gene Family and Lignin Synthesis-Related Enzyme Genes

Based on the *Arabidopsis thaliana* MYB gene family (141 genes), we identified a total of 197 MYB gene family members in *P. zhennan* through the intersection of HMM-BLAST searches and Pfam domain alignments. These genes were subsequently named according to their chromosomal arrangement. We then performed an evolutionary analysis of *Arabidopsis thaliana* and *P. zhennan* MYB family members to determine their homologous relationships (Figure 4A).

In the phenylpropanoid lignin biosynthesis model, 17 MYB transcription factors are involved, of which twelve are positive regulators and five are negative regulators (Figure 4C). We further selected 44 MYB genes from the *P. zhennan* MYB gene family related to phenylpropanoid lignin biosynthesis. They are unevenly distributed across the chromosomes, with many members of the individual gene subfamily clustering together as tandem duplicates (Figure 4D).

In addition to the MYB gene family, we similarly identified key structural genes encoding enzymes in the lignin biosynthesis pathway of *P. zhennan*. A total of 19 genes were identified, including three *PAL*s, two *C4H*s, and one gene each for *C3H*, *CCR*, and *4CL*, along with two *F5H*s, three *COMT*s, three *CAD*s, and three *CAD*-like genes.

Gene-expression levels of these MYB and structural genes were analyzed across five different tissues from three individuals of *P. zhennan* (as described in Section 4.1): leaves, current-year branches, vascular cambium of the trunk, flowers, and young fruits (Figure 4B). Overall, genes associated with the regulation of phenylpropanoid lignin biosynthesis showed the highest expression in the cambium, followed by leaves and current-year branches. Expression in young fruits was lower, and the lowest expression was observed in the flowers.

Based on the extensive characterization of the MYB gene family in *P. zhennan*, we further investigated the co-expression relationships among these genes by performing a Weighted Gene Co-expression Network Analysis (WGCNA) using transcriptome data from the five aforementioned tissues. This analysis revealed two distinct modules of co-expressed genes. Notably, the turquoise module was significantly and positively correlated with high expression

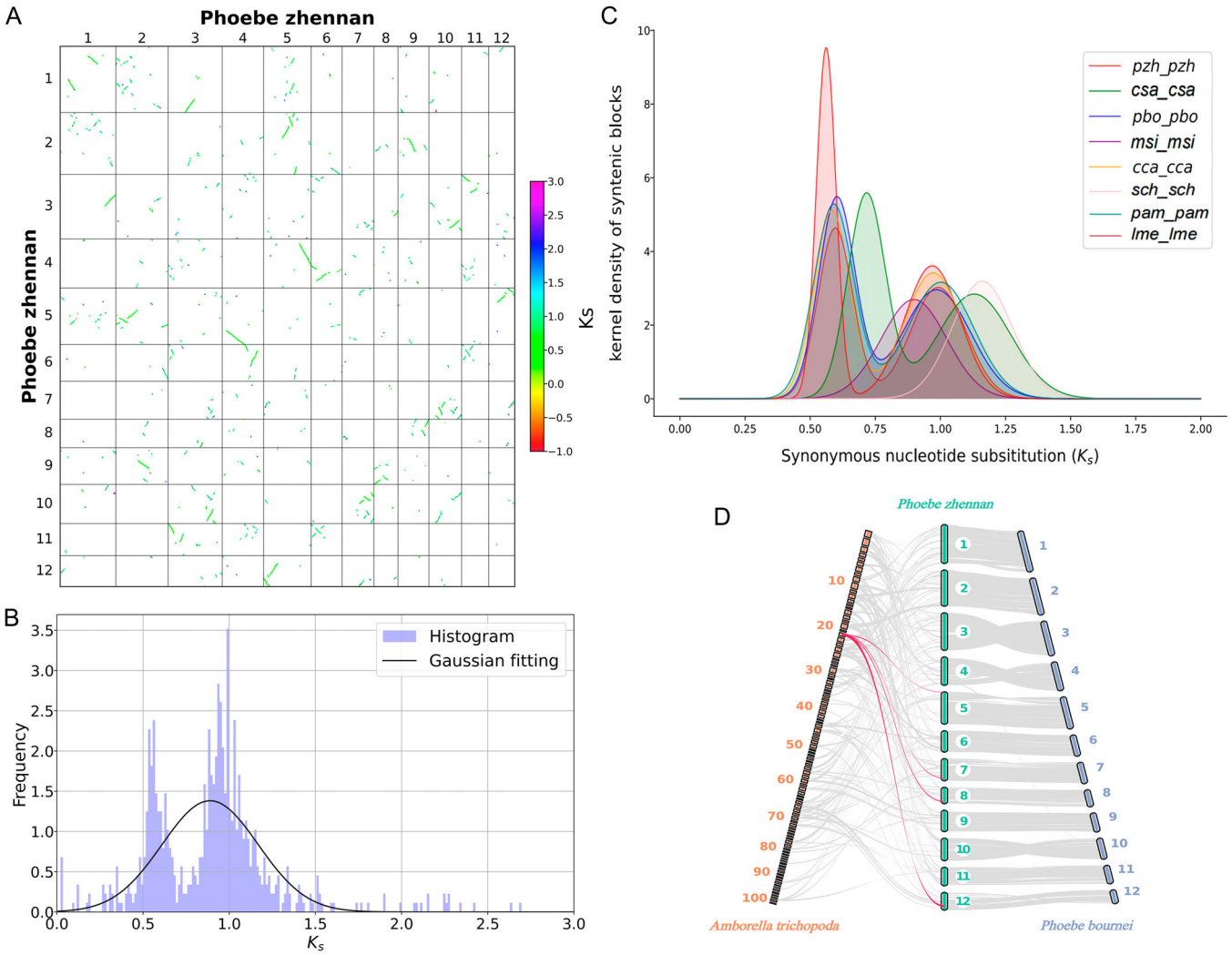


FIGURE 2 | Whole-genome duplication analysis. (A) Syntenic blocks of the *Phoebe zhennan* genome. (B) Ks distribution of syntenic blocks in the *P. zhennan* genome. The histogram shows the original Ks distribution, with the black curve representing the Gaussian fit of the raw Ks. (C) Ks distribution of syntenic blocks within the genomes of each species. pzh-*Phoebe zhennan*; csa-*Chimonanthus salicifolius*; pbo-*Phoebe bournei*; msi-*Magnolia sinica*; cca-*Cinnamomum camphora*; sch-*Saururus chinensis*; pam-*Persea americana*; lmg-*Lindera megaphylla*. (D) Synteny patterns of genomic regions in *P. zhennan*, *P. bournei*, and *Amborella trichopoda*.

(Cor = 0.83, *p*-value = 1e–04). Within this co-expression network, 16 MYB transcription factors and 19 key structural genes formed a highly interconnected module. In the WGCNA co-expression network, this indicates that these genes exhibit tightly correlated expression patterns ("rising and falling together") across all samples, suggesting they are very likely to be classified into the same gene module and to work together in the same biological pathway or complex. The strong co-expression between these MYB transcription factors and lignin biosynthetic genes provides compelling evidence that they constitute a core transcriptional regulatory network controlling lignin biosynthesis in *P. zhennan*.

3 | Discussion

3.1 | *P. Zhennan* Genome Assembly and Comparative Analysis

P. zhennan is a valuable timber species, known for its tall, straight growth form and aromatic, decay-resistant wood. In this study, we report a high-contiguity, chromosome-level genome assembly for this prized species. The assembly shows a high level of completeness (99%) based on BUSCO assessment, providing a robust foundation for subsequent genome annotation and comparative analyses. The proportion of repetitive elements annotated was 53.99%, and the LTR Assembly Index (LAI) of 14.76 indicates moderate contiguity for a genome of this size (~1 Gb), suggesting that some repetitive regions, particularly LTR retrotransposons, remain unresolved.

Through comparison with other published Lauraceae genomes, the genome size of *P. zhennan* was found to be 979.3 Mb. It is larger than those of *Persea*, *Phoebe*, and *Camphora* species, but smaller than those of *Litsea*, *Lindera*, *Cinnamomum*, and *Sextonia*. The contig N50 of the *P. zhennan* genome assembly is 10.52 Mb, and the scaffold N50 is 87.69 Mb. Both values rank among the highest reported in Lauraceae, indicating high sequencing quality and assembly continuity. The proportion of repetitive elements in *P. zhennan* (53.99%) is higher than that of *P. bournei* (48.10%), but lower than *Persea americana* (62.09%).

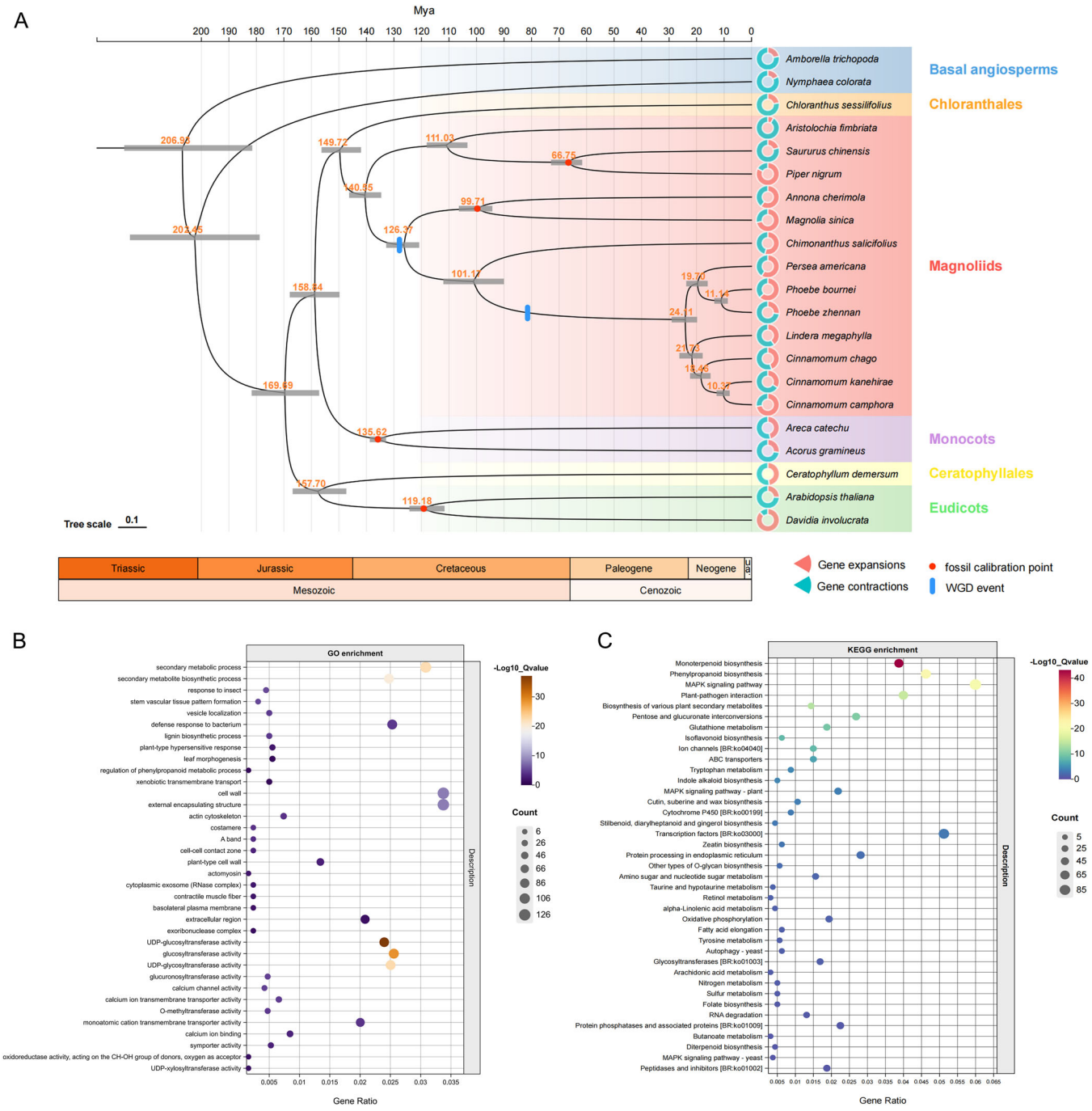


FIGURE 3 | Comparative genomic and gene family analysis. (A) Phylogenetic tree of 21 species showing divergence times. Divergence estimates (Mya, million years ago) are indicated above the nodes, and the gray nodal bars represent the 95% confidence intervals. Red dots mark the calibration points, as described in the methods section. Gene family expansions and contractions identified by CAFE5 are indicated. The red and blue circles, with corresponding numbers ($p < 0.05$), represent expansion and contraction of gene families in different species, respectively. (B) The Gene Ontology (GO) enrichment analysis of expansion genes. (C) The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of expansion genes.

Among *Camphora* species, *C. officinarum* (53.97%) and *C. micrantha* (48%) have the lowest proportions of repetitive elements. The proportion of LTRs in *P. zhennan* (34.97%) is significantly higher than that of *P. bournei* (26.34%). Compared to other reported Lauraceae species, the LTR proportion of *P. zhennan* is higher than that of *Sextonia rubra* (23.30%) but lower than that of *Cinnamomum chago* (40.71%) and *Litsea coreana* (41.52%). The genome assembly completeness of *P. zhennan* (98.9%) represents one of the highest levels among all reported Lauraceae species, second only to *Camphora longepaniculata* and *Cinnamomum chago* (99.0%) (Table 3).

3.2 | WGD Events Illuminate the Evolutionary History of *P. Zhennan* and Lauraceae

Whole-genome duplication (WGD) events can act as key drivers of species evolution (Soltis and Soltis 2016; Jiao 2018). We

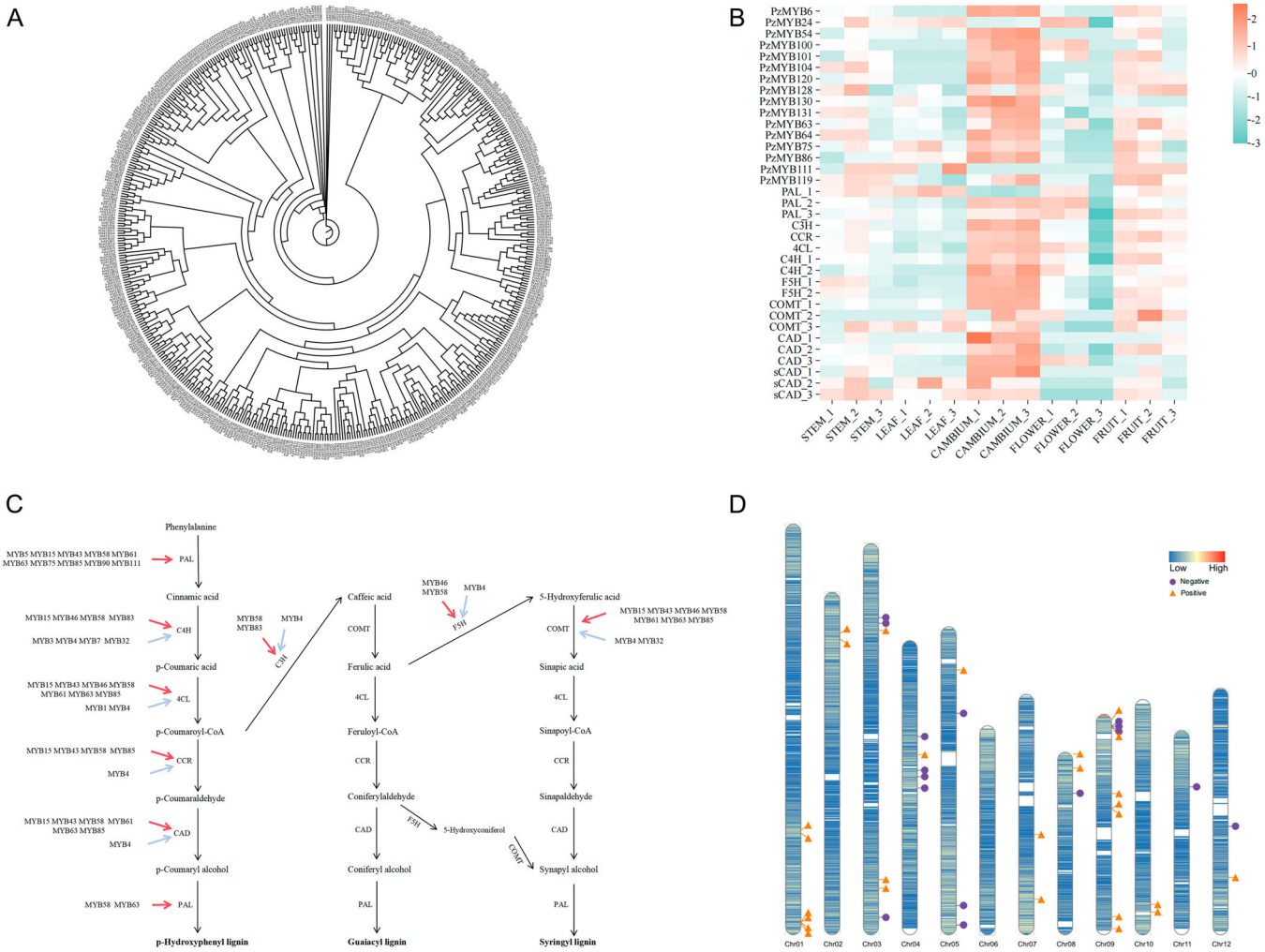


FIGURE 4 | Evolutionary analysis of phenylpropanoid lignin biosynthesis related MYB genes in *Phoebe zhennan*. (A) Maximum likelihood tree of the MYB gene family in *Arabidopsis thaliana*, *Populus trichocarpa* and *P. zhennan*. (B) Expression levels of lignin biosynthesis-related MYB TFs and genes encoding enzymes in the lignin biosynthesis in different tissues of *P. zhennan*.(C) Phenylpropanoid lignin biosynthesis pathway (Xiao et al. 2021). 4CL, 4-coumarate; CoA ligase; C3H, cinnamate 3-hydroxylase; C4H, cinnamate 4-hydroxylase; CAD, cinnamyl alcohol dehydrogenase; CCR, cinnamoyl CoA reductase; COMT, caffeate O-methyltransferase; F5H, ferulate 5-hydroxylase (D) Chromosomal localization of lignin biosynthesis-related MYB TFs in *P. zhennan*. Most genes are present in tandem repeats.

detected two WGD events in the *P. zhennan* genome (Figure 2A,B). By comparing the *P. zhennan* genome with those of other magnoliid species and constructing synteny maps, we demonstrated that the older WGD event is shared between the common ancestors of Magnoliales and Laurales, but not among all members of the magnoliid clade (Figure 2).

Earlier studies, constrained by incomplete genomic resources and sparse taxon sampling, have sometimes led to inconsistent interpretations of polyploidization history within lineages due to limited sample sizes available for comparison. For instance, Tian et al. (2023) suggested that both WGD events in *Lindera megaphylla* occurred before the divergence of Lauraceae members, while Jiao et al. (2011) confirmed that multiple ancient polyploidization events took place during the Cretaceous period (100–120 mya), including the λ event shared by the common ancestor of Magnoliales and Laurales.

The more recent WGD event in *P. zhennan* likely occurred before the divergence of Lauraceae, suggesting it is shared by all members of the family—a conclusion supported by multiple studies (Chaw et al. 2019; Chen, Sun, et al. 2020; Chen, Li, et al. 2020; Zhang et al. 2022; Li et al. 2022; Shen et al. 2022; Jiang et al. 2022). However, alternative explanations should also be considered. The observed Ks peak may be influenced by multiple temporally clustered lineage-specific duplications, which are challenging to distinguish from a single, shared WGD based on Ks data alone. Furthermore, the possibility that some Lauraceae lineages experienced additional, undetectable WGD events cannot be ruled out. Most Lauraceae members appear to have undergone two WGD events, and among currently reported genomes, *Lindera glauca* is the only species to have experienced a single WGD event after the divergence of Lauraceae (Xiong et al. 2022).

3.3 | Controversy on the Phylogenetic Position of Magnoliids

Phylogenetic analysis based on a set of single-copy orthologous genes indicates that magnoliids and monocots share a common

TABLE 3 | Comparison of genome assembly and annotation data in Lauraceae.

Species Name	Genome size (Mb)	Contig N50 (Mb)	Scalffold N50 (Mb)	GC content (%)	Number of coding genes	Repeat gene ratio (%)	LTR ratio (%)	Assembly completeness (%)
Phoebe zhennan	979.3	10.52	87.69	40.5	26262	53.99	34.97	98.9
Phoebe bournei	941.8	—	97.01	—	30096	48.10	26.34	92.1
Persea americana	913.0	7.3	26.0	38.1	42769	62.09	—	98.9
Camphora longepaniculata	766.7	35.81	63.78	38.8	39173	60.39	31.96	99.0
Camphora officinarum (= *Cinnamomum camphora*)	670.3	2.40	60.19	—	29919	53.97	27.66	95.2
Camphora micrantha (= *Cinnamomum kanehirae*)	730.7	0.90	50.4	—	27899	48.00	25.53	89.0
Cinnamomum burmannii	1177.6	1.13	94.90	39.0	41549	—	40.59	89.7
Cinnamomum chago	1061.2	—	92.10	39.4	28681	67.73	40.71	99.0
Lindera glauca	2092.2	0.71	186.54	39.6	65145	—	35.53	94.2
Lindera megaphylla	1268.6	2.61	104.72	39.4	32586	66.98	—	90.7
Litsea coreana	1139.5	4.94	97.18	—	32445	57.29	41.52	94.0
Litsea cubeba	1325.7	0.61	1.76	—	31329	—	47.64	88.4
Sextonia rubra	991.0	16.51	60.46	38.5	—	63.8	23.30	98.9

ancestor and form sister clades following their divergence from other angiosperm lineages (Figure 3A). Our phylogenomic results, which resolve magnoliids as sister to monocots, contribute to the growing consensus on one of the most enduring questions in angiosperm phylogeny. This placement finds strong support in the most comprehensive nuclear phylogenomic study of angiosperms to date, by Zuntini et al. (2024), which—using a standardized set of 353 nuclear genes across nearly 8000 genera—also recovers magnoliids (including Chloranthales) as a sister clade to monocots.

Further corroboration comes from genomic structural analyses. Qin et al. (2021) identified a shared chromosomal translocation event in magnoliids and monocots that is absent in eudicots, supporting the inference that magnoliids and monocots are sister groups, while eudicots diverged earlier. Similarly, Zhao et al. (2021) developed a novel microsynteny-based phylogenetic approach and, using over 120 high-quality genomes spanning more than 50 families and 30 orders, also supported a sister-group relationship between magnoliids and monocots.

This result is inconsistent with the studies of Chen, Sun, et al. (2020) and Han et al. (2022) on the phylogenetic position of magnoliids. Their findings indicated that magnoliids were a sister clade to eudicots rather than monocots. Chen, Sun, et al. (2020) analyzed phylogenetic trees constructed via ASTRAL and concatenated methods based on nucleotide sequences, which supported the hypothesis that magnoliids and eudicots share a recent common ancestor, forming a clade sister to monocots. Han et al. (2022) investigated the phylogenetic relationships among 18 seed plant species using a coalescence-based maximum-likelihood method to construct a phylogenetic tree. It is noteworthy that neither study included sampling from Chloranthales or Ceratophyllales—two key lineages whose inclusion can substantially affect tree topology—likely contributing to the observed inconsistencies in results.

Other studies have reported varying placements of magnoliids. Some position magnoliids at the base of the monocot–eudicot evolutionary branch (Hu et al. 2019; Rendón-Anaya et al. 2019; Chen et al. 2019; Jiang et al. 2022; Zhang et al. 2022), whereas others suggest that magnoliids form a sister clade with eudicots, implying that their phylogenetic relationship is closer (Chaw et al. 2019; Chen, Li, et al. 2020; Guo et al. 2021; Ma et al. 2021; Yang et al. 2020; Shen et al. 2022; Xiong et al. 2022). Such inconsistencies in topological structure are likely attributed to incomplete lineage sorting (ILS) and ancient hybridization events. Chen et al. (2020a) performed an ASTRAL analysis, indicating that ILS events could have occurred during the rapid early divergence of angiosperms. Ma et al. (2021) further analyzed the theta parameter across five major angiosperm lineages and found higher theta values in core eudicots and Ceratophyllales (a sister group to magnoliids). This suggests that incomplete lineage sorting (ILS) played a significant role in the evolutionary history of core angiosperms. Despite these complexities, the sister relationship between magnoliids and Ceratophyllales remains relatively stable—a finding consistent with the phylogeny we established in our study.

3.4 | Analysis of MYB Gene Families Related to Lignin Biosynthesis in *P. Zhennan*

Before this study, chromosome-level genome assemblies for *Phoebe* species were only available for *P. bournei*. These studies relied on the *P. bournei* genome to resolve the phylogenetic position of magnoliids and clarify the internal phylogeny of the Lauraceae family. They also conducted in-depth analyses of gene families in *P. bournei*, including MADS-BOX, GH3, SAUS, *MYB46, EPSP*, and TPS. Chen et al. (2020a) proposed that *MYB46* may be responsible for the tall, straight growth form of *P. bournei*, a hypothesis we further analyzed in this study.

Based on the reported phenylpropanoid lignin biosynthesis pathway in *Arabidopsis thaliana* (Geng et al. 2020; Xiao et al. 2021; Zhang et al. 2021), we identified both the MYB gene family and key structural genes involved in lignin biosynthesis in *P. zhennan* and constructed a co-expression network. A total of 16 MYB genes were located in the most significantly correlated expression module. Among them, nine were homologous to known activators and seven to repressors, suggesting a dynamic balance between activation and repression in lignin synthesis regulation. Notably, *PzMYB6* was identified as a direct homolog of *AtMYB46/83*, two key transcription factors in the second layer of regulation for secondary cell wall biosynthesis (Xiao et al. 2021). In *P. zhennan*, *PzMYB6* exhibited high expression in the vascular cambium (Figure 4B), implying its pivotal role in initiating lignification and secondary wall formation. We propose that *PzMYB6* functions as a master regulator in the lignin regulatory network of *P. zhennan*, potentially serving as the uppermost regulator that initiates secondary cell wall thickening and wood formation through direct coordination of downstream biosynthetic genes.

We also identified five genes (*PzMYB63, 64, 75, 86, and 111*) with homology to *AtMYB4*, a well-characterized repressor known to inhibit the biosynthesis of UV-protective phenolic compounds, such as coumarin derivatives (Jin et al. 2000; Zhou et al. 2009). The expansion of this gene suggests that the *MYB4*-mediated repressive mechanism is notably amplified and reinforced in *P. zhennan*. Furthermore, the presence of *PzMYB119*, a homolog of the repressor *AtMYB3*, corroborates this finding. *AtMYB3* and *AtMYB4* in *Arabidopsis thaliana* constitute a subfamily of transcriptional repressors (Kranz et al. 1998), indicating that a complex, multi-gene inhibitory module is involved in fine-tuning lignin biosynthesis in *P. zhennan*. *AtMYB58* is a direct target of *AtMYB46* and acts as a second-tier activator specifically dedicated to lignin biosynthesis (Zhong and Ye 2012). As a homolog of *AtMYB58*, *PzMYB100* is a potential downstream specialist activator within the *PzMYB6*-regulated hierarchy.

Based on stringent homology analysis, we propose a preliminary hierarchical model for the lignin regulatory network in *P. zhennan*: *PzMYB6* acts as the top-level regulator; secondary activators such as *PzMYB100* operate downstream; and a substantial repressor module comprising at least six genes (five *MYB4*-like and one *MYB3*-like) provides negative regulation. This dual architecture of "potent activation coupled with robust repression" likely enables *P. zhennan* to regulate this critical biological process with high efficiency and precision. While our

data suggest that these genes are potential candidates for regulating lignin biosynthesis, their proposed roles as putative regulators require validation by future functional studies, such as qRT-PCR, Y1H/dual-LUC, and CRISPR knockouts.

3.5 | Conservation Implications of the *P. Zhennan* Genome Resource

This study presents the first chromosome-level genome assembly of *P. zhennan* and corresponding transcriptomic data from five key tissues, offering insights into both genomic architecture and multi-tissue expression patterns. These resources provide essential genetic foundations for the conservation and sustainable use of this valuable and endangered tree species. As a rare timber tree characterized by slow growth and highly valued wood, *P. zhennan* faces severe threats from overlogging and habitat fragmentation. The reference genome generated here serves as a fundamental genetic map for future population genetics research, delineation of conservation units, assessment of genetic diversity, and the development of artificial breeding strategies.

Comparative genomic analyses identified two WGD events in *P. zhennan*, accompanied by lineage-specific gene family evolution. These evolutionary events likely contributed to its distinctive wood properties, including aromaticity, decay resistance, and the unique "golden thread" figure. The study identified 44 MYB genes linked to lignin biosynthesis and proposed a preliminary regulatory network, providing novel insights into secondary growth and actionable targets for molecular breeding. These resources enable marker-assisted selection and gene-editing strategies to improve wood traits and accelerate breeding—similar to how the chromosome-level genome of *Dalbergia odorifera* has been used to identify genes involved in heartwood formation, thereby facilitating marker-assisted selection for improved wood quality and accelerated breeding programs (Hong et al. 2020).

In the context of conservation genomics, the value of such genomic resources has been demonstrated in other endangered trees. For example, the genomes of *Taxus chinensis* var. *mairei* and *Aquilaria yunnanensis* have been used to identify biosynthetic genes involved in taxol and agarwood formation, respectively, aiding in the functional characterization of adaptive traits and informing conservation strategies (Xiong et al. 2021; Li et al. 2024). Building on these precedents, the *P. zhennan* genome will facilitate genetic diversity surveys of wild populations, the design of core germplasm collections, and the mining of functional genes related to stress resilience and wood formation. Ultimately, this genome will transition from passive protection to active, genomics-informed precision conservation and sustainable resource management.

In summary, the genomic resources presented in this study hold significant value not only for elucidating the genome evolution of Lauraceae but also for supporting genetic conservation, resource restoration, and sustainable use of *P. zhennan*. Future research utilizing this genome resource could include genetic diversity surveys of wild resources, construction of a core germplasm collection, and functional gene mining—ultimately advancing precision conservation and sustainable management of *P. zhennan* resources.

4 | Methods

4.1 | Plant Material and Genome Sequencing

Plant material of *P. zhennan* was collected from wild individuals on Emei Mountain, Sichuan Province, China (Figure 1A). Leaf samples from one individual were collected and preserved in liquid nitrogen for genome sequencing, while tissue samples (leaves, current-year branches, cambium, flowers, and young fruits) from three individuals were similarly preserved for transcriptome sequencing.

Total genomic DNA was extracted from leaf samples. HiFi libraries were constructed and sequenced on the PacBio platform. A single 20 kb library was prepared, and sequencing was performed on the PacBio Sequel II platform in CCS mode. A Hi-C library was constructed and sequenced using the DNBseq platform. Total RNA was extracted from tissue samples, and multi-throughput Iso-seq library preparation and sequencing were performed on the PacBio platform to obtain transcriptomic data. All DNA/RNA extraction and sequencing procedures were completed by BGI Genomics.

4.2 | Genome Survey

Given the substantial variation in genome heterozygosity and repetitive content among organisms—critical factors influencing assembly—we performed pre-sequencing genomic characterization. Clean reads were generated from raw data using SOAPnuke v2.1.0 (Chen et al. 2011) with the following parameters: -n 0.02 -l 20 -q 0.4 -i -G 2 –polyX 50 -Q 2 –seqType 0. GenomeScope v1.0 (Vurture et al. 2017) analyzed 21-bp K-mer frequencies from Jellyfish v2.2.6 (Marçais and Kingsford 2011) output to estimate genome size, heterozygosity rate, repetitive content, and sequencing depth.

4.3 | Genome Assembly and Quality Control

The PacBio HiFi data were filtered using CCS 6.2.0 (–minPasses 3 –min-rq 0.99 –min-length 500) and subsequently corrected and preliminarily assembled with Hifiasm v0.151 (Cheng et al. 2021) to obtain the genome contig sequences. Juicer v1.5.6 (Durand et al. 2016) was used to align paired-end Hi-C sequencing data to the assembled genome, and the proportion of valid Hi-C contacts uniquely aligned to the genome was calculated to evaluate the Hi-C library. Reads were aligned to the genome using BWA (Li and Durbin 2010), and alignment results were filtered to remove unaligned reads and those lacking correct cross-linking information. Duplicate reads were also removed. The Hi-C data were used to assemble the genome at the chromosomal level, with the genome divided into equal-length bins. The number of Hi-C read pairs between any two bins was used as the interaction strength signal, and a heatmap was generated accordingly. The completeness and redundancy of the assembled genome were evaluated using BUSCO (Simão

et al. 2015), which assesses genome completeness based on single-copy orthologous genes (orthologs) across species. A single-copy ortholog gene set was constructed based on the embryophyta_odb10 database, and the assembled genome was compared with this gene set. The accuracy and completeness of the assembly were evaluated based on the alignment ratio, completeness, and accuracy of the comparison.

4.4 | Genome Annotation and Functional Prediction

Homology-based repeat identification was conducted using the RepBase database (Bao et al. 2015). RepeatMasker (Tarailo-Graovac and Chen 2009) and RepeatProteinMask were utilized to identify and classify sequences homologous to known repeats. RepeatModeler v1.0.4 and LTRharvest (Ellinghaus et al. 2008) were employed to construct de novo repeat libraries, which were subsequently utilized by RepeatMasker for repeat prediction. Tandem Repeats Finder (Benson 1999) was used to detect tandem repeats. The abundance of each repeat type was calculated as the percentage of the genome sequence masked by that specific repeat class using RepeatMasker. We conducted homology-based predictions using annotation information from eight related species (*Aquilegia coerulea*, *Arabidopsis thaliana*, *Cinnamomum camphora*, *Daucus carota*, *Ginkgo biloba*, *Liriodendron chinense*, *Mimulus guttatus*, *Phoebe bournei*). For de novo predictions, we used Augustus (Stanke et al. 2006). RNA-Seq and full-length transcript ISOform data were aligned to the genome using HISAT2 (Kim et al. 2019) and GMAP (Wu and Watanabe 2005), respectively, and transcript assembly annotations were performed using StringTie (Pertea et al. 2015). The combined results were then integrated with homologous species gene sets using GeMoMa (Keilwagen et al. 2018) for homology annotation. Subsequently, these annotations were combined using EVM (Evidence Modeler) and compared with transposable elements (TE), retaining only genes with less than 80% overlap with TE sequences. Finally, all annotation evidence was further refined using PASA (Haas et al. 2003), resulting in the final gene set. Finally, the quality of the final annotation was assessed using BUSCO (Simão et al. 2015). To functionally annotate the gene set derived from structural annotation, DIAMOND (Buchfink et al. 2021) was employed for sequence alignment against known protein databases. Additionally, InterProScan (Zdobnov and Apweiler 2001) was utilized to identify secondary structure domains across multiple databases, including SwissProt (Boeckmann et al. 2003), TrEMBL (Boeckmann et al. 2003), KEGG (Kanehisa and Goto 2000), InterPro, NR, KOG (Koonin et al. 2004), and GO (Ashburner et al. 2000).

4.5 | Whole Genome Duplication Analysis

We used WGDI 0.6.5 (Sun, Jiao, et al. 2022) to analyze whole genome duplication (WGD) events in the *P. zhennan* genome. First, we generated a dot plot of the *P. zhennan* genome using the -dotplot parameter to visually examine the initial level of duplication. Based on this, we used the -ck command to identify syntenic blocks within the *P. zhennan* genome, and the -ks command to calculate the Ks/Ka values for each syntenic block. Subsequently, we applied the -pf parameter to fit the Ks distribution curve values and generate a peak plot. By observing the number of peaks, we determined the number of WGD events in the *P. zhennan* genome.

We also used the MCScanX v1.0.0 (Wang et al. 2012) and JCVI v1.5.7 (Tang et al. 2024) to identify syntenic blocks in the genomes of *Amborella trichopoda*, *Phoebe bourinei*, and *P. zhennan*. The alignment of the gene families was performed with MUSCLE v3.8.1551 (Edgar 2004), and Ks estimates were generated with the CODEML (Jeffares et al. 2015) program of the PAML through maximum likelihood estimation. Genome-wide syntenic and collinear blocks were identified both within and among the selected genomes. Based on the results of all vs. all BLASTp comparisons (E-value $\leq 1e{-}5$), analyses were conducted on the translated protein sequences of the six annotated gene models to establish a protein similarity database. WGDI was then used to identify conserved collinear blocks. The peaks of WGD events were fitted based on the Ks distribution. The Ks frequency distribution plot was then used to identify peaks, allowing the determination of the number and timing of WGD events. The -pf parameter was used for Gaussian fitting of the peaks in the Ks frequency distribution plot, while the -kf parameter was used to visualize the fitting results.

4.6 | Phylogenomic Analysis

To investigate the evolutionary background of *P. zhennan*, we obtained the genomes of twenty different species from publicly accessible databases, including six Lauraceae species (*Phoebe bournei* (Chen et al. 2020), *Persea americana* (Rendón-Anaya et al. 2019), *Cinnamomum camphora* (Li et al. 2023), *Cinnamomum kanehirae* (Chaw et al. 2019), *Cinnamomum chago* (Tao et al. 2024), *Lindera megaphylla* (Tian et al. 2023), seven non-Lauraceae Magnoliids (*Chimonanthus salicifolius* (Lv et al. 2020)), *Magnolia sinica* (Cai et al. 2024), *Annona cherimola* (Talavera et al. 2023), *Piper nigrum* (Hu et al. 2019), *Saururus chinensis* (Xue et al. 2023), *Aristolochia fimbriata* (Qin et al. 2021)), *Chloranthus sessilifolius* (Ma et al. 2021), two basal angiosperms (*Nymphaea colorata* (Zhang et al. 2020), *Amborella trichopoda* (Amborella Genome Project et al. 2013)), two core eudicots (*Arabidopsis thaliana*, *Davidia involucrata* (Chen and Ma 2020)), *Ceratophyllum demersum* (Yang et al. 2020), and two monocots (*Acorus calamus* (Ma et al. 2023), *Areca catechu* (Zhou et al. 2022)).

Orthologous gene groups were inferred using OrthoFinder v2.5.4 (Emms and Kelly 2019). An all-versus-all sequence search was performed using DIAMOND (Buchfink et al. 2021) to identify potential homologous genes, and the identified homologs were assigned to orthogroups using the Markov Cluster Algorithm (MCL). To construct the species tree, orthogroups containing single-copy orthologs were selected. These single-copy genes were aligned using MAFFT v.7.525 (Katoh and Standley 2013) for multiple sequence alignment (MSA), and the corresponding CDS sequences were further aligned using PAL2NAL v.14 (Suyama et al. 2006) to account for codon variation. Based on the alignments, gene trees for each single-copy gene were inferred using FastTree v2.2 (Price et al. 2009).

The species tree was reconstructed using the STRIDE (Species Tree Inference from Duplication Events) algorithm, which integrates the topologies of single-copy gene trees to generate the final phylogeny. We employed a Bayesian relaxed molecular clock approach to estimate the divergence times using MCMCTree (Reis and Yang 2011) implemented in PAML v.4.9 (Yang 2007). Four fossil calibration points were set for divergence time estimation, acquired from the TimeTree database (http://www.timetree.org): between *Magnolia sinica* and *Annona cherimola* (95.0–113.6 mya), between *Areca catechu* and *Acorus calamus* (133.0–139.1 mya), between *Piper nigrum* and *Saururus chinensis* (64.1–94.3 mya), and between *Davidia involucrata* and *Arabidopsis thaliana* (111.4–123.9 mya). Gene families in phylogenetic tree-constructing species that had experienced expansion or contraction were identified using CAFE v5 (Mendes et al. 2021).

4.7 | Identification of MYB Gene Family and Enzymatic Genes in the Lignin Biosynthesis Pathway

First, we downloaded the HMM models for the MYB gene family domains (PF00249, PF08914, PF12776, PF13837, PF13873, PF13921, PF15963) from the Pfam database (http://pfam.xfam.org/). Then, we used HMMER v3.3.2 (Eddy 1998) to conduct a preliminary identification and screening of MYB gene family members in the *P. zhennan* genome. Next, we used the NCBI Conserved Domain Database (CDD) tools (https://www.ncbi.nlm.nih.gov/) to further screen the candidate genes. The domain composition of each gene was checked in the result files, and genes without MYB domains were removed. The MYB gene family sequences from *Arabidopsis thaliana* were downloaded from TAIR (https://www.arabidopsis.org/). The MYB gene family member sequences from *Arabidopsis thaliana*, *Populus trichocarpa*, and *P. zhennan* were aligned using MAFFT v7.525 (Katoh and Standley 2013), and a phylogenetic tree was constructed using the maximum likelihood method with IQTREE v2.1.4-beta (Nguyen et al. 2015). The best-fit model (JTT + R7) was selected according to the Bayesian Information Criterion (BIC). The same identification workflow was also applied to annotate the key enzymatic genes involved in the monolignol biosynthesis pathway.

4.8 | Gene Expression Analysis of MYB TFs and Enzymatic Genes in the Lignin Biosynthesis Pathway and WGCNA

RNA-seq data for the five tissues were obtained as preprocessed data from BGI, which included adapter trimming and quality filtering according to their standard pipeline. Upon receipt, we conducted quality assessment on the provided clean reads using FastQC v0.11.9 (https://github.com/s-andrews/FastQC) to verify their high quality.

Subsequent analysis was performed as follows: First, we built the index and aligned the sequences using HISAT2 v2.2.1 (Kim et al. 2019). Then, we used HTSeq v2.0.8 (Putri et al. 2022) to count the number of reads mapped to each gene for quantifying gene expression levels. The resulting count data were merged, and differential expression analysis was performed using the DESeq. 2 (Love et al. 2014). The expression data were filtered to focus on MYB TFs related to phenylpropanoid lignin biosynthesis. Finally, expression levels of these selected MYB TFs were visualized using a heatmap generated by the pheatmap package in R.

To further investigate the co-expression relationships among MYB transcription factors and structural genes involved in lignin biosynthesis, we performed a WGCNA using the R package WGCNA v1.73 (Langfelder and Horvath 2008). Gene expression matrices for both MYB TFs and pathway structural genes were merged and transposed to construct the input matrix. After ensuring data quality and filtering out poorly expressed genes and outliers through sample clustering, we selected an appropriate soft-thresholding power (power = 6) based on the scale-free topology criterion to construct a signed co-expression network. Module identification was carried out using the blockwiseModules function. Modules were labeled with color names for further analysis.

Author Contributions

Mao Yameng: writing – original draft, writing – review and editing; conceptualization, formal analysis, investigation, visualization. **Zou Dianyang:** writing – review and editing, investigation. **Li Lang:** writing – review and editing, conceptualization, data curation, investigation, project administration, supervision, funding acquisition. **Li Jie:** writing – review and editing, supervision, funding acquisition, resources.

Acknowledgments

We would like to express our sincere gratitude to the three anonymous reviewers for their insightful comments and valuable suggestions, which have significantly improved this manuscript. This study was supported by the National Natural Science Foundation of China (31970222) and the 14th 5-Year Plan of the Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (XTBG-1450101).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in NGDC at https://ngdc.cncb.ac.cn/, reference number PRJCA048104.

References

Ashburner, M., C. A. Ball, J. A. Blake, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25, no. 1: 25–29. https://doi.org/10.1038/75556.

Amborella Genome Project, V. A. Albert, W. B. Barbazuk, et al. 2013. "The Amborella Genome and the Evolution of Flowering Plants." *Science* 342: 1241089. https://doi.org/10.1126/science.1241089.

Bao, W., K. K. Kojima, and O. Kohany. 2015. "Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes." *Mobile DNA* 6, no. 1: 11. https://doi.org/10.1186/s13100-015-0041-9.

Benson, G. 1999. "Tandem Repeats Finder: A Program to Analyze DNA Sequences." *Nucleic Acids Research* 27, no. 2: 573–580. https://doi.org/10.1093/nar/27.2.573.

Boeckmann, B., A. Bairoch, R. Apweiler, et al. 2003. "The SWISS-PROT Protein Knowledgebase and Its Supplement Trembl in 2003." *Nucleic Acids Research* 31, no. 1: 365–370. https://doi.org/10.1093/nar/gkg095.

Buchfink, B., K. Reuter, and H. G. Drost. 2021. "Sensitive Protein Alignments at Tree-Of-Life Scale Using DIAMOND." *Nature Methods* 18, no. 4: 366–368. https://doi.org/10.1038/s41592-021-01101-x.

Cai, L., D. Liu, F. Yang, et al. 2024. "The Chromosome-Scale Genome of Magnolia Sinica (Magnoliaceae) Provides Insights Into the Conservation of Plant Species With Extremely Small Populations (PSESP)." *GigaScience* 13: giad110. https://doi.org/10.1093/gigascience/giad110.

Chaw, S. M., Y. C. Liu, Y. W. Wu, et al. 2019. "Stout Camphor Tree Genome Fills Gaps in Understanding of Flowering Plant Genome Evolution." *Nature Plants* 5, no. 1: 63–73. https://doi.org/10.1038/s41477-018-0337-0.

Chen, F., D. Tholl, J. Bohlmann, and E. Pichersky. 2011. "The Family of Terpene Synthases in Plants: A Mid-Size Family of Genes for Specialized Metabolism That Is Highly Diversified Throughout the Kingdom." *Plant Journal* 66, no. 1: 212–229. https://doi.org/10.1111/j.1365-313X.2011.04520.x.

Chen, J., Z. Hao, X. Guang, et al. 2019. "Liriodendron Genome Sheds Light on Angiosperm Phylogeny and Species–Pair Differentiation." *Nature Plants* 5, no. 1: 18–25. https://doi.org/10.1038/s41477-018-0323-6.

Chen, S. P., W. H. Sun, Y. F. Xiong, et al. 2020. "The Phoebe Genome Sheds Light on the Evolution of Magnoliids." *Horticultural Research* 7: 146. https://doi.org/10.1038/s41438-020-00368-z.

Chen, Y., and T. Ma. 2020. "Genomic Analyses of a "living fossil": The Endangered Dove-Tree." *Molecular Ecology Resources* 20, no. 3: 617–840. https://doi.org/10.1111/1755-0998.13138.

Chen, Y. C., Z. Li, Y. X. Zhao, et al. 2020. "The Litsea Genome and the Evolution of the Laurel Family." *Nature Communications* 11, no. 1: 1675. https://doi.org/10.1038/s41467-020-15493-5.

Chen, Y. Z. 2013. "Cultural History and Modern Life of Gold-Thread Nanmu." *Light Industry Science and Technology* 12: 96–114.

Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li. 2021. "Haplotype-Resolved De Novo Assembly Using Phased Assembly Graphs With Hifiasm." *Nature Methods* 18, no. 2: 170–175. https://doi.org/10.1038/s41592-020-01056-5.

Cheng, W. C. 1983. "Lauraceae." In *Sylva of China*, edited by W. C. Cheng, vol. 1, 606–877. China Forestry Publishing Press.

Ding, X., J. H. Xiao, J. F. Huang, L. Li, and J. Li. 2015. "Investigation on Wild Resource of Valuable Timber Tree *Phoebe zhennan*." *Plant Divers* 37, no. 5: 629. https://doi.org/10.7677/ynzwyj201515002.

Ding, X., J. H. Xiao, L. Li, J. G. Conran, and J. Li. 2019. "Congruent Species Delimitation of Two Controversial Gold-Thread Nanmu Tree Species Based on Morphological and Restriction Site-Associated DNA Sequencing Data." *Journal of Systematics and Evolution* 57, no. 3: 234–246. https://doi.org/10.1111/jse.12433.

Dubos, C., R. Stracke, E. Grotewold, B. Weisshaar, C. Martin, and L. Lepiniec. 2010. "MYB Transcription Factors in Arabidopsis." *Trends in Plant Science* 15, no. 10: 573–581. https://doi.org/10.1016/j.tplants.2010.06.005.

Durand, N. C., M. S. Shamim, I. Machol, et al. 2016. "Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments." *Cell Systems* 3, no. 1: 95–98. https://doi.org/10.1016/j.cels.2016.07.002.

Eddy, S. R. 1998. "Profile Hidden Markov Models." *Bioinformatics* 14, no. 9: 755–763. https://doi.org/10.1093/bioinformatics/14.9.755.

Edgar, R. C. 2004. "MUSCLE: Multiple Sequence Alignment With High Accuracy and High Throughput." *Nucleic Acids Research* 32, no. 5: 1792–1797. https://doi.org/10.1093/nar/gkh340.

Ellinghaus, D., S. Kurtz, and U. Willhoeft. 2008. "LTRharvest, an Efficient and Flexible Software for De Novo Detection of LTR Retrotransposons." *BMC Bioinformatics* 9, no. 1: 18. https://doi.org/10.1186/1471-2105-9-18.

Emms, D. M., and S. Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20, no. 1: 238. https://doi.org/10.1186/s13059-019-1832-y.

Forestry, S., G. Administration, M. of Agriculture, and R. A. of China. 2021. *List of National Key Protected Wild Plants (First Batch)*. State Forestry and Grassland Administration and Ministry of Agriculture and Rural Affairs of China.

Geng, P., S. Zhang, J. Liu, et al. 2020. "MYB20, MYB42, MYB43, and MYB85 Regulate Phenylalanine and Lignin Biosynthesis During Secondary Cell Wall Formation." *Plant Physiology* 182, no. 3: 1272–1283. https://doi.org/10.1104/pp.19.01070.

Guo, X., D. Fang, S. K. Sahu, et al. 2021. "Chloranthus Genome Provides Insights Into the Early Diversification of Angiosperms." *Nature Communications* 12, no. 1: 6930. https://doi.org/10.1038/s41467-021-26922-4.

Haas, B. J., A. L. Delcher, S. M. Mount, et al. 2003. "Improving the Arabidopsis Genome Annotation Using Maximal Transcript Alignment Assemblies." *Nucleic Acids Research* 31, no. 19: 5654–5666. https://doi.org/10.1093/nar/gkg770.

Han, X., J. Zhang, S. Han, et al. 2022. "The Chromosome-Scale Genome of Phoebe *bournei* Reveals Contrasting Fates Ofterpene Synthase (TPS)-A and TPS-b Subfamilies." *Plant Communications* 3, no. 6: 100410. https://doi.org/10.1016/j.xplc.2022.100410.

Hong, Z., J. Li, X. Liu, et al. 2020. "The Chromosome-Level Draft Genome of *Dalbergia odorifera*." *GigaScience* 9, no. 8: giaa084.

Hu, L., Z. Xu, M. Wang, et al. 2019. "The Chromosome-Scale Reference Genome of Black Pepper Provides Insight Into Piperine Biosynthesis." *Nature Communications* 10, no. 1: 4702. https://doi.org/10.1038/s41467-019-12607-6.

Huang, J. H., J. H. Chen, J. S. Ying, and K. P. Ma. 2011. "Features and Distribution Patterns of Chinese Endemic Seed Plant Species." *Journal of Systematics and Evolution* 49, no. 2: 81–94. https://doi.org/10.1111/j.1759-6831.2011.00119.x.

Jeffares, D. C., B. Tomiczek, V. Sojo, and M. dos Reis. 2015. "A Beginners Guide to Estimating the Non-synonymous to Synonymous Rate Ratio of All Protein-Coding Genes in a Genome." In *Parasite Genomics Protocols. Vol. 1201 of Methods in Molecular Biology*, edited by C. Peacock, 65–90. NYHumana Press, Springer.

Jiang, R., X. Chen, X. Liao, et al. 2022. "A Chromosome-Level Genome of the Camphor Tree and the Underlying Genetic and Climatic Factors for Its Top-Geoherbalism." *Frontiers in Plant Science* 13: 827890. https://doi.org/10.3389/fpls.2022.827890.

Jiao, Y. 2018. "Double the Genome, Double the Fun: Genome Duplications in Angiosperms." *Molecular Plant* 11, no. 3: 357–358. https://doi.org/10.1016/j.molp.2018.02.009.

Jiao, Y., N. J. Wickett, S. Ayyampalayam, et al. 2011. "Ancestral Polyploidy in Seed Plants and Angiosperms." *Nature* 473, no. 7345: 97–100. https://doi.org/10.1038/nature09916.

Jin, H., E. Cominelli, P. Bailey, et al. 2000. "Transcriptional Repression by AtMYB4 Controls Production of UV-Protecting Sunscreens in Arabidopsis." *EMBO Journal* 19, no. 22: 6150–6161. https://doi.org/10.1093/emboj/19.22.6150.

Kanehisa, M., and S. Goto. 2000. "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 28, no. 1: 27–30. https://doi.org/10.1093/nar/28.1.27.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30, no. 4: 772–780. https://doi.org/10.1093/molbev/mst010.

Keilwagen, J., F. Hartung, M. Paulini, S. O. Twardziok, and J. Grau. 2018. "Combining RNA-Seq Data and Homology-Based Gene Prediction for Plants, Animals and Fungi." *BMC Bioinformatics* 19, no. 1: 189. https://doi.org/10.1186/s12859-018-2203-5.

Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. "Graph-Based Genome Alignment and Genotyping With HISAT2 and

HISAT Genotype." *Nature Biotechnology* 37, no. 8: 907–915. https://doi.org/10.1038/s41587-019-0201-4.

Koonin, E. V., N. D. Fedorova, J. D. Jackson, et al. 2004. "A Comprehensive Evolutionary Classification of Proteins Encoded Incomplete Eukaryotic Genomes." *Genome Biology* 5, no. 2: R7. https://doi.org/10.1186/gb-2004-5-2-r7.

Kranz, H. D., M. Denekamp, R. Greco, et al. 1998. "Towards Functional Characterisation of the Members of the R2R3-MYB Gene Family From *Arabidopsis thaliana*." *Plant Journal* 16, no. 2: 263–276. https://doi.org/10.1046/j.1365-313x.1998.00278.x.

Langfelder, P., and S. Horvath. 2008. "WGCNA: An R Package for Weighted Correlation Network Analysis." *BMC Bioinformatics* 9, no. 1: 559. https://doi.org/10.1186/1471-2105-9-559.

Lee, S. K., F. N. Wei, Y. T. Wei, and H. W. Li. 1979. "Materiae Ad Floram Lauracearum Sinicarum (III)." *Acta Phytotaxonomica Sinica* 17, no. 2: 45–74.

Li, D., H. Y. Lin, X. Wang, et al. 2023. "Genome and Whole-Genome Resequencing of Cinnamomum camphora Elucidate Its Dominance in Subtropical Urban Landscapes." *BMC Biology* 21, no. 1: 192. https://doi.org/10.1186/s12915-023-01692-1.

Li, D. L., Y. Q. Jin, and Q. B. Xiang. 2004. "The Geographical Distribution Research Status and Developmental Utilization Prospect of *Phoebe* Nees Plant Resource of Our Country." *Journal of Fujian Forestry Science and Technology* 31, no. 1: 5–9. https://doi.org/10.13428/j.cnki.fjlk.2004.01.002.

Li, F., S. Huang, Y. Mei, et al. 2022. "Genome Assembly Provided New Insights Into the *Cinnamomum burmannii* Evolution and D-Borneol Biosynthesis Differences Between Chemotypes." *Industrial Crops and Products* 186: 115181. https://doi.org/10.1016/j.indcrop.2022.115181.

Li, H., and R. Durbin. 2010. "Fast and Accurate Long-Read Alignment With Burrows-Wheeler Transform." *Bioinformatics* 26, no. 5: 589–595. https://doi.org/10.1093/bioinformatics/btp698.

Li, H. W., P. Y. Pai, and Y. R. Li, et al. 1982. *Lauraceae*, Vol. 31. Science Press.

Li, L., J. Li, and X. W. Li. 2011. "Taxonomic Revision of Five Species of the Genus Phoebe (Lauraceae) From China." *Plant Divers* 33, no. 2: 157. https://doi.org/10.3724/SP.J.1143.2011.10148.

Li, M., Y. Zhang, Y. Wang, Y. Yin, M. Zhou, and Y. Zhang. 2024. "Chromosome-Level Genome Assembly of *Aquilaria yunnanensis*." *Scientific Data* 11, no. 1: 790.

Li, X. W., J. Li, P. Huang, F. Wei, H. Cui, and H. van der Werff. 2008. "Lauraceae." In *Flora of China*, Vol. 7, edited by Z. Wu, P. Raven, and D. Hong, 102–254. LouisScience Press and Missouri Botanical Garden Press.

Li, L., B. Liu, Y. Song, et al. 2025. "Global Advances in Phylogeny, Taxonomy and Biogeography of Lauraceae." *Plant Diversity* 47, no. 3: 341–364. https://doi.org/10.1016/j.pld.2025.04.001.

Lin, H. R. 1988. "The Ancient Nanmu and Its Changing Distribution." *Forestry Science and Technology* 9, no. 4: 48–58. https://doi.org/10.16779/j.cnki.1003-5508.1988.04.016.

Liu, B., W. Y. Jin, L. N. Zhao, and Y. Yang. 2020. "A New Species of *Phoebe* (Lauraceae) From South-Western China." *PhytoKeys* 140: 101–106. https://doi.org/10.3897/phytokeys.140.47664.

Love, M. I., W. Huber, and S. Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With DESeq." *Genome Biology* 15, no. 12: 550. https://doi.org/10.1186/s13059-014-0550-8.

Lv, Q., J. Qiu, J. Liu, et al. 2020. "The *Chimonanthus salicifolius* Genome Provides Insight Into Magnoliid Evolution and Flavonoid Biosynthesis." *Plant Journal* 103, no. 5: 1910–1923. https://doi.org/10.1111/tpj.14874.

Ma, J., P. Sun, D. Wang, et al. 2021. "The Chloranthus Sessilifolius Genome Provides Insight Into Early Diversification of Angiosperms." *Nature Communications* 12, no. 1: 6929. https://doi.org/10.1038/s41467-021-26931-3.

Ma, L., K. W. Liu, Z. Li, et al. 2023. "Diploid and Tetraploid Genomes of Acorus and the Evolution of Monocots." *Nature Communication* 14: 3661. https://doi.org/10.1038/s41467-023-38829-3.

Marçais, G., and C. Kingsford. 2011. "A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of k-mers." *Bioinformatics* 27, no. 6: 764–770. https://doi.org/10.1093/bioinformatics/btr011.

Mendes, F. K., D. Vanderpool, B. Fulton, and M. W. Hahn. 2021. "CAFE 5 Models Variation in Evolutionary Rates Among Gene Families." *Bioinformatics* 36, no. 22/23: 5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022.

Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32, no. 1: 268–274. https://doi.org/10.1093/molbev/msu300.

Pertea, M., G. M. Pertea, C. M. Antonescu, T. C. Chang, J. T. Mendell, and S. L. Salzberg. 2015. "Stringtie Enables Improved Reconstruction of a Transcriptome From RNA-Seq Reads." *Nature Biotechnology* 33, no. 3: 290–295. https://doi.org/10.1038/nbt.3122.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2009. "FastTree: Computing Large Minimum Evolution Trees With Profiles Instead of a Distance Matrix." *Molecular Biology and Evolution* 26, no. 7: 1641–1650. https://doi.org/10.1093/molbev/msp077.

Putri, G. H., S. Anders, P. T. Pyl, J. E. Pimanda, and F. Zanini. 2022. "Analysing High-Throughput Sequencing Data in Python With HTSeq. 2.0." *Bioinformatics* 38, no. 10: 2943–2945. https://doi.org/10.1093/bioinformatics/btac166.

Qin, L., Y. Hu, J. Wang, et al. 2021. "Insights into Angiosperm Evolution, Floral Development and Chemical Biosynthesis From the *Aristolochia fimbriata* Genome." *Nature Plants* 7, no. 9: 1239–1253.

Reis, M., and Z. Yang. 2011. "Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times." *Molecular Biology and Evolution* 28, no. 7: 2161–2172. https://doi.org/10.1093/molbev/msr045.

Rendón-Anaya, M., E. Ibarra-Laclette, A. Méndez-Bravo, et al. 2019. "The Avocado Genome Informs Deep Angiosperm Phylogeny, Highlights Introgressive Hybridization, and Reveals Pathogen-Influenced Gene Space Adaptation." *Proceedings of the National Academy of Sciences* 116, no. 34: 17081–17089. https://doi.org/10.1073/pnas.1822129116.

Rohde, R., B. Rudolph, K. Ruthe, et al. 2017. "Neither Phoebe nor Cinnamomum–the Tetrasporangiate Species of Aiouea (Lauraceae)." *Taxon* 66, no. 5: 1085–1111. https://doi.org/10.12705/665.6.

Shen, T., H. Qi, X. Luan, et al. 2022. "The Chromosome-Level Genome Sequence of the Camphor Tree Provides Insights Into Lauraceae Evolution and Terpene Biosynthesis." *Plant Biotechnology Journal* 20, no. 2: 244–246. https://doi.org/10.1111/pbi.13749.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness With Single-Copy Orthologs." *Bioinformatics (Oxford, England)* 31, no. 19: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Soltis, P. S., and D. E. Soltis. 2016. "Ancient WGD Events as Drivers of Key Innovations in Angiosperms." *Current Opinion in Plant Biology* 30: 159–165. https://doi.org/10.1016/j.pbi.2016.03.015.

Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack, and B. Morgenstern. 2006. "AUGUSTUS: Ab Initio Prediction of Alternative Transcripts." Supplement, *Nucleic Acids Research* 34, no. S2: W435–W439. https://doi.org/10.1093/nar/gkl200.

Stracke, R., M. Werber, and B. Weisshaar. 2001. "The R2R3-MYB Gene Family in *Arabidopsis thaliana*." *Current Opinion in Plant Biology* 4, no. 5: 447–456. https://doi.org/10.1016/S1369-5266(00)00199-0.

Sun, P., B. Jiao, Y. Yang, et al. 2022. "WGDI: A User-Friendly Toolkit for Evolutionary Analyses of Whole-Genome Duplications and Ancestral Karyotypes." *Molecular Plant* 15, no. 12: 1841–1851. https://doi.org/10.1016/j.molp.2022.10.018.

Sun, P., H. Wang, P. Zhao, et al. 2022. "The Regulation of Xylem Development by Transcription Factors and Their Upstream MicroRNAs." *International Journal of Molecular Sciences* 23, no. 17: 10134. https://doi.org/10.3390/ijms231710134.

Suyama, M., D. Torrents, and P. Bork. 2006. "PAL2NAL: Robust Conversion of Protein Sequence Alignments Into the Corresponding Codon Alignments." Supplement, *Nucleic Acids Research* 34, no. S2: W609–W612. https://doi.org/10.1093/nar/gkl315.

Talavera, A., N. Fernandez-Pozo, A. J. Matas, J. I. Hormaza, and A. Bombarely. 2023. "Genomics in Neglected and Underutilized Fruit Crops: A Chromosome-Scale Genome Sequence of Cherimoya (Annona cherimola)." *Plants, People, Planet* 5, no. 3: 305–458. https://doi.org/10.1002/ppp3.10366.

Tang, H., V. Krishnakumar, X. Zeng, et al. 2024. "JCVI: A Versatile Toolkit for Comparative Genomics Analysis." *iMeta* 3, no. 4: e211. https://doi.org/10.1002/imt2.211.

Tao, L., S. Guo, Z. Xiong, R. Zhang, and W. Sun. 2024. "Chromosome-Level Genome Assembly of the Threatened Resource Plant *Cinnamomum chago*." *Scientific Data* 11: 447.

Tarailo-Graovac, M., and N. Chen. 2009. "Using Repeatmasker to Identify Repetitive Elements in Genomic Sequences." *Current Protocols in Bioinformatics* 25, no. 1: 4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.

Tian, X.-C., J.-F. Guo, X.-M. Yan, et al. 2023. "Unique Gene Duplications and Conserved Microsynteny Potentially Associated With Resistance to Wood Decay in the Lauraceae." *Frontiers in Plant Science* 14: 1122549. https://doi.org/10.3389/fpls.2023.1122549.

Vurture, G. W., F. J. Sedlazeck, M. Nattestad, et al. 2017. "GenomeScope: Fast Reference-Free Genome Profiling From Short Reads." *Bioinformatics* 33, no. 14: 2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

Wang, S., and Y. Xie. 2004. "Chapter 6: China species red list, lauraceae." In *China Species Red List*, Vol. 1, edited by S. Wang and Y. Xie, 330–334. BeijingHigher Education Press.

Wang, Y., H. Tang, J. D. DeBarry, et al. 2012. "MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity." *Nucleic Acids Research* 40, no. 7: e49. https://doi.org/10.1093/nar/gkr1293.

Wu, T. D., and C. K. Watanabe. 2005. "GMAP: A Genomic Mapping and Alignment Program for mRNA and EST Sequences." *Bioinformatics* 21, no. 9: 1859–1875. https://doi.org/10.1093/bioinformatics/bti310.

Xiao, R., C. Zhang, X. Guo, H. Li, and H. Lu. 2021. "MYB Transcription Factors and Its Regulation in Secondary Cell Wall Formation and Lignin Biosynthesis During Xylem Development." *International Journal of Molecular Sciences* 22, no. 7: 3560. https://doi.org/10.3390/ijms22073560.

Xiong, B., L. Zhang, L. Xie, et al. 2022. "Genome of *Lindera glauca* Provides Insights Into the Evolution of Biosynthesis Genes for Aromatic Compounds." *iScience* 25, no. 8: 104761. https://doi.org/10.1016/j.isci.2022.104761.

Xiong, X., J. Gou, Q. Liao, et al. 2021. "The Taxus Genome Provides Insights Into Paclitaxel Biosynthesis." *Nature Plants* 7, no. 8: 1026–1036.

Xue, J.-Y., Z. Li, S.-Y. Hu, et al. 2023. "The Saururus chinensis Genome Provides Insights Into the Evolution of Pollination Strategies and Herbaceousness in Magnoliids." *Plant Journal* 113, no. 5: 881–1101. https://doi.org/10.1111/tpj.16097.

Yang, Y., and B. Liu. 2015. "Species Catalogue of Lauraceae in China: Problems and Perspectives." *Biodiversity Science* 23, no. 2: 232–236. https://doi.org/10.17520/biods.2015027.

Yang, Y., P. Sun, L. Lv, et al. 2020. "Prickly Waterlily and Rigid Hornwort Genomes Shed Light on Early Angiosperm Evolution." *Nature Plants* 6, no. 3: 215–222. https://doi.org/10.1038/s41477-020-0594-6.

Yang, Z. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24, no. 8: 1586–1591. https://doi.org/10.1093/molbev/msm088.

Yang, Z., W. Y. Jin, B. Liu, D. K. Ferguson, and Y. Yang. 2021. "Big Fruits With Tiny Tepals: An Unusual New Species of Lauraceae From Southwestern China." *PhytoKeys* 179: 129–143. https://doi.org/10.3897/phytokeys.179.62050.

Zdobnov, E. M., and R. Apweiler. 2001. "InterProScan – An Integration Platform for the Signature-Recognition Methods in Interpro." *Bioinformatics* 17, no. 9: 847–848. https://doi.org/10.1093/bioinformatics/17.9.847.

Zhang, B., X. Yao, H. Chen, and L. Lu. 2022. "High-Quality Chromosome-Level Genome Assembly of *Litsea coreana* L. Provides Insights Into Magnoliids Evolution and Flavonoid Biosynthesis." *Genomics* 114, no. 4: 110394. https://doi.org/10.1016/j.ygeno.2022.110394.

Zhang, L., F. Chen, X. Zhang, et al. 2020. "The Water Lily Genome and the Early Evolution of Flowering Plants." *Nature* 577: 79–84. https://doi.org/10.1038/s41586-019-1852-5.

Zhang, Q., L. Wang, Z. Wang, et al. 2021. "The Regulation of Cell Wall Lignification and Lignin Biosynthesis During Pigmentation of Winter Jujube." *Horticulture Research* 8, no. 1: 238. https://doi.org/10.1038/s41438-021-00670-4.

Zhao, T., A. Zwaenepoel, J. Y. Xue, et al. 2021. "Whole-Genome Microsynteny-Based Phylogeny of Angiosperms." *Nature Communications* 12, no. 1: 3498.

Zhong, R., and Z. H. Ye. 2012. "MYB46 and MYB83 Bind to the SMRE Sites and Directly Activate a Suite of Transcription Factors and Secondary Wall Biosynthetic Genes." *Plant and Cell Physiology* 53, no. 2: 368–380. https://doi.org/10.1093/pcp/pcr185.

Zhou, G., H. Yin, F. Chen, et al. 2022. "The Genome of Areca Catechu Provides Insights Into Sex Determination of Monoecious Plants." *New Phytologist* 236, no. 6: 1999–2373. https://doi.org/10.1111/nph.18471.

Zhou, J., C. Lee, R. Zhong, and Z. H. Ye. 2009. "MYB58 and MYB63 Are Transcriptional Activators of the Lignin Biosynthetic Pathway During Secondary Cell Wall Formation in Arabidopsis." *Plant Cell* 21, no. 1: 248–266. https://doi.org/10.1105/tpc.108.063321.

Zou, H. Y., and D. R. Wu. 1997. "Taxonomic notes of *Phoebe bournei*." In *Population Ecology of* Phoebe bournei, edited by H. Y. Zou and D. R. Wu, 2–4. BeijingChina Forestry Publishing Press.

Zuntini, A. R., T. Carruthers, O. Maurin, et al. 2024. "Phylogenomics and the Rise of the Angiosperms." *Nature* 629, no. 8013: 843–850.