The chromosome-scale genome of *Phoebe bournei* reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies

Xiao Han1,4, Junhong Zhang1,4, Shuang Han1, Sun Li Chong1, Guanliang Meng2, Minyan Song1, Yang Wang1, Shengcai Zhou3, Chengcheng Liu1, Luhuan Lou1, Xiongzheng Lou1, Longjun Cheng1, Erpei Lin1, Huahong Huang1, Qi Yang1,* and Zaikang Tong1,*

1State Key Laboratory of Subtropical Silviculture, College of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou, Zhejiang 311300, China
2Nextomics Biosciences, Wuhan, Hubei 430000, China
3Experimental Forest Farm of Qingyuan County, Qingyuan, Zhejiang 323800, China
4These authors contributed equally to this article.

*Correspondence: Qi Yang (qiyang@zafu.edu.cn), Zaikang Tong (zktong@zafu.edu.cn)

https://doi.org/10.1016/j.xplc.2022.100410

ABSTRACT

Terpenoids, including aromatic volatile monoterpenoids and sesquiterpenoids, function in defense against pathogens and herbivores. *Phoebe* trees are remarkable for their scented wood and decay resistance. Unlike other Lauraceae species investigated to date, *Phoebe* species predominantly accumulate sesquiterpenoids instead of monoterpenoids. Limited genomic data restrict the elucidation of terpenoid variation and functions. Here, we present a chromosome-scale genome assembly of a Lauraceae tree, *Phoebe bournei*, and identify 72 full-length terpene synthase (TPS) genes. Genome-level comparison shows pervasive lineage-specific duplication and contraction of TPS subfamilies, which have contributed to the extreme terpenoid variation within Lauraceae species. Although the TPS-a and TPS-b subfamilies were both expanded via tandem duplication in *P. bournei*, more TPS-a copies were retained and constitutively expressed, whereas more TPS-b copies were lost. The TPS-a genes on chromosome 8 functionally diverged to synthesize eight highly accumulated sesquiterpenes in *P. bournei*. The essential oil of *P. bournei* and its main component, β-caryophyllene, exhibited antifungal activities against the three most widespread canker pathogens of trees. The TPS-a and TPS-b subfamilies have experienced contrasting fates over the evolution of *P. bournei*. The abundant sesquiterpenoids produced by TPS-a proteins contribute to the excellent pathogen resistance of *P. bournei* trees. Overall, this study sheds light on the evolution and adaptation of terpenoids in Lauraceae and provides valuable resources for boosting plant immunity against pathogens in various trees and crops.

Keywords: *Phoebe bournei*, genome sequence, terpene synthases, sesquiterpenoids, positive selection, pathogen resistance

Han X., Zhang J., Han S., Chong S.L., Meng G., Song M., Wang Y., Zhou S., Liu C., Lou L., Lou X., Cheng L., Lin E., Huang H., Yang Q., and Tong Z. (2022). The chromosome-scale genome of *Phoebe bournei* reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies. Plant Comm. 3, 100410.
bark are the first barrier in the defense against stem borer attacks and subsequent pathomycete infections (Nerg et al., 2004; Della Rocca et al., 2020). Terpenoids are also of significant interest because of their extensive use in the pharmaceutical, cosmetic, agriculture, and food industries (Boutanaev et al., 2015). As the most chemically, structurally, and functionally diverse compounds in plants, terpenoids are primarily restricted to specific lineages or even a single species and are therefore referred to as specialized terpenoids. Lauraceae, Piperaceae, Lamiaclaceae, Apliaceae, Myrtaceae, Rutaceae, and Zingiberaceae are representative angiosperm families with abundant terpenoids (Ebadollahi and Sendi, 2015). Lauraceae species produce volatile organic compounds (VOCs) with high economic value in the spice and perfume industries. For instance, species in Lauraceae, such as May Chang tree (Litsea cubeba), bay laurel (Laurus nobilis), and camphor tree (Cinnamomum camphora), have distinctive scents that consist mainly of diverse monoterpoids (Joshi et al., 2009). By contrast, \textit{Phoebe}, a genus closely related to \textit{Cinnamomum}, emits a scent dominated by sesquiterpenoids (Joshi et al., 2009; Xie et al., 2015; Ding et al., 2020).

Driven by selection pressure to adapt to their specific ecological niches, terpene synthase (TPS) proteins are the gatekeepers of terpenoid diversity. TPSs catalyze the cyclization and rearrangement of a few common prenyl pyrophosphate precursors (differing in the number of condensed five-carbon isoprenoid units) to generate a range of monoterpoid (C10), sesquiterpene (C15), and diterpene (C20) scaffolds (Chen et al., 2011). In plants, the TPS family is divided into seven classes: TPS-a, TPS-b, and TPS-g, which are three angiosperm-specific subfamilies; TPS-d (gymnosperm specific); TPS-h (Selaginella spp. specific); TPS-e/f, and TPS-c. The TPS-a, TPS-b, and TPS-g subfamilies predominantly comprise synthases that catalyze the formation of monoterpene and sesquiterpenes, whereas the TPS-c and TPS-e/f subfamilies mainly comprise diterpene synthases. Unlike the typically mid-sized diterpene synthases with 2–30 members, mono- and sesquiterpene synthases have undergone far more significant expansion (e.g., 69 members in grape (Martin et al., 2010) and 113 members in \textit{Eucalyptus} (Kulhem et al., 2015)) that has resulted in diverse volatile terpenoids. Elucidation of the biological characteristics of TPS genes from an evolutionary perspective is essential for the bioengineering of metabolic pathways for specific terpenoids (Mewalal et al., 2017). Genome-wide studies of the TPS family have been conducted in the model plant \textit{Arabidopsis}, the gymnosperms \textit{Picea} spp. (Keeling et al., 2011), and multiple herbaceous angiosperm species (Boutanaev et al., 2015; Zhou and Pichersky, 2020a; Booth et al., 2020; Borowski et al., 2020; Karunanithi et al., 2020). The possible molecular mechanisms and evolutionary history of terpenoids in perennial woody angiosperm plants, especially Lauraceae species, are largely unknown.

The broad-leaved evergreen tree genus \textit{Phoebe}, which belongs to the Lauraceae family, comprises approximately 100 species in tropical and subtropical Asian forests (Wei and van der Werff, 2008). \textit{Phoebe} timber has been widely used for high-quality furniture and buildings for centuries in Asia because of its pleasant scent, high durability, excellent decay resistance, and attractive appearance with a golden tint on the wood surface (Figure 1A). The preserved antique furniture and historic wooden buildings in the Forbidden City made from \textit{Phoebe} are emblematic of these excellent wood properties. \textit{Phoebe bournei}, a subtropical evergreen tree endemic to China (Figure 1A), is endangered because of the low survival rate of seedlings, intensive deforestation, and illegal logging for timber. Anti-hollow formation is highly valuable for structural wood production. In subtropical natural forests, hollow trunks are rarely observed in \textit{P. bournei} trees, especially in ancient trees with a diameter of more than 100 cm (Xie et al., 2015).

Here, we present the chromosome-scale genome assembly of \textit{P. bournei} and use it to address three main questions: the evolution and divergence of TPSs underlying terpenoid diversity between \textit{P. bournei} and other Lauraceae species, the mechanisms of sesquiterpenoid accumulation in \textit{P. bournei}, and the possible contribution of sesquiterpenoids to \textit{P. bournei} adaptation. The combination of evolutionary, transcriptomic, and metabolic analyses of the TPS family in \textit{P. bournei} will shed light on the evolution of metabolic diversity and enable further investigations of terpenoid-driven resilience to environmental stresses.

### RESULTS

**Assembly and characterization of the \textit{Phoebe} genome**

\textit{Phoebe bournei} (Hems.) Yang is a diploid (2n = 24) with an estimated genome size of 1,078.25 ± 39.81 Mb/1C, as determined by flow cytometric analysis (Supplemental Figure 1). A genome survey of \textit{P. bournei} based on a 17-mer analysis suggested an estimate of 941.77 Mb for the genome size and high heterozygosity of 1.4%. PacBio long reads (117.11 Gb, more than 100x) were used for assembly with the PacBio-preferred Falcon program and then polishing with Illumina short reads. Long-range mapping was performed to assemble the contigs onto scaffolds using 272.26 Gb of BioNano Saphyr optical mapping data, which resolved the contig conflicts and produced 463 scaffolds with a scaffold N50 of 272.26 Gb of Hi-C data. As a result, 1,057 Gb of 1.065 Gb total scaffold sequences were anchored onto 12 pseudochromosomes that represented 99.20% of the final chromosome-scale genome assembly with 85 scaffolds (scaffold N50 of 97.01 Mb) and a maximum pseudochromosome length of 137.96 Mb (Figure 1B, Supplemental Table 2, and Supplemental Figure 2). Estimation of genome completeness revealed 1266 (92.07%) complete benchmarking universal single-copy orthologs (BUSCOs) in the genome (Supplemental Table 3). In addition, 92.30–93.84% of RNA sequencing (RNA-seq) reads from the root, stem, leaf, bud, and seed were mapped to the genome (Supplemental Table 4). Repetitive elements accounted for approximately 48.10% of the \textit{P. bournei} genome (Supplemental Tables 5 and 6). This percentage was lower than that previously reported (68.51%) (Chen et al., 2020b) but similar to that in \textit{Cinnamomum kanehira}e (48%) (Chaw et al., 2019). Among these elements, long terminal repeats accounted for 26.34% of the \textit{P. bournei} genome. Copia (4.73%) and Gypsy (8.69%), which represented half of the long terminal repeats in \textit{P. bournei}, were found less often in the \textit{C. kanehirae} genome (6.10% and 10.40%). Gene annotation based on the integration of three strategies suggested a final set of 30,096 protein-coding genes (Supplemental Table 7), of which 29,716 genes
Evolution of terpene synthases in *Phoebe bournei*

(98.74%) were annotated with several conventional databases (Supplemental Table 8). The completeness assessment showed 1261 (91.7%) BUSCOs in all annotated proteins (Supplemental Table 9).

A total of 211 single-copy orthologous genes from each of 18 seed plant species were used to analyze their phylogenetic relationships. The phylogenetic topologies derived from maximum likelihood (ML) and Bayesian trees were identical, in that magnoliids were a sister clade to eudicots rather than monocots (Figure 1C and Supplemental Figure 3). In Lauraceae, *C. kanehirae* had a close relationship with *L. cubeba*, and *P. bournei* was closely related to *Persea americana*. Analysis of the whole-genome duplication (WGD) history of *P. bournei* showed that the synonymous substitutions per synonymous site ($K_s$) distribution of paralogous gene pairs in *P. bournei* had two peaks for WGD events at approximately $K_s = 0.57$ and $K_s = 0.88$ (Figure 1D), consistent with the results of four-fold degenerate synonymous site analysis (Supplemental Figure 4). Two peaks at similar $K_s$ values were also identified for...
Figure 2. Profile of volatile terpenoids in *P. bournei*.

(A) Relative contents of volatile compounds in the leaves and stems of 297 genotypes.

(B) Relative contents of total monoterpenoids and total sesquiterpenoids in various tissues (biological replicates: \( n = 3 \), mean ± SD). Composition data of monoterpenoids from the root xylem are shown (top inset).

(C) Scatterplot of volatile terpenoids from various tissues based on OPLS-DA. Pink triangles and colored dots represent tissues and terpenoids, respectively.

**legend continued on next page**
Evolution of terpene synthases in *Phoebe bournei*

*C. kanehira* and *P. americana* (Chaw et al., 2019; Rendon-Anaya et al., 2019). The *K* distribution for intergenomic orthologs indicated that the more ancient WGD event occurred approximately 144.21–146.70 million years ago, before the divergence of Laurales and Magnoliales. The more recent WGD event shared by Lauraceae plants occurred 92.90–95.39 million years ago.

**Variation in tissue-specific volatile terpenoid profiles of *P. bournei***

The effects of environment and population divergence on terpenoid biosynthesis could be significant. Therefore, to minimize sampling biases arising from interpopulation differences in *P. bournei*, we analyzed VOC profiles in a collection of 297 genotypes from 33 provenances covering the entire natural habitat that were grown in a common garden for six years (Supplemental Figure 5). The results showed that sesquiterpenoids accounted for at least 85% of the volatile compounds in leaf and stem samples of all collected *P. bournei* genotypes, whereas monoterpenoids accounted for less than 10% (Figure 2A).

To analyze the variety of terpenoids in *P. bournei*, we collected xylem and bark tissues from roots and stems, as well as leaves, from the genome-sequenced tree for further tissue-specific analysis of VOCs. There were 50 volatile compounds with a relative content above 0.5% in at least one of five analyzed *P. bournei* tissues, comprising 36 sesquiterpenoids, 13 monoterpenoids, and 1 aliphatic 3-hexenol (Supplemental Table 10). Consistent with data from the 297 genotypes, sesquiterpenoids dominated the volatile compounds in aboveground tissues, especially stem bark (96.17%; Figure 2B). Surprisingly, monoterpenoids accumulated significantly in the root xylem, predominantly as α-pinene (52.61%) and β-pinene (10.8%) (Figure 2B and Supplemental Table 10). In addition to these two monoterpenes, the remaining VOCs in the root xylem included 25.92% sesquiterpenoids and 5.63% monoterpenoids (Figure 2B). The results indicated that sesquiterpenoids were highly abundant in various tissues of *P. bournei* and represented by numerous varieties. A clear distinction of VOC profiles among tissues was revealed by orthogonal partial least squares discriminant analysis (OPLS-DA; Figure 2C and Supplemental Figure 6). Sesquiterpenes, including β-caryophyllene (21.06% and 8.19%), α-cubebene (19.76% and 40.51%), and α-copaene (10.46% and 11.92%), were the major VOC constituents in stem xylem and bark (Figure 2D and Supplemental Table 10). Hence they may contribute to the pleasant scent of *Phoebe* wood. VOC composition of the root bark contrasted with that in the root xylem. Overall, although the composition of sesquiterpenes varied among tissues, α-cubebene, α-copaene, and β-caryophyllene were among the most abundant sesquiterpenes in all studied tissues of *P. bournei*, especially the stem (Figure 2D and Supplemental Table 10).

**Antifungal activity of sesquiterpenoids in *P. bournei***

When collecting genotypes from various natural distribution areas, we observed that no *Phoebe* trunks were hollowed, even among the old trees. A previous study showed that the monoterpene d-terpineol contributed to decay resistance in *Cinnamonum* trees (Chaw et al., 2019). *P. bournei* accumulated abundant sesquiterpenoids, especially β-caryophyllene. To investigate the possible antifungal activity of the sesquiterpenes, we performed *in vitro* fungal growth inhibition assays using the essential oil of *P. bournei* and a commercial standard of β-caryophyllene (Figure 2E). *Botryosphaeria dothidea*, *Colletotrichum gloeosporioides*, and *Lasiodiplodia pseudotheobromae* were chosen for bioactivity assays because these species are among the most widespread and important canker and dieback pathogens of trees worldwide (Kamle and Kumar, 2016; Li et al., 2016; Marsberg et al., 2017). Overall, a dose of 0.25% essential oil showed moderate inhibitory activity against the mycelial growth of all the fungi except *L. pseudotheobromae* (Figure 2E). Interestingly, addition of 0.25% β-caryophyllene inhibited the mycelial growth of all three phytopathogenic fungi by as much as 75% (Figure 2E).

**Characteristics of the TPS gene family in *P. bournei***

To deepen our understanding of terpenoid biosynthesis in *P. bournei*, we identified genes encoding the proteins responsible for synthesizing a variety of terpenoids. In total, 72 full-length genes and 44 partial fragments encoding TPSs were identified in the *P. bournei* genome (Supplemental Figure 7). The coding sequences of TPS genes were further checked with transcriptomic data. Among the 72 full-length TPS genes, 10 were putative pseudogenes based on the presence of a frameshift or a premature stop codon that resulted in a truncated protein. These 10 full-length genes were also incorporated into the phylogenetic and gene expression analyses after removing the stop codons or revising the frameshifts by deleting one or two nucleotides. Using the Simple Modular Architecture Research Tool server with Protein Families Database (PFAM) domain analysis (Letunic et al., 2021), all predicted proteins encoded by the 72 genes were found to contain typical TPS domains, supporting their identification as TPS genes. However, the 44 partial TPS fragments did not contain the full-length TPS domain, indicating that the pseudogenization events of these fragments occurred a long time ago.

To define the classes of the 72 TPSs of *P. bournei*, we constructed a phylogenetic tree using the ML procedure with 649 full-length TPSs identified from 18 genome-sequenced species representing gymnosperms, basal angiosperms, magnoliids, eudicots, and monocots (Supplemental Table 11). TPS genes of *P. bournei* were grouped into six TPS clades, designated the TPS-a, b, g, e, f, and c subfamilies, with high bootstrap support values. Among these 72 TPSs, the TPS-a and TPS-b classes were the largest, with 34 and 28 members, respectively. The

(D) Relative contents of the 50 most abundant volatile compounds in various tissues. The dot size corresponds to the percentage of each compound compared with the most abundant product in the given tissues. Hexenol, monoterpenoids, and sesquiterpenoids are represented by pink, green, and blue dots, respectively. The sampled tissues were root xylem (RX), root bark (RB), stem xylem (SX), stem bark (SB), and leaves (L).

(E) Percentages of inhibition of mycelial growth by essential oil from *P. bournei* leaves and the β-caryophyllene standard. The three phytopathogenic fungi are *Botryosphaeria dothidea*, *Colletotrichum gloeosporioides*, and *Lasiodiplodia pseudotheobromae*. 
**Evolution of terpene synthases in Phoebe bournei**

Gene families expanded by tandem duplication, segmental duplication, and WGD provide raw genetic material for evolution. In *P. bournei*, only the size of the TPS-a subfamily was enlarged (i.e., 1.4 times) compared with that in the Lauraceae MRCA (Figure 3). To understand how the TPS gene family expanded in *P. bournei*, we first examined the chromosomal distribution of the 72 TPS genes and 44 TPS fragments. Except for 3 TPS genes and 2 TPS fragments assigned to three unattributed scaffold fragments, all TPS genes and fragments were assigned to 10 of the 12 chromosomes (Supplemental Figure 10 and Supplemental Table 12). The TPS distribution on chromosomes was heterogeneous. Chromosomes 2, 7, and 8 harbored the majority of the TPS genes. A total of 86% of the TPS genes (62/72) and 80% of the TPS fragments (35/44) were located in tandemly arrayed clusters. We also analyzed the sequence identities of TPS proteins in *P. bournei*. Consistent with the relationships in the phylogenetic tree, members within the same tandem cluster showed higher sequence identities than other TPS genes on the same chromosome (Supplemental Figure 11). These results indicated that tandem duplications contributed significantly to expansion of the TPS gene family in *P. bournei*.

**Extensive tandem duplication and high retention of TPSs in P. bournei**

We further reconstructed the expansion history of each gene cluster in accordance with both the gene tree and chromosome localization (Supplemental Figures 7 and 10). The most parsimonious scenarios for gene duplication, rearrangement, and loss are presented (Figure 4). In general, all 13 clusters were expanded by tandem duplications. The historical expansion events of the two large TPS-a clusters (C9 and C11) appeared to have been complex (Figure 4C). It is likely that an extensive tandem duplication and high retention of TPSs in *P. bournei*.

TPS-g and TPS-e classes had three and five members, respectively. The TPS-f and TPS-c classes had only one member each (Supplemental Table 11).

### Pervasive lineage-specific gene expansion and loss contributed to the divergence of the TPS family in Lauraceae

Phylogenetic analysis of 649 full-length TPSs from 18 species showed that the largest subfamilies were TPS-a and TPS-b (Supplemental Figure 8). Each Lauraceae species except *P. americana* contained more than 40 copies of the TPS-a and TPS-b genes (Supplemental Table 11). The TPS-a and TPS-b subfamilies predominantly comprised sesquiterpene and mono-terpene synthases, respectively. To further explore the underlying mechanism of terpenoid variation in Lauraceae species, we conducted a phylogenetic analysis of all full-length TPSs from *C. kanehirae*, *L. cubeba*, *P. americana*, and *P. bournei*. There were at least 65 ancestral TPS genes, including 24 TPS-a genes and 29 TPS-b genes, in the most recent common ancestor (MRCA) of the four Lauraceae species (Figure 3 and Supplemental Figure 9). The large number of TPSs in the MRCA may be partially attributed to two WGD events in Lauraceae species (Figure 1D and Supplemental Figure 4). The total gene numbers in the TPS-g, e, f, and c subfamilies were maintained or slightly reduced in Lauraceae species compared with their MRCA. By contrast, most of the TPS-a and TPS-b orthologs of the MRCA were lost in at least one of the four species after their divergence. This was particularly pronounced in *P. americana*, where most TPS genes were lost, especially for the TPS-b subfamily, leaving only 20 TPS genes (Figure 3).

Contemporaneously with the loss, Lauraceae species also gained numerous TPS paralogs (Figure 3). For example, *P. bournei* lost nine TPS-a orthologs and specifically duplicated 19 TPS-a paralogs, resulting in 34 TPS-a copies. Similarly, *C. kanehirae* lost 16 but gained 28 TPS-b genes and ultimately retained 41 TPS-b copies. As a result of more gain than loss, the *P. bournei* genome harbored 72 TPSs with more TPS-a copies, whereas the *C. kanehirae* genome harbored 77 TPS genes with more TPS-b copies.

### Extensive tandem duplication and high retention of TPSs in P. bournei

**Figure 3. Copy number changes of TPSs in four Lauraceae species.**

The numbers in circles and rectangles represent the numbers of TPS genes in extant and ancestral species, respectively. Numbers on branches with plus and minus symbols represent the numbers of gene gains and losses, respectively.

**TPS-a**
- P. bournei: 34
- P. americana: 12
- C. kanehirae: 24
- L. cubeba: 17

**TPS-b**
- P. bournei: 28
- P. americana: 3
- C. kanehirae: 41
- L. cubeba: 24

**TPS-g**
- P. bournei: 3
- P. americana: 2
- C. kanehirae: 4
- L. cubeba: 1

**TPS-e**
- P. bournei: 5
- P. americana: 3
- C. kanehirae: 5
- L. cubeba: 1

**TPS-f**
- P. bournei: 3
- P. americana: 2
- C. kanehirae: 2
- L. cubeba: 1

**TPS-c**
- P. bournei: 1
- P. americana: 0
- C. kanehirae: 0
- L. cubeba: 0

The numbers in circles and rectangles represent the numbers of TPS genes in extant and ancestral species, respectively. Numbers on branches with plus and minus symbols represent the numbers of gene gains and losses, respectively.

TPS-g and TPS-e classes had three and five members, respectively. The TPS-f and TPS-c classes had only one member each (Supplemental Table 11).

**Extensive tandem duplication and high retention of TPSs in P. bournei**

Gene families expanded by tandem duplication, segmental duplication, and WGD provide raw genetic material for evolution. In *P. bournei*, only the size of the TPS-a subfamily was enlarged (i.e., 1.4 times) compared with that in the Lauraceae MRCA (Figure 3). To understand how the TPS gene family expanded in *P. bournei*, we first examined the chromosomal distribution of the 72 TPS genes and 44 TPS fragments. Except for 3 TPS genes and 2 TPS fragments assigned to three unattributed scaffold fragments, all TPS genes and fragments were assigned to 10 of the 12 chromosomes (Supplemental Figure 10 and Supplemental Table 12). The TPS distribution on chromosomes was heterogeneous. Chromosomes 2, 7, and 8 harbored the majority of the TPS genes. A total of 86% of the TPS genes (62/72) and 80% of the TPS fragments (35/44) were located in tandemly arrayed clusters. We also analyzed the sequence identities of TPS proteins in *P. bournei*. Consistent with the relationships in the phylogenetic tree, members within the same tandem cluster showed higher sequence identities than other TPS genes on the same chromosome (Supplemental Figure 11). These results indicated that tandem duplications contributed significantly to expansion of the TPS gene family in *P. bournei*.

We further reconstructed the expansion history of each gene cluster in accordance with both the gene tree and chromosome localization (Supplemental Figures 7 and 10). The most parsimonious scenarios for gene duplication, rearrangement, and loss are presented (Figure 4). In general, all 13 clusters were expanded by tandem duplications. The historical expansion events of the two large TPS-a clusters (C9 and C11) appeared to have been complex (Figure 4C). It is likely that an
Figure 4. Phylogenetic trees and hypothetical evolutionary histories of the *P. bournei* TPS clusters.

(A–J) Numbers on the branches are bootstrap values. The letters T, S, and R in the schematic diagram indicate hypothetical events of tandem duplication, segmental duplication, and rearrangement, respectively. Boxes represent TPS genes that are functional (black), lost (white), or pseudogenes/fragments (gray). Full-length pseudogenes are indicated by asterisks. Frag indicates a TPS fragment. The TPS clusters numbered C1, C2, etc., are shown in Supplemental Figure 10.
An ancestral cluster including seven tandemly arrayed TPSs on chromosome 8 underwent a unique segmental duplication event (the segment was longer than 2 Mb), probably as a consequence of local gene duplication by unequal crossover. Subsequently, each descendant cluster experienced a series of tandem duplications and rearrangements (Figure 4C). Both the TPS-a and TPS-b subfamilies experienced an extensive expansion in *P. bournei* via tandem duplication. Notably, more copies of TPS-a are preserved, and more copies of TPS-b are pseudogenized in the genome, which is supported by the large numbers of pseudogenes and fragments belonging to the TPS-b subfamily (including 27 fragments and 7 pseudogenes).

### Positive selection on genes of the TPS-a and TPS-b subfamilies in *P. bournei*

To infer the influences of selection on the expansion of the TPS-a and TPS-b subfamilies in *P. bournei*, we conducted molecular evolutionary analysis. The \( \omega \) (\( d_\text{N}/d_\text{S} \)) values for TPS-a and TPS-b genes were estimated using Phylogenetic Analysis by Maximum Likelihood (PAML) software with the one-ratio and free-ratio models (Supplemental Figure 12). Likelihood ratio tests of the two models indicated that the one-ratio model should be rejected in favor of the free-ratio model \( (P < 0.001) \) for both the TPS-a and TPS-b subfamilies, suggesting that the branches were under various selection pressures (Supplemental Table 13). The \( \omega \) values estimated by the free-ratio model were larger than one, indicating that the genes were under positive selection (Supplemental Figure 12). For all TPS-a genes in *P. bournei*, positive selection was detected on three branches (Clades I, II, and III, Supplemental Figure 12A). The branches under positive selection within Clade I were composed of \( \text{PbTPS-a17, PbTPS-a18, PbTPS-a19, PbTPS-a22, and PbTPS-a23} \) from the C9 and C11 clusters. These genes were specifically expanded in *P. bournei* compared with the other Lauraceae species (Supplemental Figure 9). Positive selection was also detected for several branches of TPS-b genes in *P. bournei* (Supplemental Figure 12B).

### Divergence of TPS expression in *P. bournei*

Gene expression divergence often occurs at or shortly after duplication, especially for tandem and dispersed duplicates (Ganko et al., 2007). To explore the functional divergence of TPSs in *P. bournei*, we examined the tissue-specific expression profiles of TPS genes in root xylem, root bark, stem xylem, stem bark, and leaves via transcriptomic analysis (Figure 5B). In general, the expression patterns of TPS genes in the xylem and bark tissues of the stem differed from those of the root (Figure 5B). Moreover, the TPS genes exhibited a unique expression profile in leaves compared with the other tissues. Twenty-seven of the 34 TPS-a genes showed a high expression level in root and/or stem tissue, and half of the TPS-b genes were underexpressed in all tissues.

---

Figure 5. The phylogenetic relationships and expression patterns of *P. bournei* TPSs.

(A) Phylogenetic tree of full-length TPS proteins of *P. bournei*. The ML procedure using the JTT+F+I model with 100 bootstrap replicates was used for phylogenetic reconstruction. The numbers at each node in the phylogenetic tree are bootstrap values, and only bootstrap values >50% are shown. Putative pseudogenes are indicated with asterisks. The TPS-a, TPS-b, TPS-g, TPS-e, TPS-f, and TPS-c subfamilies are shaded in different colors.

(B) Tissue-specific expression patterns of TPSs in root xylem (RX), root bark (RB), stem xylem (SX), stem bark (SB), and leaves (L). Expression data are provided as log2(fragments per kilobase per million mapped fragments [FPKM]+1) values.
we analyzed. The tissue-specific expression patterns of the TPSs were in line with the mono- and sesquiterpenoid compositions of the corresponding tissues (Figures 2D and 5B). Our results suggested that most TPS-a genes (which mainly encode sesquiterpene synthases) were constitutively expressed in stems and roots with high expression levels, whereas most TPS-b genes (which mainly encode monoterpene synthases) were expressed only under specific conditions or in specific tissues (Figure 5B).

In addition, we analyzed the TPS expression profiles in P. bournei seedlings under drought and cold stresses. All 72 TPS genes exhibited divergent expression patterns either in specific tissues or in response to specific stresses (Figure 5B) and Supplemental Figure 13). Furthermore, consistent with the gene expression divergence results, the cis-elements upstream of the 72 TPS genes varied greatly (Supplemental Figure 14).

Functional characterization of TPS-a genes
To better understand the functional divergence of TPS genes and verify the contribution of TPS-a genes to sesquiterpene accumulation in P. bournei, we characterized the enzymatic function of all 10 TPS-a genes from the C9 and C11 clusters, except the pseudogene PbTPS-a24, in tobacco (Figure 6). There were three paralogous gene pairs among these 10 genes: PbTPS-a20/a25, PbTPS-a21/a26, and PbTPS-a22/a23 (Figure 6A).

The sesquiterpenoids produced heterologously in tobacco leaves were similar to the major products observed in P. bournei, such as α-cubebene, α-copaene, β-copaene, β-caryophyllene, and elemene (Figures 2D and 6 and Supplemental Figure 15). For the gene pair PbTPS-a20/a25, the overexpression of PbTPS-a25, but not PbTPS-a20, in tobacco significantly enhanced the emission of β-caryophyllene, indicating the divergent functions of the paralogous genes (Figure 6). Likewise, for the gene pair PbTPS-a22/a23, the accumulation of α-cubebene, α-copaene, and β-copaene was detected in transgenic tobacco with PbTPS-a22, but not PbTPS-a23 (Figure 6). For the gene pair PbTPS-a21/a26, PbTPS-a21 catalyzed the formation of the bicyclic terpenoids β-caryophyllene and guai-6,9-diene as major products and the monocyclic terpenoids elemene and ylangene as minor products. PbTPS-a26 exhibited a distinct enzymatic function by enhancing the emissions of the tricycles α-copaene and β-copaene, indicating that neofunctionalization or subfunctionalization had occurred in the PbTPS-a21/a26 gene pair (Figure 6). In summary, we identified PbTPS-a22 as the main synthase for α-cubebene

Figure 6. Functional characterization of P. bournei TPS-a genes.
(A) Chromatograms of GC–MS traces of products from transient overexpression of TPS-a genes in Nicotiana benthamiana leaves. The phylogenetic tree in the left panel presents the relationships of all TPS-a genes on chromosome 8 of P. bournei. Except for that of the pseudogene PbTPS-a24, noted with an asterisk, the chromatograms of the α-copaene and β-caryophyllene standards, the empty vector (control), and the TPSs are shown in the right panel. The peaks of sesquiterpenoid products are shown in blue. Mass spectra for all detected TPS products are given in Supplemental Figure 15. The products were: 1, β-elemene; 2, α-cubebene; 3, ylangene; 4, α-copaene; 5, β-copaene; 6, β-caryophyllene; 7, γ-elemene; and 8, Guai-6,9-diene.
(B) Corresponding carbon skeletons are shown. Monocyclic, bicyclic, and tricyclic sesquiterpenoids are shaded in pink, yellow, and blue, respectively.
and PbTPS-a21 and PbTPS-a25 as the main synthases for β-caryophyllene, the two most abundant sesquiterpenes in the stem xylem of *P. bournei* (Figures 2D and 6).

**DISCUSSION**

In this study, we used PacBio, Bionano, and Hi-C data to assemble a chromosome-scale genome of *P. bournei*. For species with high heterozygosity, a chromosome-scale genome assembly helps to reduce sequence redundancy and improve genome accuracy and completeness. Compared with a contig-scale genome (Chen et al., 2020b), the decreased sequence redundancy was mainly achieved by removing repetitive sequences. The percentages of repetitive elements in the other three Lauraceae plants with chromosome-assembled genomes ranged from 48% to 55.47% (48% for *C. kanehira*, 50% for *P. americana*, and 55.47% for *L. cubeba*). The content of repetitive elements in the *P. bournei* genome was 48.10%, which also falls within this range. We used additional transcriptomic data from five tissues of *P. bournei* to correct the annotation of gene models. With the chromosome-assembled genome and the additional transcriptomic data, the quality of the *P. bournei* genome in our study was greatly improved compared with that in a previous report (Chen et al., 2020b). The improvements were evident in the increased average length of genes (10 606.6 versus 9 695.55 bp) and coding sequences (CDS) (1264.20 versus 1258.80 bp) and the more complete gene functional annotation (98.74% versus 95.79%) and BUSCO assessment (91.7% versus 81.1%) (Supplemental Tables 7–9).

The phylogenetic placement of magnoloids has a conflicting and undefinable phylogenetic topology. The sparse taxon sampling and limited gene numbers used for phylogenetic analysis are crucial restricting factors that result in the phylogenetic discordance of magnoloids among major Mesangiospermae lineages (Chaw et al., 2019; Chen et al., 2019, 2020c; Hu et al., 2019; Rendon-Anaya et al., 2019; Shang et al., 2020). In addition, incomplete lineage sorting during early diversification significantly impacts the elucidation of topological variation in Mesangiospermae lineages (Ma et al., 2021). The results of this study showed that magnoloids were a sister clade to eudicots rather than monocots (Figure 1C). A recent study with sufficient supporting evidence also indicated that the monocot lineage was sister to major Mesangiospermae lineages, including magnoloids and eudicots (Ma et al., 2021).

The genera *Phoebe* and *Persea* belong to the Tribe Persea and specifically accumulate sesquiterpenoids (Wei and van der Werff, 2008; Meléndez-González and Espinosa-García, 2018). By contrast, monoterpenoids are more abundant in species from other Lauraceae genera, such as *Cinnamomum*, *Laurus*, *Litsea*, and *Cryptocarya* (Joshi et al., 2009). Over the course of this project, newly assembled genome sequences became available for several Lauraceae species, including *C. kanehira*, *L. cubeba*, and *P. americana* (Chaw et al., 2019; Rendon-Anaya et al., 2019; Chen et al., 2020c), which made it possible to investigate the evolution of the TPS family in Lauraceae with reliable data. The high abundance of sesquiterpenoids detected in the vegetative organs of *P. bournei* was attributed to the large size and high expression level of the TPS-a gene family in *P. bournei* (Figures 3 and 5). We found that pervasive lineage-specific expansions and contractions of the TPS family contributed to the extreme terpenoid variation within Lauraceae (Figure 4 and Supplemental Figure 9). Similar to the findings of studies on the TPS family in other angiosperm species such as tomato, rice, sorghum, *C. kanehira*, and *L. cubeba* (Chen et al., 2011, 2020a, 2020c; Chaw et al., 2019; Zhou and Pichersky, 2020a), tandem duplication was the main pattern for TPS family expansion in *P. bournei* and provided raw genetic material for functional divergence and rapid biological evolution (Chantret et al., 2005; Lan et al., 2009). Interestingly, our results demonstrate that these two subfamilies of the TPS gene family (TPS-a and TPS-b) have not evolved in the same way in *P. bournei* (Figure 7). As shown by the MRCA analysis, 9 TPS-a orthologs were lost and 19 TPS-a paralogs were specifically duplicated, resulting in 34 TPS-a copies in *P. bournei* (Figure 3). By contrast, 10 TPS-b orthologs were lost and 9 TPS-b paralogs were specifically duplicated, resulting in 28 TPS-b copies in *P. bournei*. If we subtracted the pseudogenes among these 34 TPS-a and 28 TPS-b genes, then 31 TPS-a and 22 TPS-b genes encoding putative functional TPSs remained. Our results support 1.4 times (31/22) more putative functional TPS-a genes than TPS-b genes in *P. bournei*, indicating a more pronounced expansion of the TPS-a subfamily than the TPS-b subfamily after speciation.

Although the TPS-a and TPS-b families underwent similar degrees of tandem duplication based on the analysis of hypothetical evolutionary histories in *P. bournei* (Figure 4), the TPS-b subfamily had a higher rate of gene loss than the TPS-a subfamily. As evidenced by 27 fragments and 7 pseudogenes in the *P. bournei* genome, the TPS-b subfamily was contracted compared with the TPS-a subfamily. Diversifying selection is the main driver of lineage-specific expansions and contractions of gene families (McDowell et al., 1998; Alunni et al., 2007). The loss of TPS-b genes may be less harmful than that of TPS-a genes. Consistent with the gene expression divergence results, the cis-elements upstream of the 72 TPS genes varied to a great extent (Figure 5 and Supplemental Figure 14). Although the enzymatic function analysis showed that PbTPS-a25, PbTPS-a21, and PbTPS-a26 catalyzed the formation of β-caryophyllene, one of the most abundant sesquiterpenoids in *P. bournei* (Figure 6), PbTPS-a25, PbTPS-a21, and PbTPS-a26 showed divergent gene expression patterns. Although we cannot define the function of β-caryophyllene synthase as derived either from its ancestor or from a neo-functionalization event, these three genes functionally diverged in enzymatic function and gene expression. The contrasting evolutionary mechanisms and expression patterns between TPS subfamilies have together contributed to the accumulation of sesquiterpenoids instead of monoterpenoids in *P. bournei* (Figure 7).

The main constituents in leaf essential oils of *L. cubeba* and *C. kanehira* are monoterpenoids (Cheng et al., 2015; Chen et al., 2020c). By comparison, volatile analysis of avocado (*P. americana*) trees showed similar results to our study (Supplemental Table 10), in that sesquiterpenoids such as β-caryophyllene, α-cubebene, and α-copaene were dominant (Ogunbini et al., 2007; Niogret et al., 2013). Recent studies of *Chimonanthus praecox* showed that monoterpenes were the major components of its volatile compounds (Tian et al., 2019), and expansion of the TPS-b subfamily contributed to monoterpane accumulation in *C. praecox* (Shang et al., 2020).
Therefore, the relationship between the accumulation of sesquiterpenes and the high preservation of TPS-a genes in *P. bournei* is probably not coincidental. Admittedly, the number of compared genomes is limited because of the inadequacy of available genome data. Further evidence is needed from more plant species via combined analysis of TPS family evolution and metabolite profiles.

Compared with the other TPS subfamilies, the TPS-a and TPS-b genes exhibited rapid expansion in angiosperms (Chen et al., 2011), especially in Lauraceae species (Supplemental Figure 8). Why did this rapid expansion of TPS-a and TPS-b genes occur in angiosperms? One possible reason is driven by functional requirements (Nam et al., 2004). Perennial trees are subjected to many more abiotic and biotic stresses during their long life cycle than are annual plants. The diseases caused by herbivores and pathogens are devastating to forests, especially in modern forestation areas with a single or closely related species (Ploetz et al., 2017; Aylward et al., 2019). Many tree species in Lauraceae accumulate essential oils with specific scents, each of which exhibits a unique pattern of monoterpene and sesquiterpene composition (Guo et al., 2016; Chaw et al., 2019; Chen et al., 2020c). These volatile monoterpenes and sesquiterpenes play an essential role in pollinator attraction and defense against herbivory and pathogens (Pichersky and Raguso, 2018). Lauraceae trees dominate many subtropical forest sites in terms of abundance and coexisting species diversity (Ye et al., 2019). The chemical dissimilarity between cohabitant species could help reduce the probability of specialist biological enemies being transmitted between host plants (Pichersky and Raguso, 2018).

Defense against biological enemies is the best-established function of plant terpenoids (Pichersky and Raguso, 2018). However, the assignment of specific roles to specific terpenes has been difficult. *Phoebe* trees show good decay resistance and rarely have hollowed or cankered trunks in natural forests (Li and Min, 2020). The genus *Persea* is closely related to the genus *Phoebe* but is mainly distributed in America (Wei and van der Werff, 2008). In contrast to those in *P. bournei*, both the TPS-a and TPS-b subfamilies in *P. americana* were mostly contracted, with only 12 and 3 genes preserved in the genome, respectively (Figure 3). The loss of functional TPS genes in *Persea* may increase the risk of pathogen infection. In fact, epidemic laurel

---

**Figure 7. Schematic overview of the contrasting fates of the TPS-a and TPS-b subfamilies in *P. bournei* contributing to terpenoid variation.**

Although the TPS-a and TPS-b subfamilies are both expanded by tandem duplication, more TPS-a copies are retained and constitutively expressed, whereas more TPS-b copies are lost, resulting in a high abundance of sesquiterpenoids instead of monoterpenoids. VOCs with abundant sesquiterpenoids, especially β-caryophyllene, produced in stems may confer excellent pathogen resistance to *P. bournei* trees.
Plant Communications

wilt disease has killed over 300 million red bay trees (*Persea borbonia*) in the United States and currently poses an economic threat to avocado (*P. americana*) (Ploetz et al., 2017). Laurel wilt is caused by the fungus *Raffaelea lauricola*, a symbiont of an invasive Asian ambrosia beetle (*Xyleborus glabratu* Eichhoff) that prefers native laureaceous hosts such as *Phoebe* and camphor trees in China (Fraedrich et al., 2015). Despite the presence of the sesquiterpenes α-copaene and α-cubebene, which accumulate to high levels in *P. bournei* and are kairomones for attracting the beetle *X. glabratu*, there are no reports of laurel wilt in *Phoebe* trees in China (Ploetz et al., 2017).

Interestingly, β-caryophyllene and its oxide have been confirmed to inhibit the growth of a wide range of microbes and function as allelochemical components to induce resistance via jasmonic acid signaling (Chang et al., 2008; Silva et al., 2019; Karunanithi et al., 2020). β-caryophyllene, which is synthesized by *PbTPS-a21* and *PbTPS-a25*, exhibited high antipathogenic activity against three common woody plant pathogenic fungi (Figure 2E), suggesting that the abundance of sesquiterpenoids, especially β-caryophyllene, in the stem plays a vital role in the adaptation and pathogen defense of *P. bournei* (Figure 7). Notably, *PbTPS-a26* was the synthase for α-copaene, whereas its paralogous protein *PbTPS-a21* was sub- or neofunctionalized, synthesizing β-caryophyllene instead (Figure 6). The divergence of multifunctional duplicates of *PbTPS-a21/a26* might provide proof of the “coevolutionary arms race” hypothesis that assumes terpenoids diverge as a result of changes in their biotic partners and enemies (Pichersky and Raguso, 2018).

The genome-level comparison showed pervasive lineage-specific tandem duplication and contractions of the TPS family that contributed to the extreme terpenoid variation within Lauraceae. We demonstrated that the TPS-a and TPS-b subfamilies experienced contrasting fates during the evolution of *P. bournei*. The abundant sesquiterpenoids produced by TPS-a proteins in the stem contributed to the excellent pathogen resistance of *P. bournei* trees. Although the ecological functions of each specific terpenoid in *Phoebe* have not yet been elucidated, there is excellent potential for selecting effective antipest/pathogen compounds from *Phoebe*. In conclusion, this study sheds light on the evolution of terpenoid divergence in Lauraceae and the contribution of sesquiterpenoids to adaptation in *P. bournei*.

Evolution of terpene synthases in *Phoebe bournei*

MgCl2 isolation buffer (200 mM Tris, 4 mM MgCl2, 6H2O, 0.5% Triton X-100 [pH 7.5]) and 500 µl 1% polyvinyl pyrrolidone (PVP). The cell nuclei suspension was filtered through a 50-µm nylon filter into a 1.5-ml centrifuge tube and incubated on ice for 5 min. After 5 min of centrifugation at 1000 g, the supernatant was discarded. Then, 100 µl of precooled isolation buffer was added for resuspension. The nuclei were stained by adding 150 µl of precooled 50 µg ml−1 propidium iodide in the dark for 5 min at 4°C. The DNA content of at least 10 000 stained nuclei was analyzed on a CytoFLEX Flow Cytometer (Beckman Coulter, USA). The genome size of *P. bournei* was estimated using *Populus trichocarpa* as the internal standard.

Genome sequencing and assembly

The DNA library was constructed with a DNA Template Prep Kit, DNA Polymerase Binding Kit, and MagBead Kit (PacBio, Menlo Park, CA, USA). The 20-kb DNA library for long reads was sequenced on a PacBio RS II instrument using the optimal combination of P6 polymerase and C4 reagent. The raw data were filtered for quality control. DNA libraries for the genome survey, *Phoebe*, and *Hi-C* analyses were sequenced on the Illumina HiSeq X Ten platform with 150-bp paired-end reads. PacBio long-read sequences were assembled using FALCON v.1.8.7 (Chin et al., 2016). The consensus sequences of the primary assembly were then corrected and polished using GenomicConsensus v.2.3.0 ([https://opensourcelibs.com/lib/genomicconsensus](https://opensourcelibs.com/lib/genomicconsensus)) with PacBio reads and Pilon v.1.22 (Walker et al., 2014) with Illumina reads. The PacBio assembly was scaffolded using BioNano data and gap closed with PBJelly v.15.8.24 (English et al., 2012). A chromosome-level assembly was achieved with Hi-C data. Genome completeness was assessed using BUSCO v.3.0.2 and the Embryophyta 10 dataset and by alignment of RNA-seq datasets from the root, stem, leaf, bud, and seed tissues of *P. bournei* with STAR v.2.7.3 (Dobin et al., 2013). More details on genome assembly are described in the Supplemental notes.

Genome annotation and functional prediction

Repetitive sequences in the *P. bournei* genome were explored by homolog searching and *ab initio* prediction. For the first method, two databases, Repbase (Bao et al., 2015) and PGSB-REdat (Spannagl et al., 2016), were used for annotation with RepeatMasker v.4.0.8 (Bedell et al., 2000), RepeatProteinMask (Bedell et al., 2000), and Tandem repeats finder (Benson, 1999). For the second method, a de novo repeat database was built with RepeatModeler v.1.0.9 (Bedell et al., 2003) and then applied to the *P. bournei* genome with RepeatMasker. The repeat-masked genome was then subjected to protein-coding gene prediction. Three strategies were used to identify gene models. First, the *ab initio* and evidence-directed predictor Augustus (Stanke and Waack, 2003) was trained on *Arabidopsis thaliana* and used to predict gene models along the *P. bournei* genome. Second, protein-homology-based prediction was performed with GeMoMa (Keilwagen et al., 2016) and GeneWise (Birney and Durbin, 2000) using homolog protein sets of six representative plants, *A. thaliana*, *Oryza sativa*, *Populus trichocarpa*, *Amborella trichopoda*, *Eucalyptus grandis*, and *Vitis vinifera*, which were downloaded from Phytozome. Third, transcriptome data were used for annotation with Program to Assemble Spliced Alignments (PASA) (Haas et al., 2008). Finally, the three sets of gene models were merged and filtered using the EVidenceModeler (EVM) pipeline to generate the final gene models for the *P. bournei* genome. The gene annotation was processed by eggNOG-mapper ([Huerta-Cepas et al., 2017] with the Non-Redundant Protein Sequence Database (NR), Swissprot, Trembl, InterProScan, Gene Ontology, Eukaryotic Orthologous Groups of protein (KOG), and Kyoto Encyclopedia of Genes and Genomes databases. BUSCO v3.0.2 analysis of identified genes was performed using the Embryophyta 10 dataset.

Comparative genomic analyses

To elucidate the phylogenetic positions of *Phoebe* and magnoliids on the tree of life, we selected *P. bournei*, *P. americana*, *C. kanehirae*, *L. cubeba*, and...
Evolution of terpene synthases in *Phoebe bournei*

*Liriodendron chinense*, and *Piper nigrum* as representatives of magnoliids and compared these with five eudicots and five monocots, including the relatively early evolved monocot *Zostera marina* and the basal eudicots *Aquilegia coerulea*, *Macleaya cordata*, and *Nelumbo nucifera*. *Gnetum montanum* and *Amborella trichopoda* were used as outgroups. Single-copy orthologous genes were identified using OrthoMCL v.2.0.9 (Li et al., 2003). Protein sequences were aligned with MAFFT v.7.407 (Katoh and Standley, 2013). Conserved alignments were subsequently filtered with Gblocks v.0.91b (Talavera and Castresana, 2007). A single-copy gene phylogenetic tree was constructed by the ML method using RAxML v.8.2.12 (Stamatakis, 2006). Then, all single-copy orthologous genes were integrated into a coalescent species tree with ASTRAL v.5.6.1 (Zhang et al., 2018a) using the PROTGAMMALGF model and 500 bootstrap replicates. In addition, all conserved alignments were concatenated into a superalignment matrix. The ML species trees based on the supermatrix were constructed with RAxML v.8.2.12 (Stamatakis, 2006), and a Bayesian phylogenetic species tree was constructed using MrBayes v.3.1.2 (Ronquist and Huelsenbeck, 2003).

Whole-genome duplication analysis

The longest transcript of each gene for representative plants was selected for WGD analysis. For the analysis of intraspecies paralogous genes, we performed a self-to-self alignment with BLAST v.2.6.0+ (− Evalu 1e-10 -max_target_seqs 5), and then collinear blocks were identified using MCScanX (Wang et al., 2012). Synonymous substitution rates (Ks) of codons and four-fold degenerate synonymous site distances between anchored genes were calculated with KaKs_Calculator v.2.0 (Wang et al., 2010) and a custom Perl script, respectively. For analysis of interspecies orthologous genes, the Ks values of different species pairs were calculated separately based on single-copy orthologous genes drawn from the OrthoMCL analysis. The ggplot2 and gggridges packages were used to compute and draw the kernel density estimate. The times of WGD events were calculated based on a synonymous substitution rate of 3.02E-9 synonymous substitutions per year for the basal angiosperms (Cui et al., 2006).

Determination of volatile components

The volatile components from five tissues, including the root xylem, root bark, stem xylem, stem bark, and leaf of *P. bournei*, were extracted by modified headspace solid phase microextraction (SPME) and determined by gas chromatography–mass spectrometry (GC–MS) (Deng et al., 2004; Cechci and Aefi, 2013; Chen et al., 2020c). The SPME fiber used for volatile component extraction was 100-μm polydimethylsiloxane (Sigma–Aldrich, St. Louis, MO, USA) and was conditioned before use as recommended by the manufacturer. A 1.2-g fresh sample was minced against the National Institute of Standards and Technology MS Library (Supplemental Figure 9). Subsequently, we identified nodes presented the divergence points between the four species based on two clades. Ancestral genes were identified according to a previous report (Kim et al., 2006). First, we constructed a phylogenetic tree with protein sequences of all full-length TPSs from *C. kanehirae*, *L. cubeba*, *P. americana*, and *P. bournei* (Supplemental Figure 9). Subsequently, we identified nodes that led to species-specific clades and predicted that these nodes represented the divergence points between the four species based on two criteria: (a) the bootstrap value was higher than 50%; and (b) the relationship among the four species-specific clades was consistent with the
The hypothetical expansion histories of TPS clusters were reconstructed by reconciling both the gene tree and the positions of genes within clusters, according to previous reports (Lan et al., 2009; Ren et al., 2014). Individual genes were assigned to tandem gene clusters if the genes or fragments were closely distributed on the same chromosome; the distances between two TPS genes/fragments were mostly less than 2.7 Mb, except for TPS fragments 11 and 16. Moreover, those genes or fragments were grouped together with a bootstrap value higher than 50 on the phylogenetic tree (Supplemental Figure 7). The tandem duplication events of each cluster were deduced according to the relationship of the cluster-specific phylogenetic tree shown in Figure 4. We assumed a rearrangement event occurred if two gene copies were derived from a tandem duplication but were not located tandemly in the same chromosome region.

Expression analysis of TPS genes

Samples were harvested from leaves, stem xylem, stem bark, root xylem, and root bark of the sequenced P. bournei treelet and frozen quickly in liquid nitrogen before RNA extraction. One-year-old seedlings of P. bournei were hydroponically cultured with drought-stressed condition and without (nonstressed condition) 10% polyethylene glycol PEG 6000. Leaf and root samples were collected for RNA extraction at 0 h, 1 h, 1 day, and 3 days after treatment. Three biological replicates were included for each sample. Two pedigrees (WP and WY) collected from the southern and northern edges of the P. bournei distribution area were considered cold-sensitive and cold-resistant, respectively. Seedlings of these two pedigrees were treated with cold stress (Zhang et al., 2018b). Library construction was performed following the manufacturer’s protocols. The Illumina HiSeq X Ten platform was used for 150-bp paired-end sequencing. The raw data were filtered with fastp v.0.20.0 (Chen et al., 2018) and then mapped to the P. bournei genome using HISAT2 v.2.1.0 (Kim et al., 2015). Gene expression levels were calculated using featureCounts v.1.6.4 (Liao et al., 2014), and differentially expressed genes were detected using the R package edgeR v.3.26.8 (Robinson et al., 2010) based on the criteria \( \log(\text{fold-change}) > 1 \) and false discovery rate \(<0.05.\)

TPS gene cloning and functional characterization

TPS genes in the sequenced P. bournei were amplified with specific primers (Supplemental Table 14) and introduced into the pK2GW7 vector with a 35S::PtTPS overexpression cassette by Gateway-assisted vector construction (Han et al., 2020). The empty and overexpression vectors were transformed into Agrobacterium tumefaciens GV3101 and infiltrated into the leaves of tobacco (Nicotiana benthamiana) for transient expression analysis. After 48 h of growth, the transformed leaves were collected for headspace SPME/GC–MS analysis. The extraction of tobacco VOCs was performed using a procedure similar to that described above but with a few modifications (Wu et al., 2020). In brief, the injector temperature was set to 250°C, and the column flow rate was 1.5 ml min\(^{-1}.\) The program setting for the oven was as follows: held at 50°C for 2 min, gradually increased to 180°C at 3°C min\(^{-1}.\) held at 180°C for 1 min, gradually increased to 230°C at 10°C min\(^{-1}.\) held at 230°C for 5 min, increased to 250°C at 20°C min\(^{-1}.\) and held at 250°C for 3 min. The retention index was calculated using a homologous series of n-alkanes (C7–C30). The authentic standards of \( \beta \)-caryophyllene and \( \alpha \)-copaene were obtained from Sigma–Aldrich. The experiments were performed with three biological replicates.

Evolution of terpene synthases in Phoebe bournei

Most graphs were plotted with the R package ggplot2. Analysis of variance was performed in R using the base package. Final figures were assembled using PowerPoint and Adobe Illustrator.

Graphical and statistical software

This work was supported by the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (grant 2021C02070-10), the National Natural Science Foundation of China (grants 32171828 and 32101545), and the State Key Laboratory of Subtropical Silviculture (grant ZY20180204).

AUTHOR CONTRIBUTIONS

Z.T., Q.Y., X.H., and J.Z. designed the research. X.H., S.Z., L.L., X.L., and L.C. performed the sampling. S.H., S.L.C., M.S., Y.W., C.L., E.L., H.H., and X.H. performed the experiments. Q.Y., G.M., X.H., and S.H. analyzed the data. Q.Y., X.H., J.Z., and Z.T. wrote the manuscript. All authors read and approved the final manuscript. X.H. and J.Z. contributed equally.

ACKNOWLEDGMENTS

We acknowledge Kean-Jin Lim, Yanhui Gao, and Xueyu Liu for experimental assistance and manuscript preparation. Yongjun Wang from Zhejiang A&F University and Shuaifei Chen from the China Eucalypt Research Centre, CAF are gratefully acknowledged for the kind gift of fungus isolates. No conflict of interest is declared.

Received: February 16, 2022
Revised: July 5, 2022
Accepted: July 8, 2022
Published: July 16, 2022

REFERENCES

Alumni, B., Kevei, Z., Redondo-Nieto, M., Kondorosi, A., Mergaert, P., and Kondorosi, E. (2007). Genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in Medicago truncatula. Mol. Plant Microbe Interact. 20:1138–1148.

Aylward, J., Roets, F., Dreyer, L.L., and Wingfield, M.J. (2019). Teratosphaeria stem canker of Eucalyptus: two pathogens, one devastating disease. Mol. Plant Pathol. 20:8–19.

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6:11.

Bedell, J.A., Korf, I., and Gish, W. (2000). MaskerAid: a performance enhancement to RepeatMasker. Bioinformatics 16:1040–1041.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27:573–580.

Binney, E., and Durbin, R. (2000). Using GeneWise in the Drosophila annotation experiment. Genome Res. 10:547–548.
Evolution of terpene synthases in Phoebe bournei

Booth, J.K., Yuen, M.M.S., Jancsik, S., Madilao, L.L., Page, J.E., and Bohlmann, J. (2020). Terpene synthases and terpene variation in cannabis sativa. Plant Physiol. 184:130–147.

Borowinski, N., Hamilton, J.P., Liao, P., Wood, J.C., Dudareva, N., and Buell, C.R. (2020). Genome sequencing of four culinary herbs reveals terpenoid genes underlying chemodiversity in the Nepetoideae. DNA Res. 27:dsaa025.

Boutanaev, A.M., Moses, T., Zi, J., Nelson, D.R., Mugford, S.T., Peters, R.J., and Osbourn, A. (2015). Investigation of terpene diversification across multiple sequenced plant genomes. Proc. Natl. Acad. Sci. USA 112:E81–E88.

Cecchi, T., and Alfei, B. (2013). Volatile profiles of Italian monovarietal extra virgin olive oils via HS-SPME-GC-MS: newly identified compounds, flavors molecular markers, and terpenic profile. Food Chem. 141:2025–2035.

Chang, H.T., Cheng, Y.H., Wu, C.L., Chang, S.T., Chang, T.T., and Su, Y.C. (2008). Antifungal activity of essential oil and its constituents from Calocedrus macrolepis var. formosana Florin leaf against plant pathogenic fungi. Bioresour. Technol. 99:6266–6270.

Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Jouard, P., et al. (2005). Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and Aegilops). Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and Aegilops). Plant Cell 17:1033–1045.

Chaw, S.M., Liu, Y.C., Wu, Y.W., Wang, H.Y., Lin, C.Y.I., Wu, C.S., Ke, H.M., Chang, L.Y., Hsu, C.Y., Yang, H.T., et al. (2019). Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. Nat. Plants 5:65–73.

Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E. (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 J. 68:212–229.

Chen, H., Kollner, T.G., Li, G., Wei, G., Chen, X., Zeng, D., Qian, Q., and Chen, F. (2020a). Combinatorial evolution of a terpene synthase gene cluster explains terpene variations in Oryza. Plant Physiol. 182:480–492.

Chen, J., Hao, Z., Guang, X., Zhao, C., Wang, P., Xue, L., Zhu, Q., Yang, L., Sheng, Y., Zhou, Y., et al. (2019). Liriodendron genome sheds light on angiosperm phylogeny and species-pair differentiation. Nat. Plants 5:18–25.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastsp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:884–890.

Chen, S.P., Sun, W.H., Xiong, Y.F., Jiang, Y.T., Liu, X.D., Liao, X.Y., Zhang, D.Y., Jiang, S.Z., Li, Y., Liu, B., et al. (2020b). The Phoebe genome sheds light on the evolution of magnoliids. Hortic. Res. 7:146.

Chen, Y.C., Li, Z., Zhao, Y.X., Gao, M., Wang, J.Y., Liu, K.W., Wang, X., Wu, L.W., Jiao, Y.L., Xu, Z.L., et al. (2020c). The Litsea genome and the evolution of the laurel family. Nat. Commun. 11:1675.

Cheng, S.-S., Lin, C.-Y., Yang, C.-K., Chen, Y.-J., Chung, M.-J., and Chang, S.-T. (2013). Chemical polymorphism and composition of leaf essential oils of Cinnamomum kanheiariae using gas chromatography/mass spectrometry, cluster analysis, and principal component analysis. J. Wood Chem. Tech. 36:207–219.

Chin, C.S., Peluso, P., Sediazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O’Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods 13:1050–1054.

Cui, L., Wall, P.K., Leebens-Mack, J.H., Lindsay, B.G., Soltis, D.E., Doyle, J.J., Soltis, P.S., Carlson, J.E., Arumuganathan, K., Barakat, A., et al. (2006). Widespread genome duplications throughout the history of flowering plants. Genome Res. 16:738–749.

Della Rocca, G., Danti, R., Hernando, C., Gijarro, M., Michelozzi, M., Carrillo, C., and Madrigal, J. (2020). Terpenoid accumulation links plant health and flammaribility in the cypress-bark canker pathosystem. Forests 11:651.

Deng, C., Zhang, X., Zhu, W., and Qian, J. (2004). Gas chromatography-mass spectrometry with solid-phase microextraction method for determination of methyl salicylate and other volatile compounds in leaves of Lycopersicon esculentum. Anal. Bioanal. Chem. 378:518–522.

Ding, W., Liping, N., Xing, H., Wei, Z., Zhoua, Q., Nong, R., and Chen, J. (2020). Essential oil extracted from leaf of Phoebe bournei (Hemsli) yang: chemical constituents, antitumor, antibacterial, hypoglycemic activities. Nat. Prod. Res. 34:2524–2527.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

Ebadollahi, A., and Sendi, J.J. (2015). A review on recent research results on bio-effects of plant essential oils against major Coleopteran insect pests. Toxin Rev. 34:76–91.

Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput. Biol. 7:e1002195.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., et al. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One 7:e47768.

Fradrich, S.W., Harrington, T.C., and Best, G.S. (2015). Xyleborus glabratus attacks and systemic colonization by Raffaelea lauricola associated with dieback of Cinnamomum camphora in the southeastern United States. For. Pathol. 45:60–70.

Ganko, E.W., Meyers, B.C., and Vision, T.J. (2007). Divergence in expression between duplicated genes in Arabidopsis. Mol. Biol. Evol. 24:2298–2309.

Gershenzon, J., and Dudareva, N. (2007). The function of terpene natural products in the natural world. Nat. Chem. Biol. 3:408–414.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.

Guo, S., Geng, Z., Zhang, W., Liang, J., Wang, C., Deng, Z., and Du, S. (2016). The chemical composition of essential oils from Cinnamomum camphora and their insecticidal activity against the stored product pests. Int. J. Mol. Sci. 17:1836.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVIDencEModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9:R7.

Han, X., An, Y., Zhou, Y., Liu, C., Yin, W., and Xia, X. (2020). Comparative transcriptome analyses define genes and gene modules differing between two Populus genotypes with contrasting stem growth rates. Biotechnol. Biofuels 13:139.

Hu, L., Xu, Z., Wang, M., Fan, R., Yuan, D., Wu, B., Wu, H., Qin, X., Yan, L., Tan, L., et al. (2019). The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. Nat. Commun. 10:4702.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34:2115–2122.

Joshi, S.C., Padalia, R.C., Bisht, D.S., and Mathela, C.S. (2009). Terpenoid diversity in the leaf essential oils of Himalayan Lauraceaee species. Chem. Biodivers. 6:1364–1373.
**Plant Communications**

**Kamle, M., and Kumar, P.** (2016). Colletotrichum gloeosporioides: pathogen of anthracnose disease in mango (mangifera indica L.). In Current Trends in Plant Disease Diagnostics and Management Practices–Kumar, V.K. Gupta, A.K. Tiwari, and M. Kamle, eds. (Springer International Publishing), pp. 207–219.

**Karunanithi, P.S., Berrios, D.I., Wang, S., Davis, J., Shen, T., Fiehn, O., Maloof, J.N., and Zerbe, P.** (2020). The foxtail millet (Setaria italica) terpene synthase gene family. Plant J. 103:781–800.

**Katoh, K., and Standley, D.M.** (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

**Keeling, C.I., Weisshaar, S., Ralph, S.G., Jancsik, S., Hamberger, B., Dullat, H.K., and Bohlmann, J.** (2011). Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (Picea spp.). BMC Plant Biol. 11:43.

**Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., and Hartung, F.** (2016). Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. 44:e69.

**Kim, D., Langmead, B., and Salzberg, S.L.** (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12:357–360.

**Kim, J., Shiu, S.H., Thoma, S., Li, W.H., and Patterson, S.E.** (2006). Patterns of expansion and expression divergence in the plant polygalacturonase gene family. Genome Biol. 7:R87.

**Kulheim, C., Padovan, A., Hefer, C., Krause, S.T., Kollner, T.G., Myburg, A.A., Degenhardt, J., and Foley, W.J.** (2015). The Eucalyptus terpene synthase gene family. BMC genomics 16:450.

**Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K.** (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35:1547–1549.

**Lan, T., Yang, Z.L., Yang, X., Liu, Y.J., Wang, X.R., and Zeng, Q.Y.** (2009). Extensive functional diversification of the Populus glutathione S-transferase supergene family. Plant Cell 21:3749–3766.

**Lefort, V., Longueville, J.E., and Gascuel, O.** (2017). SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 45:D458–D460.

**Li, G., Liu, F., Li, J., Liu, Q., and Chen, S.** (2016). Characterization of Botryosphaeria dothidea and Lasiodiplodia pseudotheobromae from English walnut in China. J. Phytopathol. 164:348–353.

**Li, L., Stoeckert, C.J., Jr., and Roos, D.S.** (2003). OrthoMCL: identification of orthologous groups for eukaryotic genomes. Genome Res. 13:2178–2189.

**Li, T., and Min, X.** (2020). Dormancy characteristics and germination requirements of Phoebe bournei seed. Scientia Horticulturae 260:108903.

**Liao, Y., Smyth, G.K., and Shi, W.** (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.

**Ma, J., Sun, P., Wang, D., Wang, Z., Yang, J., Li, Y., Mu, W., Xu, R., Wu, Y., Dong, C., et al.** (2021). The Chloranthus sessilifolius genome provides insight into early diversification of angiosperms. Nat. Commun. 12:6929.

**Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al.** (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 47:W636–W641.

**Marsberg, A., Keimer, M., Jami, F., Nagel, J.H., Postma-Smidt, A., Naidoo, S., Wingfield, M.J., Crous, P.W., Spatafora, J.W., Hesse, C.N., et al.** (2017). Botryosphaeria dothidea: a latent pathogen of global importance to woody plant health. Mol. Plant Pathol. 18:477–488.

**Martin, D.M., Aubourg, S., Schouwney, M.B., Daviet, L., Schalk, M., Toub, O., Lund, S.T., and Bohlmann, J.** (2010). Functional annotation, genome organization and phylogeny of the grapevine (Vitis vinifera) terpene synthase gene family based on genome assembly, FlcDNA cloning, and enzyme assays. BMC Plant Biol. 10:226.

**Martin, D.M., Gershenzon, J., and Bohlmann, J.** (2003). Induction of volatile terpene biosynthesis and diurnal emission by methyl jasmonate in foliage of Norway spruce. Plant Physiol. 132:1586–1599.

**McCaskill, D., and Croteau, R.** (1997). Prospects for the bioengineering of isoprenoid biosynthesis. Adv Biochem Eng Biotechnol 55:107–146. https://doi.org/10.1007/BFb0102064.

**McDowell, J.M., Dhandaydham, M., Long, T.A., Aarts, M.G., Goff, S., Holub, E.B., and Dangi, J.L.** (1998). Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis. Plant Cell 10:1861–1874.

**Méndez-González, C., and Espinosa-García, F.J.** (2018). Metabolic profiling of Persea americana cv. Hass branch volatiles reveals seasonal chemical changes associated to the avocado branch borer, Coputus aguacatea. Scientia Horticulturae 240:116–124.

**Mewalal, R., Rai, D.K., Kainer, D., Chen, F., Kulheim, C., Peter, G.F., and Tuskan, G.A.** (2017). Plant-derived terpenes: a feedstock for specialty biofuels. Trends Biotechnol. 35:227–240.

**Mitchell, A.L., Atwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., et al.** (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 47:D351–D360.

**Nam, J., Kim, J., Lee, S., An, G., Ma, H., and Nei, M.** (2004). Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. Proc. Natl. Acad. Sci. USA 101:1910–1915.

**Nerg, A.M., Heijari, J., Noidt, U., Viritinen, H., Vuorinen, M., Kainulainen, P., and Holopainen, J.K.** (2004). Significance of wood terpenoids in the resistance of Scots pine provenances against the old house borer, Hylotrupes bajulus, and brown-rot fungus, Coniophora puteana. J. Chem. Ecol. 30:125–141.

**Niogret, J., Epsky, N.D., Schnell, R.J., Boza, E.J., Kendra, P.E., and Heath, R.R.** (2013). Terpenoid variations within and among half-sibling avocado trees, Persea americana Mill. (Lauraceae). PLoS One 8:e73601.

**Ogounbuinu, A.O., Ogunwaned, I.A., Flaminid, G., and Cionid, P.L.** (2007). Volatile compounds of Persea americana Mill from Nigeria. J. Essent. Oil Bearing Plants 10:133–138.

**Pichersky, E., and Raguso, R.A.** (2018). Why do plants produce so many terpenoid compounds? New Phytol. 220:692–702.

**Ploetz, R., Kendra, P., Choudhury, R., Pollins, J., Campbell, A., Garrett, K., Hughes, M., and Dreden, T.** (2017). Laurel wilt in natural and agricultural ecosystems: understanding the drivers and scales of complex pathosystems. Forests 8:48.

**Price, M.N., Dehal, P.S., and Arkin, A.P.** (2010). FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.

**Ren, L.L., Liu, Y.J., Liu, H.J., Qian, T.T., Qi, L.W., Wang, X.R., and Zeng, Q.Y.** (2014). Subcellular relocalization and positive selection play key roles in the retention of duplicate genes of Populus class III peroxidase family. Plant Cell 26:2404–2419.

**Rendón-Anaya, M., Ibarra-Laclette, E., Méndez-Bravo, A., Lan, T., Zheng, C., Carretero-Paulet, L., Perez-Torres, C.A., Chacón-López, A., Hernandez-Guzman, G., Chang, T.H., et al.** (2019). The avocado genome informs deep angiosperm phylogeny, highlights...
Evolution of terpene synthases in *Phoebe bournei*

introgressive hybridization, and reveals pathogen-influenced gene space adaptation. Proc. Natl. Acad. Sci. USA 116:17081–17089.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Shang, J., Tian, J., Cheng, H., Yan, Q., Li, L., Jamal, A., Xu, Z., Xiang, L., Saksi, C.A., Jin, S., et al. (2020). The chromosome-level wintersweet (Chimonanthus praecox) genome provides insights into floral scent biosynthesis and flowering in winter. Genome Biol. 21:200.

Silva, F.F.A.d., Alves, C.C.F., Oliveira Filho, J.G.d., Vieira, T.M., Crotti, A.E.M., and Miranda, M.L.D. (2019). Chemical constituents of essential oil from *Murraya paniculata* leaves and its application to in vitro biological control of the fungus *Sclerotinia sclerotiorum*. Food Sci. Technol. 39:413–417.

Spannagl, M., Nussbaumer, T., Bader, K.C., Martis, M.M., Seidel, M., Kugler, K.G., Gundlach, H., and Mayer, K.F.X. (2016). PGSB PlantsDB: updates to the database framework for comparative plant genome research. Nucleic Acids Res. 44:D1141–D1147.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19 (Suppl 2):ii215–ii225.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

Tian, J.P., Ma, Z.Y., Zhao, K.G., Zhang, J., Xiang, L., and Chen, L.Q. (2019). Transcriptomic and proteomic approaches to explore the differences in monoterpene and benzenoid biosynthesis between scented and unscented genotypes of wintersweet. Physiol. Plant 166:478–493.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakhikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs.Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics 8:77–80.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40:e49.

Wei, F., and van der Werff, H. (2008). *Phoebe*. In Flora of China, Z.Y. Wu, P.H. Raven, and D.Y. Hong, eds., 7 (Science Press), pp. 189–200. St. Louis: Missouri Botanical Garden Press.

Wernersson, R., and Pedersen, A.G. (2003). RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res. 31:3537–3539.

Wu, L., Zhao, Y., Zhang, Q., Chen, Y., Gao, M., and Wang, Y. (2020). Overexpression of the 3-hydroxy-3-methylglutaryl-CoA synthase gene LcHMGS effectively increases the yield of monoterpenes and sesquiterpenes. Tree Physiol. 40:1095–1107.

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17:32–43.

Ye, J., Li, D., and Hampe, A. (2019). Differential Quaternary dynamics of evergreen broadleaved forests in subtropical China revealed by phyleogeography of *Lindera aggregata* (Lauraceae). J. Biogeogr. 46:1112–1123.

Zang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018a). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.

Zhang, J., Zhu, Y., Pan, Y., Huang, H., Li, C., Li, G., and Tong, Z. (2018b). Transcriptomic profiling and identification of candidate genes in two *Phoebe bournei* ecotypes with contrasting cold stress responses. Trees-Struct. Funct. 32:1315–1333.

Zhou, F., and Pichersky, E. (2020a). The complete functional characterisation of the terpene synthase family in tomato. New Phytol. 226:1341–1360.

Zhou, F., and Pichersky, E. (2020b). More is better: the diversity of terpene metabolism in plants. Curr. Opin. Plant Biol. 55:1–10.