Research paper

Genomic insights into ecological adaptation of oaks revealed by phylogenomic analysis of multiple species

Tian-Rui Wang [a, 1], Xin Ning [a, 1], Si-Si Zheng [a, 1], Yu Li [a, b], Zi-Jia Lu [a, c], Hong-Hu Meng [d], Bin-Jie Ge [a], Gregor Kozlowski [a, e, f], Meng-Xiao Yan [a, **], Yi-Gang Song [a, b, *]

[a] Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China
[b] College of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou 311300, China
[c] College of Life Sciences, Shanghai Normal University, Shanghai 200234, China
[d] Plant Phylogenetics and Conservation Group, Centre for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, China
[e] Department of Biology and Botanic Garden, University of Fribourg, Fribourg, Switzerland
[f] Natural History Museum Fribourg, Fribourg, Switzerland

ARTICLE INFO

ABSTRACT

Understanding the ecological adaptation of tree species can not only reveal the evolutionary potential but also benefit biodiversity conservation under global climate change. *Quercus* is a keystone genus in Northern Hemisphere forests, and its wide distribution in diverse ecosystems and long evolutionary history make it an ideal model for studying the genomic basis of ecological adaptations. Here we used a newly sequenced genome of *Quercus gilva*, an evergreen oak species from East Asia, with 18 published Fagales genomes to determine how Fagaceae genomes have evolved, identify genomic footprints of ecological adaptability in oaks in general, as well as between evergreen and deciduous oaks. We found that oak species exhibited a higher degree of genomic conservation and stability, as indicated by the absence of large-scale chromosomal structural variations or additional whole-genome duplication events. In addition, we identified expansion and tandem repetitions within gene families that contribute to plant physical and chemical defense (e.g., cuticle biosynthesis and oxidosqualene cyclase genes), which may represent the foundation for the ecological adaptation of oak species. Circadian rhythm and hormone-related genes may regulate the habits of evergreen and deciduous oaks. This study provides a comprehensive perspective on the ecological adaptations of tree species based on phylogenetic, genome evolutionary, and functional genomic analyses.

1. Introduction

Global climate change has profound effects on the survival and distribution of species and poses a serious threat to biodiversity and natural ecosystems (Walther et al., 2002; Díaz and Malhi, 2022). Forest trees dominate many terrestrial ecosystems and play an increasingly vital role in protecting biodiversity, maintaining the global carbon cycle balance, and responding to global climate change (Aitken et al., 2008). Forest trees spread and settled across diverse habitats primarily through ecological adaptations (Aitken et al., 2008; de Lafontaine et al., 2018). Therefore, exploring ecological adaptation mechanisms is important for predicting forest tree responses to climate change and developing management strategies to assist biodiversity conservation (Waldvogel et al., 2020).

Numerous studies have been conducted on the ecological adaptation footprints of tree species by assessing genetic variations associated with climate variables and evaluating millions of variants across the genome of a species (Dauphin et al., 2021; Gougherty et al., 2021). However, only a few of these variants are

expected to be related to ecological adaptation, and the mechanisms underlying ecological adaptation in trees at the genomic level are poorly understood. With advances in genomic technologies, whole-genome data have provided higher-resolution results and greatly accelerated adaptation research over the last decade (Ellegren, 2014). Predicting climate change impacts on a single species may be insufficient for analyzing the extensive and profound impacts of global climate change on species, communities, and even ecosystems (Capblancq et al., 2020). Genomic sequences of multiple closely related species provide important information about biological factors such as the evolutionary history, genome evolution, and functional gene identification, thereby clarifying the basis of species' ecological adaptations (He et al., 2022; Sun et al., 2022a). Therefore, ecological adaptation mechanisms should be studied at the multispecies level based on whole-genome data, and phylogenomics provides a comprehensive and reliable platform for the study of genome evolution and adaptation of a species (Tang et al., 2022; Shao et al., 2023).

Quercus (oak) belongs to the family Fagaceae, and is one of the most speciose and widely distributed genera, with approximately 450 species in the Northern Hemisphere (Hipp et al., 2020; Kremer and Hipp, 2020). The genus evolved at approximately 56 million years ago (Ma) and has experienced numerous climatic events (such as the climatic optimum during the early Eocene, climate warming during the Miocene, and the succession of glacial and interglacial periods during the Pleistocene), eventually spreading to form its present diversity pattern (Hipp et al., 2020; Kremer and Hipp, 2020; Jin et al., 2023). Owing to their remarkable adaptability to environmental stresses, oaks have become dominant species in temperate and subtropical forests across North America, Europe, and Asia, and they play a crucial role in shaping biodiversity and carbon sequestration to mitigate climate warming (Sork et al., 2022). Oaks exhibiting remarkable adaptability in response to drought (Du et al., 2020), cold (Meng et al., 2019), and diverse environmental stress (Rellstab et al., 2016; Sork et al., 2016; Martins et al., 2018; Leroy et al., 2020; Gugger et al., 2021). However, it remains unclear how oaks evolved with such extensive ecological adaptability, and whether there are differences in the adaptive evolution between different types of oaks (i.e., evergreen and deciduous oaks), which inhabit different environmental conditions. In recent years, oak genome resources have been increasing rapidly (Plomion et al., 2018; Fu et al., 2022; Sork et al., 2022; Zhou et al., 2022a). Thus, as an ecologically adaptable species with a long evolutionary history, oak is a good model for investigating the mechanisms of genome response and ecological adaptation under global climate change using phylogenomics. Deciphering the ecological adaptation mechanisms of evergreen and deciduous oaks holds profound significance in gaining a deeper understanding of how trees respond to global climate change.

Here, we used a newly sequenced genome of *Quercus gilva*, an evergreen oak species from East Asia, with 18 published Fagales genomes to determine how Fagaceae genomes have evolved. We then integrated oak genomes and transcriptomes to explore the genomic footprints and gene regulation underlying the wide ecological adaptability of oaks. Finally, we investigated the genomic basis underlying the different ecological adaptations between evergreen and deciduous oaks.

2. Materials and methods

2.1. Plant materials, DNA and RNA extraction

Quercus gilva was sampled from Fujian province (29.94°N, 116.87°E; Fig. S1) for genome sequencing. High-quality genomic DNA was isolated from fresh leaves using the cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1986). DNA quality and concentration were tested by 0.75% agarose gel electrophoresis, a NanoDrop One spectrophotometer (Thermo Fisher Scientific), and a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). Leaves of *Q. gilva* were collected and used for RNA sequencing. RNA purity was checked using a Kaiao K5500® Spectrophotometer (Kaiao, Beijing, China). RNA integrity and concentration were assessed using an RNA Nano 6000 Assay Kit on a Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

2.2. Genome sequencing and RNA sequencing

Illumina sequencing paired-end (PE) libraries with insert fragments of approximately 350 bp were prepared using the Nextera DNA Flex Library Prep Kit. Sequencing was performed using the Illumina HiSeq XTen platform with 150 bp PE reads. Raw reads were cleaned to discard low-quality reads with adapters, unknown nucleotides (Ns), or more than 20% low-quality bases using the SOAPnuke v.2.1.4 tool (Chen et al., 2018a). An Oxford Nanopore library was prepared using the SQK-LSK109 ligation kit. The purified library was sequenced using an Oxford Nanopore PromethION sequencer at Wuhan Benagen Tech Solutions Company Limited, Wuhan, China.

A total of 2 μg RNA was used to generate libraries with the NEBNext® Ultra ® RNA Library Prep Kit for Illumina®. RNA concentration was measured using the Qubit® RNA Assay Kit in Qubit® 3.0. Libraries were sequenced on an Illumina platform, and 150-bp PE reads were generated. Raw reads were further cleaned by removing adapters and low-quality bases.

2.3. Genome assembly and assessment

We first estimated the size of the *Quercus gilva* genome using a K-mer analysis-based approach together with Illumina PE short reads. GCE v.1.0.0 (Liu et al., 2013) was used to count the K-mer in the DNA samples and estimate the genome size. The initial assembly sequences were obtained using NextDenovo (https://github.com/Nextomics/NextDenovo). After two rounds of error correction of the splicing results, the final assembly sequences were obtained based on two rounds of Illumina reads using Pilon v.1.23 (Walker et al., 2014). Finally, the quality of the genome assembly was evaluated by Illumina mapping rate and coverage (Simão et al., 2015).

2.4. Hi-C library construction and assembly of the chromosome

A 2% formaldehyde solution (40 mL) was used to crosslink DNA samples of *Quercus gilva*. DNA was restriction digested with DpnII to cut crosslinked DNA samples and produce sticky ends on 200–600bp fragments. After the library was constructed, Qubit 2.0 and Agilent 2100 were used to detect the library concentration and insert size, respectively. Finally, the Illumina platform was used for high-throughput sequencing with a PE sequencing read length of 150 bp. High-quality clean data were obtained by removing adapters and low-quality reads using Fastp v.0.21.0 (Chen et al., 2018b). Clean data were mapped to the assembled draft genome using HICUP v.0.7.2 (Wingett et al., 2015). The contig sequences were then clustered into multiple chromosomal groups using the ALLHIC algorithm. Contig sequences were further oriented and manually ordered and oriented using JuiceBox v.1.8.8 (Durand et al., 2016). Finally, the N base was used to complete the gap, and the final chromosome-level genome was obtained. Contigs with no interactions were considered unmounted fragments.

2.5. Genome annotation

Homologous, ab initio, and transcription predictions were combined to predict the gene structure in *Quercus gilva*. First, related species (*Quercus suber*, *Juglans regia*, and *Amborella trichopoda*) were selected for homology prediction using Exonerate v.2.4.0 (https://github.com/nathanweeks/exonerate). Genes were then predicted de novo from DNA sequences with ab initio approaches based on Augustus v.3.3.2 (Stanke and Waack, 2003), Genscan v.1.0 (Burge and Karlin, 1997), and GlimmerHMM v.3.0.4 (Majoros et al., 2004), with default settings. Finally, transcripts of the RNA-seq data were reconstructed using StringTie v.2.1.1 (Shumate et al., 2022) and the coding frame was predicted using TransDecoder v.5.5.0 (Haas et al., 2013).

All assembled genes identified by the above methods were combined using MAKER v.2.31.10 (http://yandell.topaz.genetics.utah.edu/cgi-bin/maker_license.cgi). Gene function annotations were performed based on sequence and motif similarities. First, the protein sequences encoded by the genes in the gene set were compared with the Uniport, Nr databases, and Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathway database (Kanehisa et al., 2004) using the BLAST + program v.2.6.0 (Camacho et al., 2009). Functional information on the sequences and metabolic pathway information of the proteins was then obtained. KEGG annotations were associated with KEGG orthology and pathway classifications using KOBAS v.2.0 (Xie et al., 2011). The corresponding relationships between each protein family recorded in the UniProt database and functional nodes in the Gene Ontology (GO) database were used to predict the biological functions of the protein sequences encoded by genes (Ashburner et al., 2000). InterProScan v.5.33 (Jones et al., 2014) was used to compare various sub-databases of InterPro to obtain the conserved sequences, motifs, and domains of proteins. The structural domain was further predicted using hmmscan v.3.1 (https://www.ebi.ac.uk/Tools/hmmer/). We used BUSCO to evaluate the completeness of the genome annotations.

2.6. Gene family analysis

Gene family clustering was performed using the OrthoFinder program v.2.3.12 (Emms and Kelly, 2019). All-to-all blast searches using the BLAST + program v.2.6.0 (Camacho et al., 2009) were conducted on the protein-coding genes of 21 species, including six species of oaks (*Q. gilva*, *Q. aquifolioides*, *Q. mongolica*, *Q. robur*, *Q. lobata*, *Q. suber*), *Castanea mollissima*, *C. crenata*, *Fagus sylvatica*, *Betula platyphylla*, *Carpinus fangiana*, *Corylus mandshurica*, *Alnus glutinosa*, *Morella rubra*, *Juglans regia*, *Platycarya strobilacea*, *Pterocarya stenoptera*, *Carya illinoinensis*, *Casuarina equisetifolia*, *Amborella trichopoda*, and *Vitis vinifera* (Table S1). Gene families of the four oak species (*Q. gilva*, *Q. lobata*, *Q. suber*, and *Q. robur*) were selected for Venn analysis based on a cluster of protein sequences.

We also evaluated GO term functional enrichment and KEGG pathway analysis of the significantly shared and unique gene families in *Q. gilva* species. All target genes were first mapped to GO terms in the database. Gene numbers of every term were calculated and enriched significant GO terms for shared and unique gene families compared to the genome background were identified using the clusterProfiler package (Wu et al., 2021). Significantly enriched pathways in the target genes were identified based on KEGG pathway enrichment analyses. The adjusted p value was calculated to screen for significantly enriched GO terms and KEGG pathways ($p < 0.05$).

2.7. Phylogenetic reconstruction and divergence time prediction

We constructed phylogenetic trees using 649 single-copy genes shared in 90% of 21 Fagaceae species (see section 2.6). Protein sequences of single-copy gene families shared by these 21 species were fully aligned using MUSCLE v.3.8.31 (Edgar, 2004). Positions with gaps and missing data for each single-copy gene family were compared and filtered using TrimAI v.1.4 (Capella-Gutiérrez et al., 2009). Trimmed sequences were concatenated to form the final alignment matrix. A concatenated alignment was then used to construct a maximum likelihood (ML) phylogenetic tree using RAxML v.8.2.11 with the PROTGAMMAWAG matrix-based model and 1000 bootstrap replicates (Stamatakis, 2014). Because recombination or incomplete lineage sorting may cause divergence in the evolutionary history of different genomic regions or genes, we simultaneously constructed phylogenetic trees using the ASTRAL-III v.5.6.3 method based on coalescence strategy (Zhang et al., 2018). In ASTRAL-III, we generated a summary species tree based on the 649 single-copy gene trees, with posterior probabilities calculated at each node (Sayyari and Mirarab 2016). Divergence time of the derived phylogenetic tree was inferred using the MCMCtree program (Puttick, 2019), implemented via phylogenetic analysis using the ML method, with *Arabidopsis trichopoda* as an outgroup. MCMCtree analysis was executed using the following parameters: burn-in, 8,000,000; and sample number, 3,000,000. For the divergence time estimation, we calibrated the model using fossil dates of the Fagaceae stem (Xiang et al., 2014) (95.5—101.2 Ma), Fagaceae crown (Xiang et al., 2014; Zhou et al., 2022b) (81—83.5 Ma), oaks (Hipp et al., 2020) (55.18—60.52 Ma), and Juglandoideae (Kozlowski et al., 2018) (65—55 Ma).

2.8. Expansion and contraction of gene families

Analysis of orthologous gene family expansion and contraction was performed by comparing the cluster sizes of Fagaceae 21 species using CAFÉ v.4.2 (Han et al., 2013). We also analyzed gene families for signs of expansion or contraction in these 21 species based on ML modeling of gene gain and loss. Rapid expansion/contraction was indicated by statistical significance and non-random expansion/contraction at $p < 0.05$, as described by Han et al. (2013). We derived an ML phylogenetic tree using single-copy orthologs across the 21 plant species used for gene family construction, as described above. The p value < 0.05 was used to identify gene families that expanded and contracted at a specific lineage or species. Orthogroups that were specifically contracted or expanded in plants producing MIAs were further investigated around *Quercus gilva*.

2.9. Whole-genome duplication

To investigate genome evolution across oak species, we searched for WGD events in four species—four oak species (*Quercus gilva*, *Q. aquifolioides*, *Q. mongolica*, and *Q. lobata*). To avoid the effect of assembly quality, we removed *Q. robur* and *Q. suber*, which were poorly assembled. In addition, we also investigated WGD events for *Fagus* (*F. sylvatica*) and *Castanea* (*C. mollissima*). First, protein sequences were compared using BLAST v.2.6.0+ with the default parameter set, and then colinear blocks of the genome in these species were analyzed using an improved version of colinearScan (*icl*) model of WGDI (Sun et al., 2022b). We then estimated the *Ka* (number of substitutions per non-synonymous site), *Ks* (number of substitutions per synonymous site), and *Ka*/*Ks* values for gene pairs generated by different modes of duplication using PAML in conjunction with the YN00 model for each node in the gene family phylogenies (Yang, 2007). A dot plot of the synteny blocks with the *Ks* values was obtained using the block *Ks* (*bk*) model in WGDI (Sun et al., 2022b). We used *Ks* fit (*kf*) model in WGDI to map the distribution of synonymous nucleotide substitutions.

To compare the genome structure of the studied oaks, we assembled and performed a collinearity search for four oak species with high-quality assembled genomes (*Quercus gilva*, *Q. aquifolioides*, *Q. mongolica*, *Q. lobata*), one *Castanea* species (*C. mollissima*), and one *Fagus* species (*F. sylvatica*). Next, we performed all-against-all LAST (Kiełbasa et al., 2011) and chained the LAST hits with a distance cut-off of ten genes, which required at least five gene pairs per synteny block. The syntenic "depth" function implemented in MCScanX was applied to estimate the duplication history in the respective genomes. Genomic synteny was visualized and compared using the Python JCVI library. The resulting dot plots were inspected to confirm the polyploidy level of the *Q. gilva* genome to other genomes.

2.10. Functional gene identification

We compared the genes of five oak species (*Quercus suber*, *Q. gilva*, *Q. lobata*, *Q. mongolica*, and *Q. robur*; genome data were the same as those in the previous analysis). *Q. aquifolioides* was removed due to the low annotation quality. The reference sequence, hidden Markov models (HMMs), and identified genes are listed in Tables S2 and S3.

2.10.1. Identification of plant cuticle biosynthesis genes
Amino acid sequences of the wax genes of *Arabidopsis thaliana* and glycerol-3-phosphate acyltransferase (GPAT) genes were obtained from the TAIR database (https://www.arabidopsis.org/). The amino acid sequences of ten wax biogenesis genes from *A. thaliana* were used as bait genes and used the BLAST GUI in TBtools (Chen et al., 2020) to identify genes in oaks (Table S2). In addition, HMMs of the gene domains related to epidermal wax biosynthesis were downloaded from the Protein Family (Pfam) database 35.0 (http://pfam.xfam.org/) (Table S2). HMMER 3.0 (Finn et al., 2011) was used to search for all putative genes related to cuticular wax biosynthesis in the genomes of the five species (e-value cut-off of 1e-5). Candidate genes identified by BLASTP and HMMER were further checked using the NCBI online BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Two cytochrome P450 (*CYP*) genes are involved in cutin biosynthesis: *CYP86A* and *CYP77A*. The HMM profile of Pfam PF00067 was used to extract full-length *CYP* candidates from the five oak genomes using HMMER 3.0 (Kumar et al., 2017), which were filtered by length (i.e., 400 to 600 amino acids).

2.10.2. Identification of OSC genes
The amino acid sequences of *QsOSC1-3* in *Quercus suber* (Busta et al., 2020) were used as queries and used BLAST GUI in TBtools to screen genes (Table S3). We then searched and filtered out hits that presented the longest open reading frame that was less than 1500bp (virtually, all plant *OSCs* characterized thus far are longer than 2000 bp) or had gaps in the SDCTAE motif (Busta et al., 2020). The filtered genes were tentatively identified as pseudogenes. All the identified *OSC* gene sequences were aligned using MAFFT v.7.310 (Katoh and Standley, 2013). An ML phylogenetic tree was constructed using IQ-TREE v.2.0.3 software with 1000 ultrafast bootstrap replicates (Nguyen et al., 2015). *OSC1—5* genes were identified based on their phylogenetic relationships with *QsOSC1—3*. iTOL V6 (Interactive Tree of Life, https://itol.embl.de/). was used to visualize the phylogenetic tree.

In addition, we employed ParaAT v.2.0 and KaKs_Calculator v.3.0 to calculate the selection pressure on key genes involved in the biosynthesis pathways of cuticle and triterpenoids in five oak species (Zhang et al., 2012; Zhang, 2021). Using ParaAT v.2.0, we aligned the protein sequences in parallel and then reverse-translated these alignments into the corresponding DNA sequences. Finally, we selected the YN model to calculate the final $Ka/$ Ks value in KaKs_Calculator v.3.0 with significance set at $p < 0.05$ (Zhang, 2021; Tian et al., 2015).

2.11. Analysis of gene synteny

Syntenic gene pairs were identified using an all-vs-all BLAST search using LAST v.1389 (Frith et al., 2010). Syntenic gene pairs were then filtered to remove pairs with scores below 0.7 and clustered into syntenic blocks using MCScan Toolkit v.1.1 (Wang et al., 2012), which was implemented in Python. Shared tandem duplication among different species was confirmed by the syntenic gene pairs of the identified genes and the upstream and downstream genes.

2.12. RNA-seq analysis

We downloaded RNA-seq data of leaves and other tissues from five oak species (Table S4; Fig. S2). Raw data were first processed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to filter adapters and low-quality sequences. The RNA-seq reads were mapped to the reference genome using HISAT2 v.2.1.0 (Kim et al., 2019). The expressed genes were assembled and quantified using StringTie v.2.1.4 (Pertea et al., 2015) to calculate the gene expression levels for each sample, which were expressed as fragments per kilobase of transcript per million fragments mapped (FPKM). The expression matrix was extracted using the R package Ballgown (Frazee et al., 2015) and the expression heatmap was visualized using the TBtools HeatMap tool.

2.13. Positively selected genes

Positively selected genes within the leaf habits (deciduous: *Quercus lobata*, *Q. mongolica*, and *Q. robur*; and evergreen: *Q. suber*, *Q. aquifolioides,* and *Q. gilva*) were identified using the CodeML module in PAML v.4.9. First, single-copy gene families of the selected species for the deciduous and evergreen leaf habits were obtained. Protein sequences of each gene family were then compared using MAFFT (Katoh and Standley, 2013) (parameter: maxiterate: 1000). PAL2NAL was used to invert the protein sequence into a codon alignment sequence (Suyama et al., 2006). Next, the branch-site model in the CodeML module was used to perform likelihood ratio tests. Model A (meaning that the foreground branch ω represents positive selection, i.e., ω > 1) and the NULL Model (meaning that the ω value of any locus is not allowed to be greater than 1) were tested using the chi 2 program in PAML. Deciduous and evergreen oaks trees were used as foreground branches to identify positively selected genes of deciduous and evergreen oaks species. Finally, genes with Ka/Ks values greater than 1 that presented a significant p value of less than 0.05 were identified as positively selected genes.

3. Results

3.1. Genome assembly of Quercus gilva

We assembled a chromosome-level genome of *Q. gilva* using Illumina short reads, Nanopore long-read sequencing, and high-throughput chromosome conformation capture (Hi-C) technologies (Table 1). A final assembly of 894.8 Mb anchored onto 12 pseudochromosomes with a contig N50 of 45.4 Mb was obtained (Tables 1 and S5-S7; Figs. S1 and S3). The new *Q. gilva* genome assembly had a larger N50 value and 100.0% anchoring rate of chromosome, an improvement over previous assemblies (Plomion et al., 2018; Fu et al., 2022; Sork et al., 2022; Zhou et al., 2022a). The benchmarking universal single-copy orthologues (BUSCO)

Table 1
Genome assembly and annotation of *Quercus gilva*.

	Item	Value
Assembly	Assembly level	Chromosome
	Total length (Mb)	894.8
	Number of contigs	42
	N50 of contigs (Mb)	45.5
	Percentage of sequence anchored on chromosome (%)	100
	GC content (%)	35.2
Annotation	Number of protein-coding genes	40,177
	Average gene length (kb)	6.8
	Average CDS length (kb)	1.0
	Average exons per gene	4.5
	Repeat sequences length (Mb)	516.1
	Percentage of repeat sequences (%)	57.7
	Number of tRNA	647
	Number of rRNA	248
	Number of snRNA	772
	Complete BUSCOs (%)	97.8

Note: BUSCO: benchmarking universal single-copy orthologs; tRNA: transport RNA; rRNA: ribosomal RNA; snRNA: small Nucleolus RNA.

evaluation (97.1%) and a mapping rate (97.0%) of Illumina sequencing data revealed high genome completeness (Tables S5 and S7). Finally, 36,985 protein-coding genes were clustered into 22,565 gene families in the *Q. gilva* genome (Fig. S4; Tables S1, S8 and S9).

The shared and species-specific gene families of *Quercus gilva* were obtained by comparing the gene families of *Q. gilva* with those of three other high-quality oak genomes (*Q. robur, Q. suber,* and *Q. lobata*). A total of 10,838 gene families were shared among the four oak species, and 6375 gene families were unique to *Q. gilva* (Fig. S1). The shared gene families of the four oak species were enriched in flavone and flavonol biosynthesis, calcium ion binding, monoterpenoid biosynthesis, quercetin, salicylic acid metabolic processes, glycosyltransferases, and acyltransferases (Tables S10 and S11). Species-specific gene families of *Q. gilva* were particularly enriched in the citrate cycle (TCA cycle), carbon metabolism, and phenylpropanoid biosynthesis (Tables S12 and S13).

3.2. Phylogeny and evolutionary history of Fagaceae

We combined our newly sequenced *Quercus gilva* genome with those of 18 published Fagales species to generate a well-resolved and strongly supported phylogeny that showed relationships between the major clades and families of Fagales. Concatenation and coalescence strategies yielded phylogenetic trees with consistent topologies (Figs. 1, S4, S5). Fagaceae was supported as a sister group of the core Fagales. Within the core Fagales, Juglandaceae, and Myricaceae formed one lineage that was strongly supported as a sister to the lineage formed by Casuarinaceae and Betulaceae. Within the oak genus, three species of section *Quercus* were included in the Subgenus *Quercus* which is sister to the clade of the Subgenus *Cerris*. Section *Cyclobalanopsis* (*Q. gilva*) is sister to a clade consisting of sections *Cerris* and *Ilex*.

The stem ages of Fagaceae and the core Fagales were estimated at 82.6 Ma (95% HPD: 81.1–83.5) and 91.3 Ma (95% HPD: 86.6–95.5), respectively. Our time estimates are similar to those of previous fossil calibrations (Xiang et al., 2014; Xing et al., 2014; Larson-Johnson, 2016). The diversification of the crown group of the core Fagales (91.3 Ma) was between previous dating: 93.4 Ma (Xiang et al., 2014) and 88 Ma (Larson-Johnson, 2016). The oak genus was estimated to have split from *Castanea* at 56.8 Ma (95% HPD: 55.0–60.1). Within oak species, the crown age of section *Quercus* was estimated at 38.9 Ma (95% HPD: 25.9–48.0), with the

Eurasian lineage originating at 29.4 Ma (95% HPD: 17.2–40.6). The Subgenus *Cerris* was estimated to have a crown age of 54.6 Ma (95% HPD: 52.0–58.4), followed by the origin of section *Cerris* + *Ilex* at 49.9 Ma (95% HPD: 44.0–55.4) (Fig. 1). The crown age of the Subgenus *Cerris* is slightly older than previously reported 54.15 Ma (Hipp et al., 2020) and 50.53 Ma (Jiang et al., 2019). The crown age of section *Quercus* (38.9 Ma) is much older than that of Ai et al. 2022(17.0 Ma), whereas it is similar to the mean age between the crown and stem calibration analysis (28.11 and 45 Ma) of Hipp et al. (2020). Section *Quercus* was estimated to have arrived in Eurasia during the Oligocene, as proposed by Hipp et al. (2020). Thus, our results support the view that all *Quercus* sections originated between the early and late Eocene (Hipp et al., 2020).

3.3. Gene families and genome structure of Quercus

The expansion and contraction of gene families have been well-documented as crucial driving forces in lineage splitting and functional diversification (Chen et al., 2013). For *Quercus*, we found that the expanded gene family fluctuated from 885 (*Q. robur*) to 2195 (*Q. suber*), while the contracted gene family fluctuated from 725 (*Q. suber*) to 3566 (*Q. robur*). Among the oak species, *Q. robur* contracted the most, with 3566 gene families undergoing contraction. *Q. suber* expanded the most, with 2195 gene families undergoing expansion. We further detected 1278 gene families in *Q. gilva* that underwent expansion and 754 gene families that underwent contraction (Fig. 2A). Enrichment analyses revealed that the expanded gene families of *Q. gilva* were specifically enriched in cutin, suberin, wax, terpenoid, flavone, and flavonol biosynthesis ($p < 0.05$) (Tables S14 and S15).

To investigate the genome structure stability of *Quercus* species, macrosynteny analysis was performed by comparing the genomes of oak species with those of *Castanea* and *Fagus* species. Collinearity analysis showed that a nearly one-to-one correspondence syntenic relationship occurred, and large-scale chromosomal rearrangement or inversion was not observed among the four oaks (*Q. lobata*, *Q. mongolica*, *Q. aquifolioides*, and *Q. gilva*). Owing to the low assembly quality of the *F. sylvatica* genome (Mishra et al., 2018), the collinearity between *F. sylvatica* and *C. mollissima* was poor and lower than that between *C. mollissima* and *Q. lobata* (Fig. 2B). The collinear relationship between the 12 chromosomes of the four oak species showed a 3:3 syntenic depth relationship in homologous regions with *V. vinifera* chromosomes (Fig. S6A–E). In addition, *F. sylvatica* and *C. mollissima* also showed a 3:3 syntenic depth relationship in homologous regions with *V. vinifera* (Fig. S6F–I).

Whole-genome duplication (WGD) events among oak species, *F. sylvatica* and *C. mollissima* were further analyzed by measuring the synonymous substitutions per synonymous site (*K*s) age distributions (Figs. 2C and S6D–I). *K*s values for the genomes of the paralogs of oak species, *Fagus* and *Castanea* species showed a similar peak at *K*s = 1.41 (1.38–1.44). Shared *K*s values indicate a common typical gamma (γ) event corresponding to whole-genome triplication. Our results showed that oak, *Fagus*, and *Castanea* species did not undergo additional WGD events.

3.4. Genomic basis of ecological adaptability in oaks

Oaks are long-lived sessile trees that survive within a wide range of heterogeneous environments and under various abiotic and biotic threats throughout their lifespans (Plomion et al., 2018; Sork et al., 2022). To further investigate the genomic basis underlying ecological adaptability of oaks, we conducted comparative genomics and functional genomics analyses. Five oak species were analyzed, including two evergreen species (*Quercus gilva* and *Q. suber*) and three deciduous species (*Q. mongolica*, *Q. lobata*, and
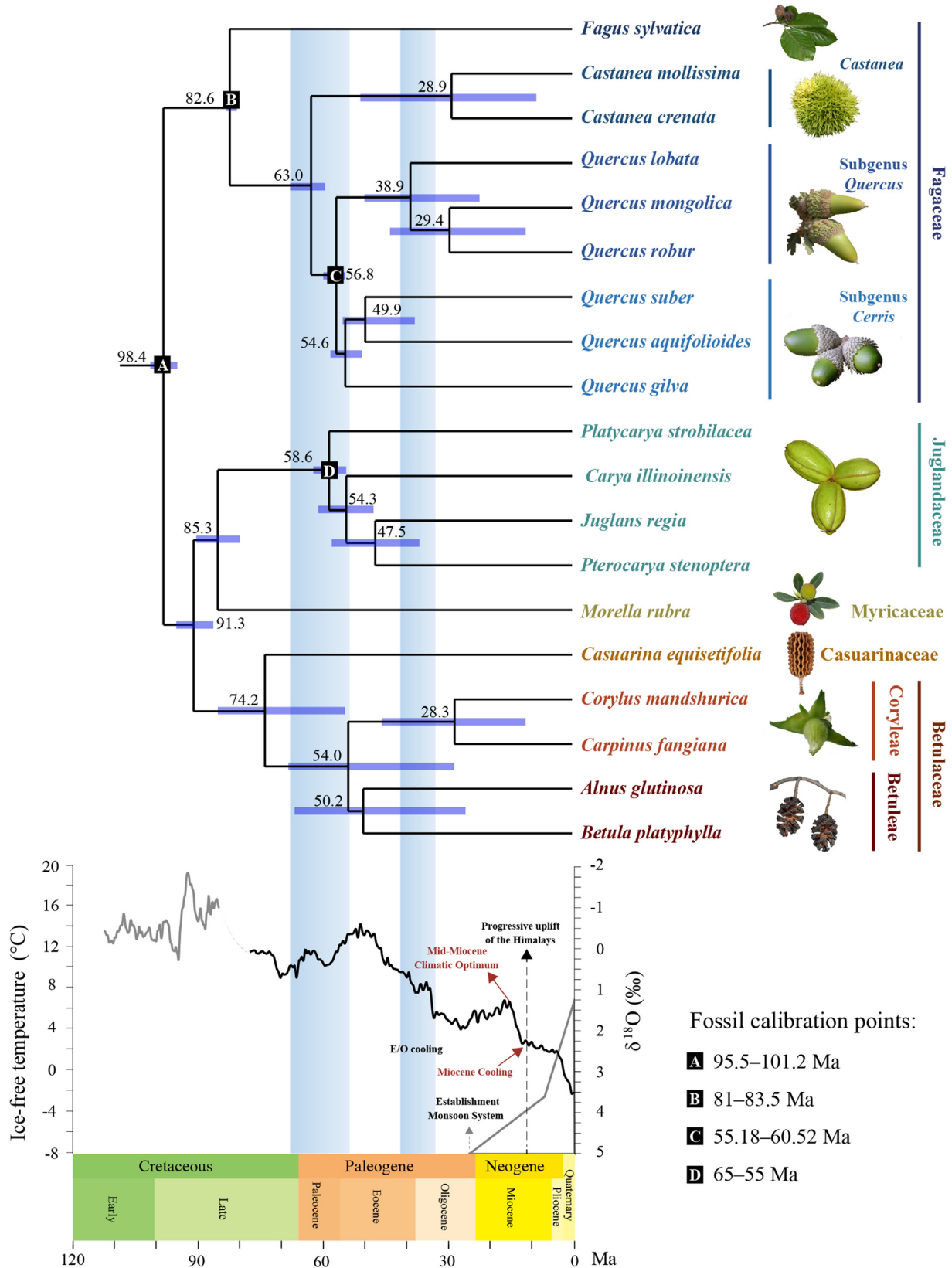
Fig. 1. Maximum likelihood phylogenetic tree of 19 Fagales species based on 649 single-copy genes. Numerical values at each node show estimated time of divergence. Fossil calibration times are marked A, B, C, and D (see Methods). Blue bars across nodes represent 95% highest posterior density (HPD) intervals around the mean divergence time estimates. A stacked deep-sea benthic foraminiferal oxygen-isotope curve shows the major global temperature trends (Cramer et al., 2009).
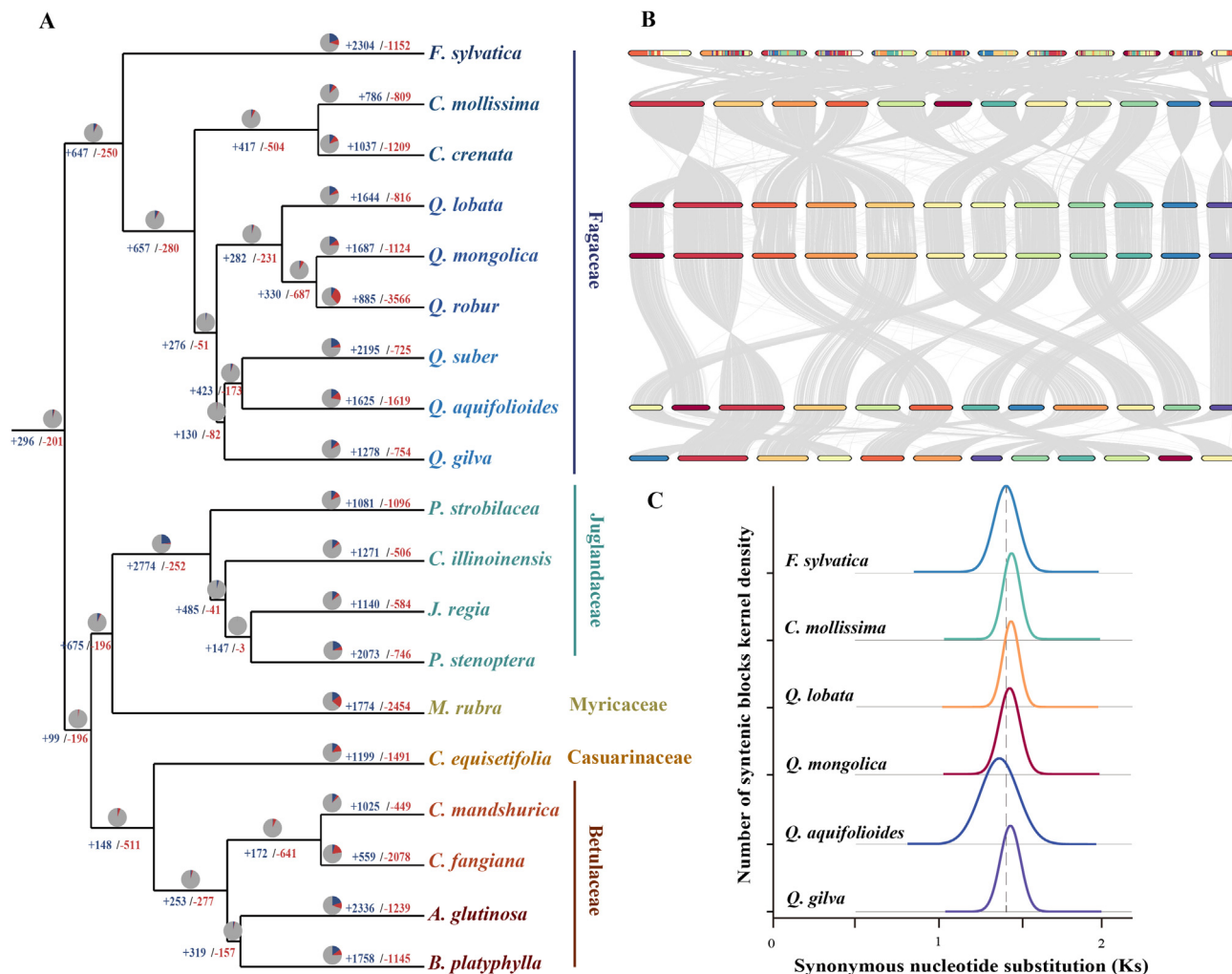
Fig. 2. Evolutionary analyses of *Quercus gilva* genome (A) Expansion and contraction of gene families of *Q. gilva* and 20 other species (B) MCScanX identified synteny blocks (involving ≥ 5 collinear genes) between six Fagaceae species. Species names correspond to those in (A) (C) *Ks* distribution of syntenic blocks for six Fagaceae species.

Q. robur). *Q. aquifolioides* was removed due to the low annotation quality (see also section 2.10).

3.4.1. Leaf cuticle biosynthesis in oaks

The plant cuticle is a hydrophobic barrier that allows plants to survive harsh conditions, such as drought, UV radiation, and water scarcity (Yeats and Rose, 2013). Cutin is the predominant chemical component of the cuticle, comprising covalently crosslinked C16 or C18 hydroxy fatty acids and their derivatives and glycerol (Pollard et al., 2008; Beisson et al., 2012). The cytochrome P450 86 A (*CYP86A*) family is involved in the terminal catalysis of fatty acid hydroxylation. Two to four genes homologous to *CYP86A* are found in the five oak species (Table S2). The *CYP77A* family performs mid-chain hydroxylation of the cutin monomer (Fich et al., 2016). One to two genes homologous to *CYP77A* were identified in the five oak species (Table S2). The final reaction is catalyzed by a fatty acyl moiety to form lysophosphatic acid (LPA) by glycerol-3-phosphate acyltransferase (GPAT) (Singer et al., 2016). Twenty-three to thirty-six genes homologous to *GPAT* were found in the five oak species (Table S2). *GPAT* shares a tandem repeat region in five species, and the tandem repeat region contains 1—7 homologous genes (Table S16).

Wax components typically constitute 20%—60% of the cuticle mass (Heredia, 2003) and protect plants from both biotic and abiotic stresses (Eigenbrode and Pillai, 1998). Cuticular wax is a complex mixture of straight-chain C20 to C60 aliphatics (Jetter et al., 2006). Three main stages of plant cuticular wax biosynthesis are observed: fatty acid elongation, alcohol formation (acyl reduction pathway), and alkane formation (decarbonylation pathway). In the fatty acid elongation stages, the homologous genes of the five species did not expand significantly and were relatively conserved among the species. The copy numbers of the four genes were 1—2 (*KCS6*), 1—3 (*KCR1*), 1 (*PAS2*), and 1—3 (*ECR*) (Figs. 3 and 4A; Table S2). Genes in the alcohol- and alkane-forming stages were significantly expanded. In the alcohol-forming stage, one to seven genes were homologous to *FAR3* in the five oak species. Except for *Q. suber*, the four other oak species shared a tandem repeat region (Fig. 4B and Table S16). *WSD1* shared two tandem repeat regions in the five oaks, with 2—9 homologous genes (Fig. 4; Tables S2 and S16). During the alkane-forming stage, *CER1* shares a tandem repeat region in five species, with 3—12 homologous genes (Fig. 4; Tables S2 and S16). *CER3* has 1—2 homologous genes in the five oaks and cytochrome *b*5 (*CYTB5*) has 3—14 homologous genes in the five oaks (Fig. 4A; Table S2). *MAH1* shared one tandem repeat region with 2—7 homologous genes in the five species (Fig. 4; Tables S2 and S16). In this study, both evergreen and deciduous oaks exhibited expansion in the cuticle biosynthesis pathway, and tandem repeats were present in the synteny blocks of five oaks,

indicating that the cuticular wax and cutin biosynthesis had expanded in the most recent ancestors of oaks. Both evergreen and deciduous oaks were retained, and there were no significant differences in the number of homologous genes related to cuticular wax and cutin biosynthesis between the evergreen and deciduous oaks (Table S17). This finding suggests the accumulation of wax and cutin for the formation of the cuticle layer, as a strong physical barrier, is crucial for both evergreen and deciduous oaks (Eigenbrode and Pillai, 1998).

Cuticle biosynthesis-related genes are highly expressed in the buds and leaves of the five oaks (Figs. 3, S7, and S8). To investigate the gene regulation of cuticle biosynthesis in leaves, we conducted a transcriptomic analysis of bud-to-leaf development in an evergreen oak (Q. gilva) and deciduous oak (Q. robur). Both in Q. gilva and Q. robur, cutin and wax biosynthesis-related genes were highly expressed in the buds and young leaves. However, in Q. gilva, these genes were the most highly expressed in the buds, whereas in Q. robur, they were the most highly expressed in young leaves. Since Q. gilva sprouts in summer, intense sunshine, high temperatures, and drought likely induce the biosynthesis of cuticles in buds to protect tender buds and young leaves. By contrast, Q. robur sprouts in spring, which presents more moderate sunshine and temperature, thus leading to gradual increases in cutin and wax biosynthesis. However, drought did not increase cuticle biosynthesis in mature leaves. As revealed in Q. robur, cuticle biosynthesis-related genes were not differentially expressed between the control and drought groups.

3.4.2. Triterpene biosynthesis in oaks

Plant triterpenes are secondary metabolites that play a crucial role in reducing the damage caused by diseases, insect pests, and other organisms (Gershenzon and Dudareva, 2007; Busta et al., 2020). The first diversification step in triterpene biosynthesis is the catalysis of 2, 3-oxidized squalene into a four- or five-ring structure by oxidosqualene cyclases (OSCs) (Xu et al., 2004; Phillips et al., 2006). Cyclized triterpenes are further oxidized by cytochrome P450-dependent monooxygenases and subsequently glycosylated to form various saponins (Moses et al., 2013) (Fig. 5A). We identified OSC genes in the five oaks using three experimentally verified OSC genes in Q. suber as reference sequences (Busta et al., 2020) (Figs. 5B and S9; Table S3). Compared with Arabidopsis (with 14 OSC genes from the reviewed Swiss-Prot), the OSC gene family species were significantly expanded in the five oaks. Copy numbers were 10−21 (OSC1), 3−8 (OSC2), 3−8 (OSC3), 4−10 (OSC4), and 0−3 (OSC5) (Table S3). All five species shared tandem repeats (Fig. 5C; Table S18). OSC1 had two shared tandem repeats (Fig. 5C). OSC2, OSC3, and OSC4 each contained one tandem repeat (Fig. 5C). Expanded and diverse OSC homologous genes in oaks have laid the foundation for the production of diverse triterpenoids. Significant differences in the number of OSC homologous genes were observed between evergreen and deciduous species ($p < 0.05$), and evergreen species had more gene copies than deciduous species (Table S17). Due to the continuous growth ecological habits of evergreen oaks, evergreen oaks were more affected by environmental stress. More OSC gene copies are able to produce and accumulate diverse triterpenoids as defense compounds against biotic and abiotic stresses (Gershenzon and Dudareva, 2007; Busta et al., 2020; Picard et al., 2023).

Triterpene synthesis genes are expressed in various tissues of oak species to varying degrees, including in the leaves, stems, flowers, and roots (Figs. 5C and S8), which suggests that triterpenes play an important role in whole-plant phytochemical defense. For example, in oak leaves, pentacyclic triterpenoids are the most abundant class and account for approximately 49%−61% of the cuticular wax, which protects plants against herbivory and

stabilizes the cuticle of heat-stressed plants (Bueno et al., 2020). Several OSC genes are highly expressed in Q. gilva and Q. robur during leaf development, such as QgOSC1.6, QrOSC1.5, QgOSC2.2, and QrOSC2.4. OSC1 and OSC2 are involved in the synthesis of lupeol and amyrin and may be the main triterpenoids or precursors of triterpenoids in the leaf cuticle of oaks (Bueno et al., 2020; Simões et al., 2020) (Figs. 5C and S8). However, the temporal expression trend of these genes differed. For example, during bud-to-leaf development, OSC1 expression in Q. gilva initially increased, reached the highest level in young leaves, and then decreased, which was similar to that of QgOSC1.6 (Figs. 5C and S8); OSC2 expression in Q. gilva was highest in the buds and then decreased, which was similar to that of QgOSC2.2 (Figs. 5C and S8); and OSC3 expression in Q. robur was similar to that of QrQSC3P3 - QrQSC3P6 (Figs. 5C and S8). These findings indicate that the synthesis of different triterpenes exhibits temporal heterogeneity during leaf development.

To investigate whether the key genes involved in the cuticle and triterpenoids biosynthesis pathways in five oak species underwent selective pressures, we conducted positive selection analysis. Our results revealed that these genes are predominantly subject to purifying selection within oak species (Table S25), indicating that these genes are relatively conserved in their function (Sigalova et al., 2019). Cuticle, wax, and terpene compounds play crucial roles in plant responses to biotic and abiotic stresses such as drought, radiation, and insect pests (Yeats and Rose, 2013; Busta et al., 2020). Therefore, these genes have undergone strong purifying selection, maintained their functional stability and possibly further shaped the ecological adaptability of oak species.

3.5. Positively selected genes between deciduous and evergreen oaks

The ecological habits of oak species differ greatly. Two well-known habits of oaks include evergreen and deciduous. We identified genetic differences in growth status and life habits (e.g., dormancy) between evergreen and deciduous oaks by identifying positively selected genes in deciduous oak species (e.g., Quercus lobata, Q. mongolica, and Q. robur) and evergreen oak species (e.g., Q. gilva, Q. suber, and Q. aquifolioides). A total of 1526 genes were positively selected in deciduous oak species and 1277 in evergreen oak species (Tables S19 and S20). The positively selected genes of deciduous oaks were functionally enriched in the response to abscisic acid (ABA), auxin-activated signaling pathway, starch biosynthesis process, and circadian rhythm−plant pathways (Fig. 6A and B; Tables S21 and S22). In contrast, positively selected genes in evergreen oaks were functionally enriched for glycolysis/ gluconeogenesis, carbon fixation in photosynthetic organisms, TCA cycle, and circadian rhythm−plant pathways (Fig. 6A−C; Tables S23 and S24). These positively selected genes may have contributed to the environmental adaptability of the oak species.

The circadian rhythm-plant pathway, which is responsible for different life habits as adaptations to seasonal changes, was enriched in both evergreen and deciduous oaks. Seven circadian clock genes were positively selected in evergreen and deciduous oaks. Three genes, LATE ELONGATED HYPOCOTYL (LHY), PSEUDO-RESPONSE REGULATORS 7 (PRR7), and SUPPRESSOR OF PHYA-105 (SPA1), were positively selected in both deciduous and evergreen oaks. The CCA1 HIKING EXPEDITION (CHE) and ELONGATED HY-POCOTYL 5 (HY5) genes were positively selected in deciduous oaks. TIMING OF CAB EXPRESSION1 (TOC1) and ZEITLUPE (ZTL) were positively selected in evergreen oaks. These circadian clock genes are likely involved in sensing and regulating the seasonal responses of oaks (Fig. 6D). For example, photoreceptors (PHYTOCHROME: PHY, CRYPTOCHROME: CRY, and ZTL) sense changes in photoperiod
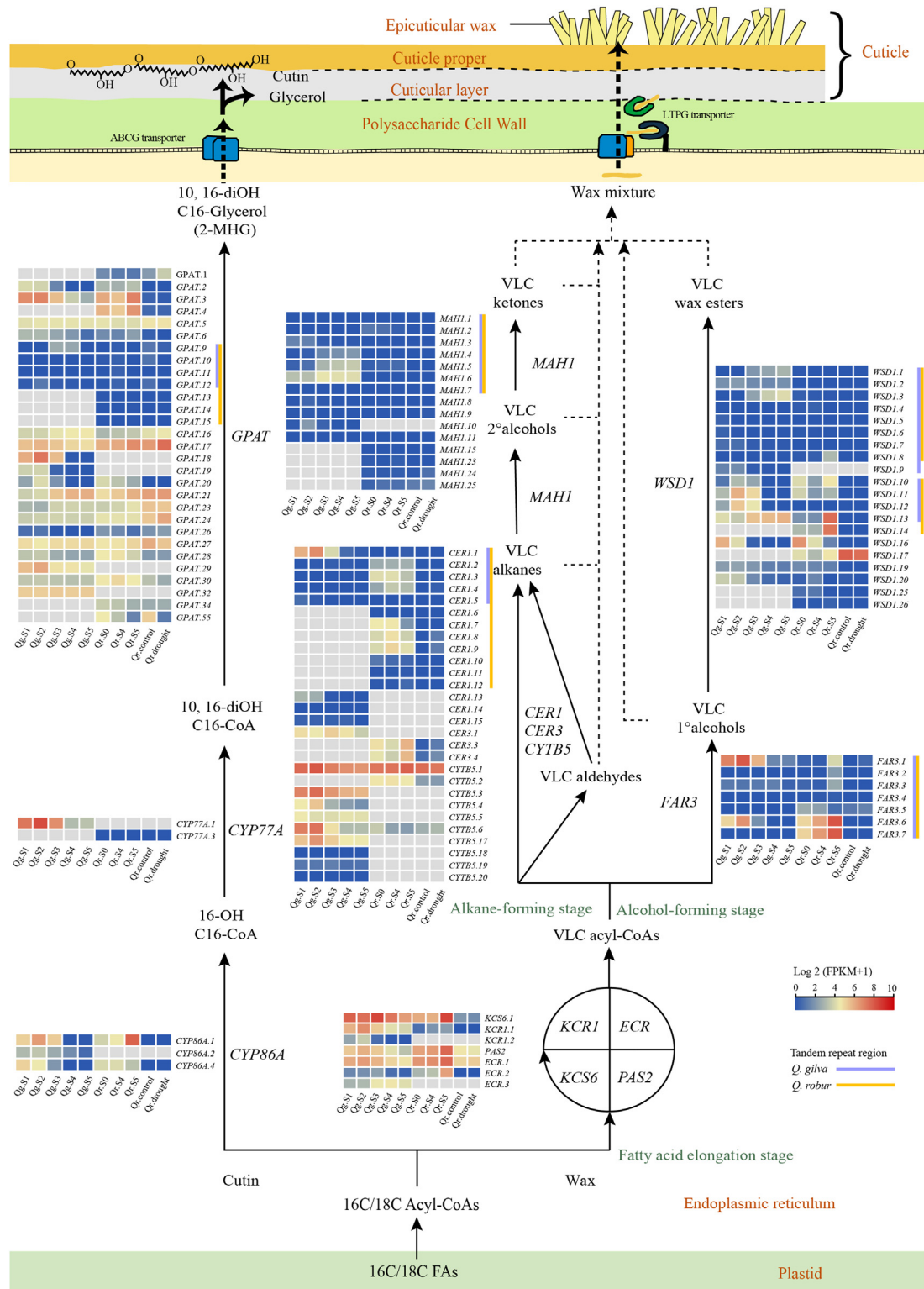
Fig. 3. Cuticular wax and cutin biosynthesis pathways and heat maps of genes related to *Quercus gilva* and *Q. robur*. Arrangement of genes between species were in a syntenic correspondence. Right vertical line of the gene name refers to the tandem repeat area of the gene (orange = *Q. robur*, purple = *Q. gilva*). Leaf development stages corresponding to transcriptome samples: Qg. S1: buds appear; Qg. S2: leaf sprouts; Qg. S3: young leaf unfolds; Qg. S4: leaf develops; Qg. S5: mature leaf; Qr. S0: buds appear; Qr. S4: leaf and stem development; Qr. S5: mature leaf; Qr. control: Control group of leaf drought stress; Qr. drought: Leaves were subjected to drought stress; Qg, *Q. gilva*; Qr, *Q. robur*. CYP86A, Cytochrome P450 86A; *CYP77A, Cytochrome P450 77A*; *GPAT, Glycerol-3-phosphate acyltransferase*; *KCS6, 3-ketoacyl-CoA synthase 6*; *KCR1, Very-long-chain 3-oxoacyl-coa reductase 1*; *PAS2, Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase PASTICCINO 2A*; *ECR, Very-long-chain enoyl-CoA reductase*; *FAR3, Fatty acyl-CoA reductase 3*; *WSD1, Wax ester synthase/diacylglycerol acyltransferase 1*; *CER1, ECERIFERUM 1*; *CER3, ECERIFERUM 3*; *CYTB5, Cytochrome b5*; *MAH1, Midchain alkane hydroxylase 1*; FAs, Fatty acids; VLC, Very long chain.
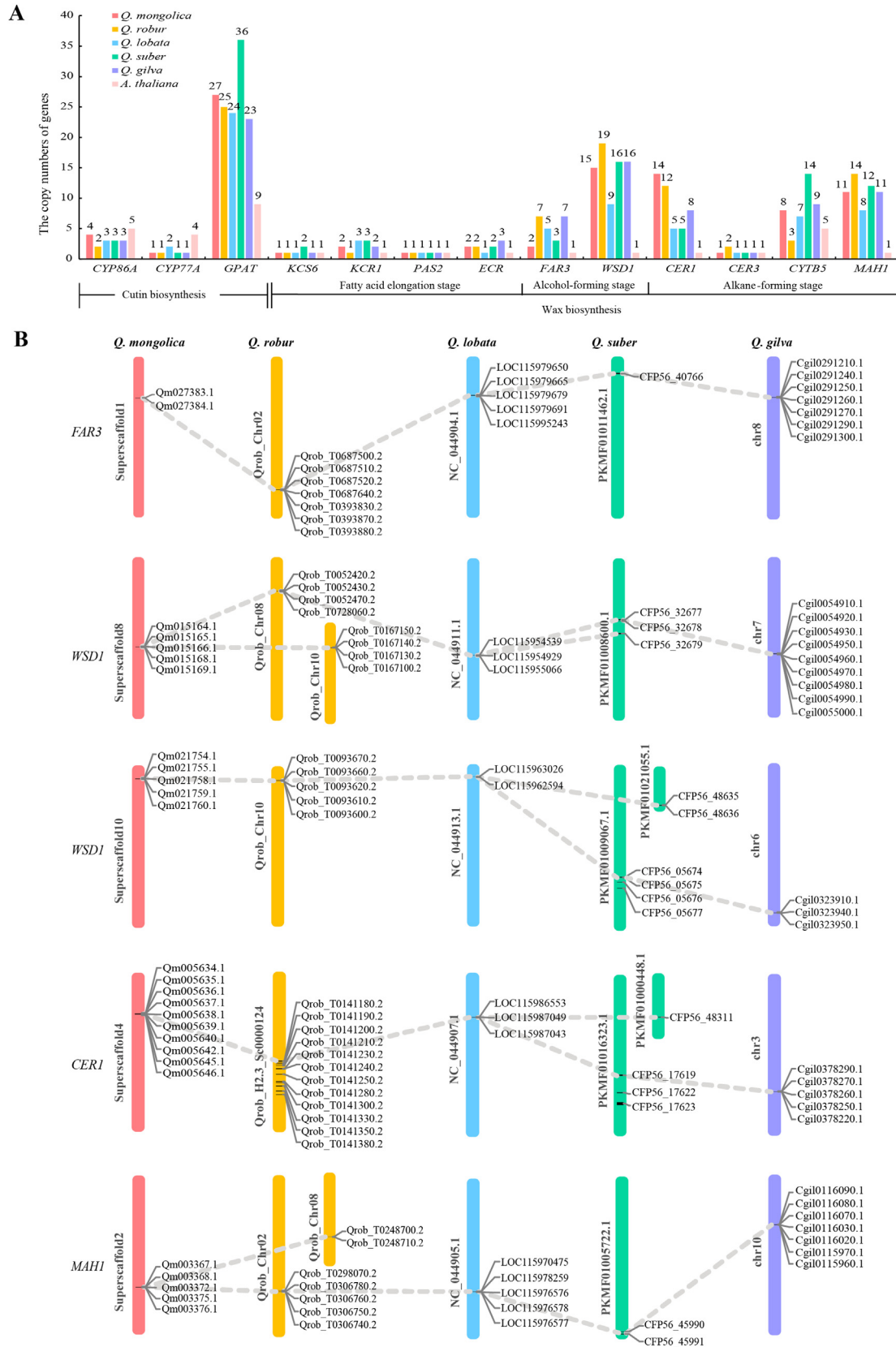
Fig. 4. Identification and comparison of key genes in the pathway of cutin and cuticular wax biosynthesis in five *Quercus* species and chromosomal localization of tandem repeat regions. (A) Identification and comparison of key genes involved in cutin and cuticular wax biosynthesis pathway in five oaks. (B) Chromosome localization of tandem repeat regions of genes *FAR3*, *WSD1*, *CER1* and *MAH1* in five oaks. Grey dashed lines represent one-to-one correspondence between syntenic regions of five oaks.
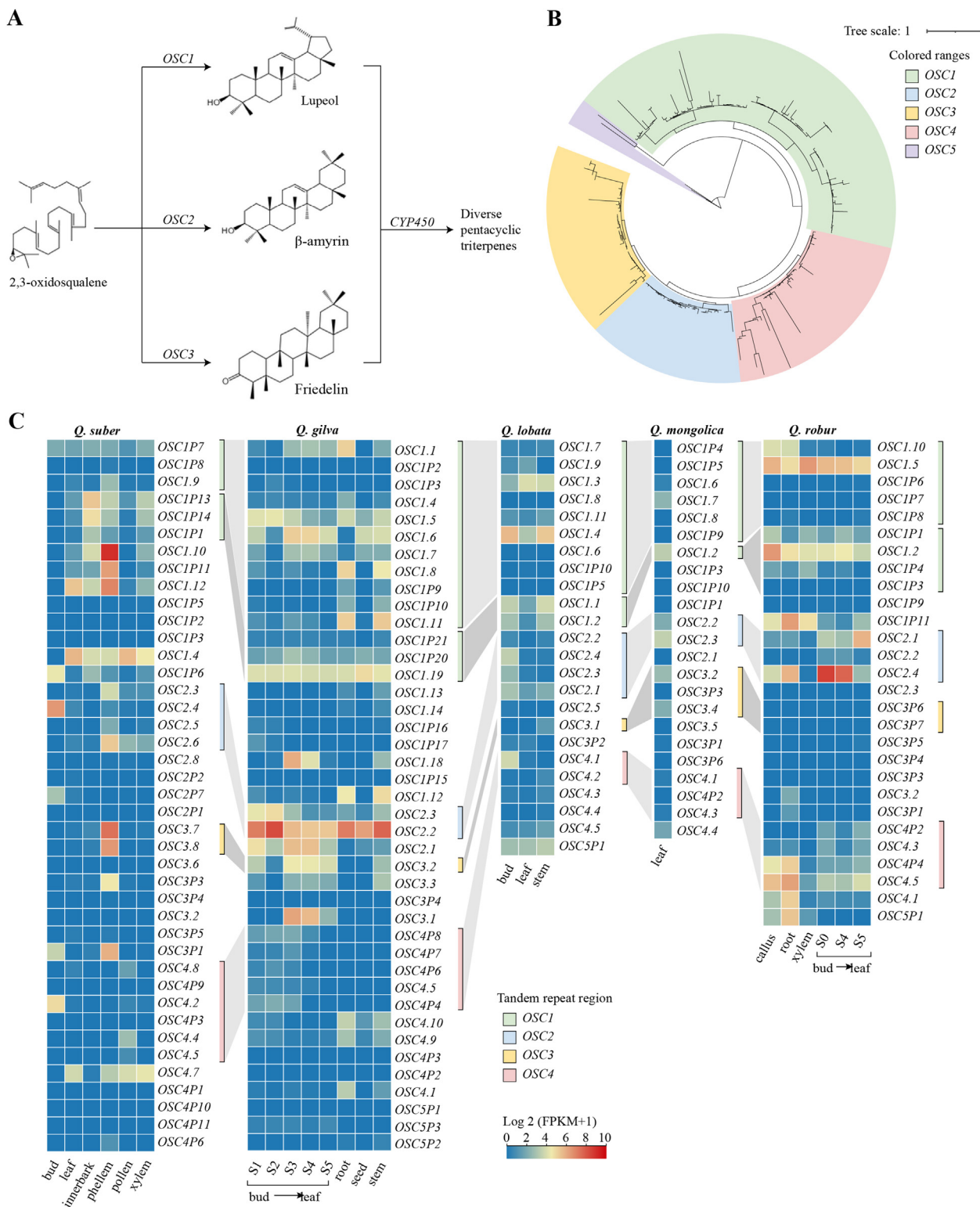
Fig. 5. *OSC* (*Oxidosqualene cyclases*) pathway of pentacyclic triterpenes, phylogenetic tree and heat map of *OSC1—5* gene expression in different tissues of the five oaks. (A) Pathway of pentacyclic triterpenes involving *OSC1*, *OSC2*, and *OSC3*. *OSC1*, *Oxidosqualene cyclases 1*; *OSC2*, *Oxidosqualene cyclases 2*; *OSC3*, *Oxidosqualene cyclases 3*. (B) *OSC* phylogenetic tree of the five oaks. *OSC4* is the *OSC* gene with unknown function, and *OSC5* is the gene with unknown function at the base of the phylogenetic tree. (C) Heat map of *OSC1—5* gene expression in different tissues of the five oaks during the synthesis of pentacyclic triterpene (the same color brackets on the right side of the heat map indicate the same shared tandem repeat regions, grey color represent one-to-one correspondence between syntenic blocks of five oaks). *OSC1P*, *Oxidosqualene cyclases 1 pseudogene*; *OSC2P*, *Oxidosqualene cyclases 2 pseudogene*; *OSC3P*, *Oxidosqualene cyclases 3 pseudogene*. The pseudogenes were identified, which filtered out hits that presented a longest open reading frame less than 1500bp or had gaps in the SDCTAE motif.
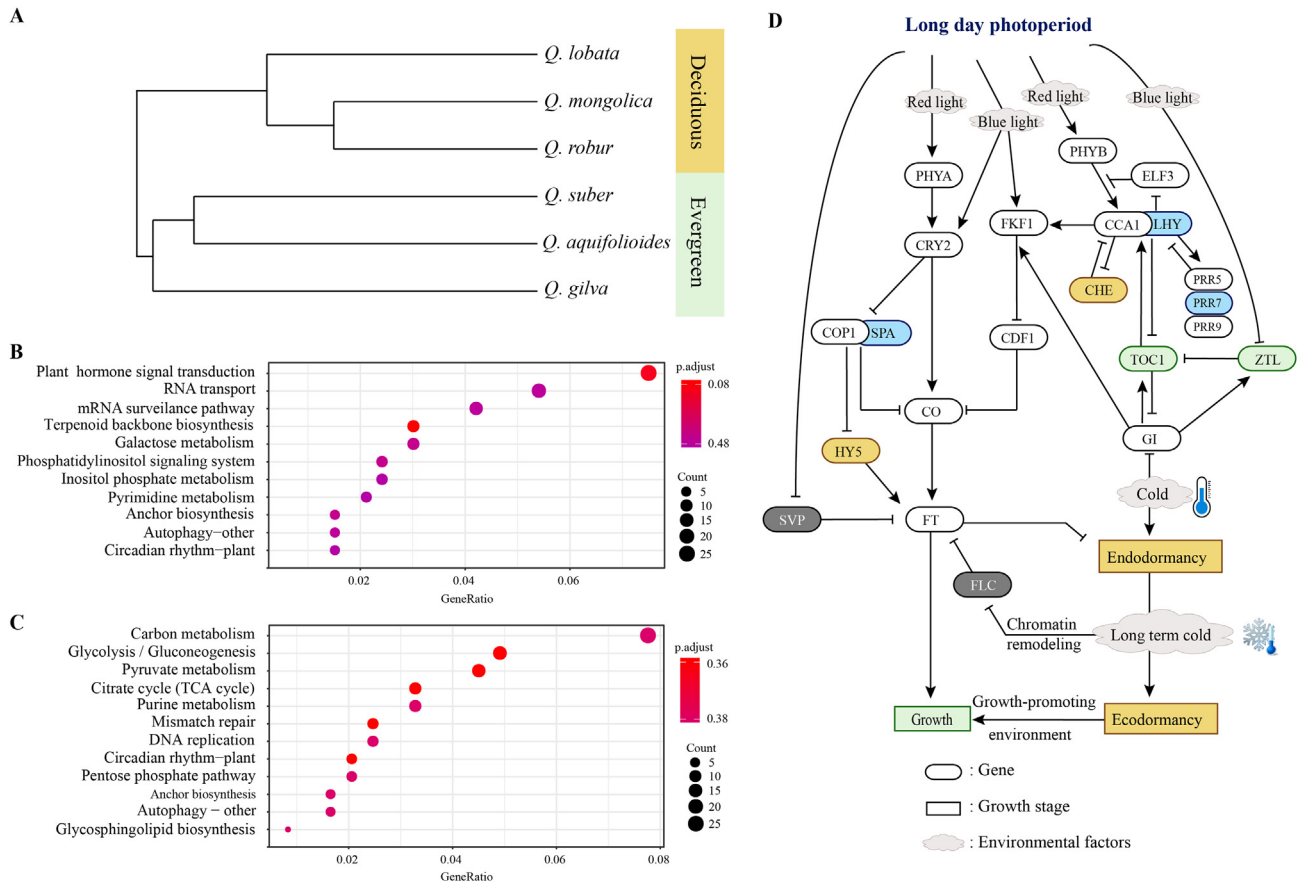
Fig. 6. Positively selected gene analysis of *Quercus*. (A) Phylogenetic tree of deciduous-related species (*Q. lobata*, *Q. mongolica* and *Q. robur*) and evergreen species (*Q. suber*, *Q. aquifolioides* and *Q. gilva*). The Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathway enrichment analysis of positively selected genes in (B) deciduous *Quercus* with evergreen *Quercus* as background branches and (C) evergreen *Quercus* with deciduous *Quercus* as background branches. (D) Model representing a simplified overview of the multiple pathways between environmental input signals and dormancy regulation and growth based on information gathered from annuals and perennials. The gray ovals are non-circadian clock genes, the rest are circadian clock genes. The blue ovals indicate the common positively selected genes of the deciduous and evergreen groups, the yellow ovals indicate the positively selected genes of the deciduous groups, the green ovals are the positively selected genes of the evergreen groups, and the white ovals are genes that have not been positively selected. Lines with arrows and lines with T-ends indicate activation and inhibition, respectively.

and temperature, and these signals are then transmitted to the core circadian oscillator composed of CIRCADIAN CLOCK ASSOCIATED 1(CCA1)/LHY, PRRs, and TOC1 to generate circadian oscillations (Guo et al., 1998; Wang and Tobin, 1998). Subsequently, the signals are transmitted to output pathways, such as the flowering positive regulators CONSTANS(CO)/FLOWERING LOCUS T (FT) and HY5, which regulate flower growth, and some dormancy-related SHORT VEGETATIVE PHASE (SVP) genes, which control dormancy (Hartmann et al., 2000). Seasonal sensing and regulation play important roles in the growth and survival of woody plants (Bhalerao and Fischer, 2017). Different circadian clock genes have been selected in evergreen and deciduous oaks, and they likely correspond to the different seasonal responses in the trees, suggesting that distinct regulatory mechanisms are adopted to adapt to various ecological niches. Our results provide a basis for further research on the mechanisms by which evergreen and deciduous oaks respond to seasonal changes.

4. Discussion

4.1. Comparison of genome evolution in Fagaceae

Fagaceae is a significant keystone family whose members are economically important because of their valuable timber and edible fruits (Wang et al., 2009). Fagaceae diverged during the Middle Cretaceous but mainly diversified after the Cretaceous-Paleogene (K-Pg) boundary (Xiang et al., 2014). The K-Pg boundary extinction provided abundant new niches, and dramatic climate changes, such as a warm climate during the Paleogene, which may also have contributed to the diversification of Fagaceae (Zachos et al., 2001; Cramer et al., 2009; Xiang et al., 2014). As a result, Fagaceae comprises more than 900 species (70% of all extant Fagales species) and is spread throughout the Northern Hemisphere, making it the most evolutionarily successful family of plants (Kremer et al., 2012).

Interestingly, we found that species within Fagaceae (including oaks, *Fagus*, and *Castanea*) have not undergone additional WGD or large-scale structural variation events (e.g., chromosomal rearrangements). However, previous studies have demonstrated that the Juglandaceae family has undergone a recent WGD event in addition to the shared ancient γ event (Zhang et al., 2020; Ding et al., 2023). There has been no additional WGD event in the genome of Betulaceae family, while higher chromosome structural variations caused by chromosomal rearrangements have been discovered (Li et al., 2021). Thus, the genome of Fagaceae may have undergone a different sort of evolution, resulting in a more conserved and stable genome structure. For example, local gene duplication has been found in the oak genome (Plomion et al., 2018; Ai et al., 2022). Considering the wide distribution and numerous

species of the Fagaceae family, future research should encompass a broader range of Fagaceae species to delve deeper into the evolutionary mechanisms underlying the Fagaceae genome structure.

4.2. Genomic footprint for the wide ecological adaptations of oaks

Most oaks are long-lived organisms that adapt to diverse habitats and face a wide range of abiotic and biotic threats throughout their lifespans (Plomion et al., 2018; Sork et al., 2022). Thus, plant defenses provide an important guarantee for oaks to respond to biological and abiotic stresses. Plants exert their defense mechanisms mainly through physical and chemical defense features and gene regulation (Agrawal and Fishbein, 2006; Neilson et al., 2013). The expansion and tandem repetition of genes involved in physical and chemical defense and gene regulation are likely the genomic footprints for the wide ecological adaptation of oaks. Oak species have developed bark and leaf cuticles against biological stresses, such as pathogens and herbivores, and abiotic stresses, such as drought and ultraviolet radiation (Bueno et al., 2020; Vasilchenko et al., 2022), and they represent strong physical defenses. Cuticle biosynthesis genes involved in bark synthesis (Fernández-Piñán et al., 2021) are significantly expanded in oaks, with five homologous genes (*GPAT*, *FAR3*, *WSD1*, *CER1*, and *MAH1*) sharing tandem repeat regions, and they are highly expressed in buds and young leaves. Oaks are rich in diverse secondary metabolites, such as terpenoids, flavonoids, tannins, and salicylic acid, which represent a significant chemical defense mechanism for the growth and development of plants and coping with stressful environments (Picard et al., 2023). Genes involved in secondary metabolism, such as the biosynthesis and modification of terpenoids, flavonoids (e.g., quercetin), and salicylic acid, were significantly shared and enriched among oak species. For example, *OSC* genes related to pentacyclic triterpene synthesis were significantly expanded in oaks, with four homologous genes (*OSC1–4*) sharing tandem repeat regions, and were highly expressed in various tissues. Third, plant disease resistance-related genes (*R* genes) are key components of gene interactions between plants and pathogens (Gururani et al., 2012). Previous studies have reported that *R* genes detected in oaks and their local tandem repetition play key roles in the disease resistance pathway of oaks (Plomion et al., 2018; Ai et al., 2022; Sork et al., 2022). It is worth noting that our study only encompasses species from four out of eight sections of *Quercus*. Hence, it remains imperative to include species from the remaining four sections in future research to provide a more comprehensive understanding of the general evolutionary patterns of oaks.

4.3. Differences in ecological adaptation between evergreen and deciduous oaks

Deciduous oak species mainly grow in spring and summer and resist adverse environmental conditions through natural and ecological dormancy in autumn and winter, whereas evergreen oaks grow throughout their lifespan (Escudero et al., 2017). In addition to circadian rhythm-related genes, other positively selected genes also explain the growth habits of evergreen and deciduous oaks. In deciduous oaks, the response to ABA, auxin-activated signaling pathway, and starch biosynthesis process are under positive selection and may be related to the phenology of deciduous leaves and winter dormancy. ABA and auxin are important hormones involved in defoliation and dormancy (Ueno et al., 2013). Reduced sunlight in autumn increases ABA synthesis and stimulates ethylene production and callose deposition in the plasmodesmata, which not only results in seasonal defoliation but also impedes intercellular communication in the shoot apical meristem, thus leading to dormancy (Rinne et al., 2011). Several studies have shown that auxins promote the decomposition of callose during phloem dormancy and accelerate the release of dormancy. Auxin signaling also plays an important role in regulating bud dormancy (Hao et al., 2019). Starch pathway selection in deciduous oak species may also be related to its dormancy habit. Owing to the short growing season of deciduous oaks, most perennials accumulate large amounts of starch stored in tissue cells before winter, which serves as the nutrient basis for new germination and growth from winter to the following year (Richardson et al., 2010).

In evergreen oaks, the TCA cycle, glycolysis/gluconeogenesis, and carbon fixation in photosynthetic organs are significantly enriched, which may be related to their persistent growth habits. Plants maintain sustainable growth by producing glucose and breaking it down into energy materials through a series of enzymatic and biochemical reactions (Zhang and Fernie, 2018). The gene family associated with carbon fixation is enriched in evergreen oaks, which produce glucose through cyclic reactions (Braakman and Smith, 2012). Glycolysis and TCA, the main pathways of carbohydrate catabolism, produce a large amount of energy for plant growth (Plaxton, 1996). Evergreen oaks produce more glucose through carbon fixation; furthermore, this glucose can generate more energy through glycolysis and TCA catabolism to maintain the growth and development of evergreen oaks. Compared with the dormant deciduous oaks, evergreen oaks continue to grow throughout their lives and, thus, consume more energy and substances to survive. The associated pathways identified in evergreen oaks may provide sufficient energy and metabolic substances for sustained growth.

Data availability

All sequencing data used in this study have been deposited in the National Genomics Data Center (https://ngdc.cncb.ac.cn) with the accession number GWHDOCW00000000 for the reference genome and gene annotations, and PRJCA018224 for the RNA-seq data of *Quercus gilva*.

CRediT authorship contribution statement

Tian-Rui Wang: Writing – original draft, Software, Methodology, Formal analysis, Data curation. **Xin Ning:** Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Si-Si Zheng:** Writing – original draft, Visualization, Validation, Funding acquisition, Formal analysis. **Yu Li:** Visualization, Validation, Formal analysis. **Zi-Jia Lu:** Formal analysis. **Hong-Hu Meng:** Writing – review & editing, Supervision, Conceptualization. **Bin-Jie Ge:** Writing – review & editing, Visualization, Investigation. **Gregor Kozlowski:** Writing – review & editing, Supervision. **Meng-Xiao Yan:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Yi-Gang Song:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pld.2024.07.008.

References

Agrawal, A.A., Fishbein, M., 2006. Plant defense syndromes. Ecology 87, S132—S149.

Ai, W., Liu, Y., Mei, M., et al., 2022. A chromosome-scale genome assembly of the Mongolian oak (*Quercus mongolica*). Mol. Ecol. Resour. 22, 2396—2410.

Aitken, S.N., Yeaman, S., Holliday, J.A., et al., 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol. Appl. 1, 95—111.

Ashburner, M., Ball, C.A., Blake, J.A., et al., 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25—29.

Beisson, F., Li-Beisson, Y., Pollard, M., 2012. Solving the puzzles of cutin and suberin polymer biosynthesis. Curr. Opin. Plant Biol. 15, 329—337.

Bhalerao, R.P., Fischer, U., 2017. Environmental and hormonal control of cambial stem cell dynamics. J. Exp. Bot. 68, 79—87.

Braakman, R., Smith, E., 2012. The emergence and early evolution of biological carbon-fixation. PLoS Comput. Biol. 8, e1002455.

Bueno, A., Sancho-Knapik, D., Gil-Pelegrín, E., et al., 2020. Cuticular wax coverage and its transpiration barrier properties in *Quercus coccifera* L. leaves: does the environment matter? Tree Physiol. 40, 827—840.

Burge, C., Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78—94.

Busta, L., Serra, O., Kim, O.T., et al., 2020. Oxidosqualene cyclases involved in the biosynthesis of triterpenoids in *Quercus suber* cork. Sci. Rep. 10, 8011.

Camacho, C., Coulouris, G., Avagyan, V., et al., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Capblancq, T., Fitzpatrick, M.C., Bay, R.A., et al., 2020. Genomic prediction of (mal) adaptation across current and future climatic landscapes. Annu. Rev. Ecol. Evol. Syst. 51, 245—269.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972—1973.

Chen, C., Chen, H., Zhang, Y., et al., 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant 13, 1194—1202.

Chen, S., Krinsky, B.H., Long, M., 2013. New genes as drivers of phenotypic evolution. Nat. Rev. Genet. 14, 645—660.

Chen, S., Zhou, Y., Chen, Y., et al., 2018b. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884—i890.

Chen, Y., Chen, Y., Shi, C., et al., 2018a. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience 7, gix120.

Cramer, B.S., Toggweiler, J.R., Wright, J.D., et al., 2009. Ocean overturning since the Late Cretaceous: inferences from a new benthic foraminiferal isotope compilation. Paleoceanogr. Paleoclimatol. 24, PA4216.

Dauphin, B., Rellstab, C., Schmid, M., et al., 2021. Genomic vulnerability to rapid climate warming in a tree species with a long generation time. Global Change Biol. 27, 1181—1195.

de Lafontaine, G., Napier, J.D., Petit, R.J., et al., 2018. Invoking adaptation to decipher the genetic legacy of past climate change. Ecology 99, 1530—1546.

Díaz, S., Malhi, Y., 2022. Biodiversity: concepts, patterns, trends, and perspectives. Annu. Rev. Environ. Resour. 47, 31—63.

Ding, Y., Pang, X., Cao, Y., et al., 2023. Genome structure-based Juglandaceae phylogenies contradict alignment-based phylogenies and substitution rates vary with DNA repair genes. Nat. Commun. 14, 617.

Doyle, J.J., Doyle, J.L., 1986. A rapid dna isolation procedure from small quantities of fresh leaf tissues. Phytochem. Bull. 19, 11—15.

Du, F.K., Wang, T.R., Wang, Y.Y., et al., 2020. Contrasted patterns of local adaptation to climate change across the range of an evergreen oak *Quercus aquifolioides*. Evol. Appl. 13, 2377—2391.

Durand, N.C., Shamim, M.S., Machol, I., et al., 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95—98.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792—1797.

Eigenbrode, S.D., Pillai, S.K., 1998. Neonate plutella xylostella responses to surface wax components of a resistant cabbage (*Brassica oleracea*). J. Chem. Ecol. 24, 1611—1627.

Ellegren, H., 2014. Genome sequencing and population genomics in non-model organisms. Trends Ecol. Evol. 29, 51—63.

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238.

Escudero, A., Mediavilla, S., Olmo, M., et al., 2017. Coexistence of deciduous and evergreen oak species in mediterranean environments: costs associated with the leaf and root traits of both habits. In: Gil-Pelegrín, E., Peguero-Pina, J.J., Sancho-Knapik, D. (Eds.), Oaks Physiological Ecology. Exploring the Functional Diversity of Genus *Quercus* L. Springer, pp. 195—237.

Fernández-Piñán, S., Boher, P., Soler, M., et al., 2021. Transcriptomic analysis of cork during seasonal growth highlights regulatory and developmental processes from phellogen to phellem formation. Sci. Rep. 11, 12053.

Fich, E.A., Segerson, N.A., Rose, J.K.C., 2016. The plant polyester cutin: biosynthesis, structure, and biological roles. Annu. Rev. Plant Biol. 67, 207—233.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29—W37.

Frazee, A.C., Pertea, G., Jaffe, A.E., et al., 2015. Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat. Biotechnol. 33, 243—246.

Frith, M.C., Hamada, M., Horton, P., 2010. Parameters for accurate genome alignment. BMC Bioinformatics 11, 80.

Fu, R.R., Zhu, Y.X., Liu, Y., et al., 2022. Genome-wide analyses of introgression between two sympatric Asian oak species. Nat. Ecol. Evol. 6, 924—935.

Gershenzon, J., Dudareva, N., 2007. The function of terpene natural products in the natural world. Nat. Chem. Biol. 3, 408—414.

Gougherty, A.V., Keller, S.R., Fitzpatrick, M.C., 2021. Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. Nat. Clim. Change 11, 166—171.

Gugger, P.F., Fitz-Gibbon, S.T., Albarrán-Lara, A., et al., 2021. Landscape genomics of *Quercus lobata* reveals genes involved in local climate adaptation at multiple spatial scales. Mol. Ecol. 30, 406—423.

Guo, H., Yang, H., Mockler, T.C., et al., 1998. Regulation of flowering time by *Arabidopsis* photoreceptors. Science 279, 1360—1363.

Gururani, M.A., Venkatesh, J., Upadhyaya, C.P., et al., 2012. Plant disease resistance genes: current status and future directions. Physiol. Mol. P. 78, 51—65.

Haas, B.J., Papanicolaou, A., Yassour, M., et al., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494—1512.

Han, M.V., Thomas, G.W.C., Lugo-Martinez, J., et al., 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol. Biol. Evol. 30, 1987—1997.

Hao, X., Tang, H., Wang, B., et al., 2019. Gene characterization and expression analysis reveal the importance of auxin signaling in bud dormancy regulation in tea plant. J. Plant Growth Regul. 38, 225—240.

Hartmann, U., Höhmann, S., Nettesheim, K., et al., 2000. Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. Plant J. 21, 351—360.

He, Z., Feng, X., Chen, Q., et al., 2022. Evolution of coastal forests based on a full set of mangrove genomes. Nat. Ecol. Evol. 6, 738—749.

Heredia, A., 2003. Biophysical and biochemical characteristics of cutin, a plant barrier biopolymer. Biochim. Biophys. Acta Mol. Cell Res. 1620, 1—7.

Hipp, A.L., Manos, P.S., Hahn, M., et al., 2020. Genomic landscape of the global oak phylogeny. New Phytol. 226, 1198—1212.

Jetter, R., Kunst, L., Samuels, A.L., 2006. Composition of Plant Cuticular Waxes. Annual Plant Reviews, 23. Biology of the Plant Cuticle, pp. 145—181.

Jiang, X.L., Hipp, A.L., Deng, M., et al., 2019. East Asian origins of European holly oaks (*Quercus* section *Ilex* Loudon) via the Tibet-Himalaya. J. Biogeogr. 46, 2203—2214.

Jin, D.M., Yuan, Q., Dai, X.L., et al., 2023. Enhanced precipitation has driven the evolution of subtropical evergreen broad-leaved forests in eastern China since the early Miocene: evidence from ring-cupped oaks. J. Syst. Evol. 62, 677—686.

Jones, P., Binns, D., Chang, H.Y., et al., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236—1240.

Kanehisa, M., Goto, S., Kawashima, S., et al., 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. 32, D277—D280.

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

Kiełbasa, S.M., Wan, R., Sato, K., Horton, P., et al., 2011. Adaptive seeds tame genomic sequence comparison. Genome Res. 21, 487—493.

Kim, D., Paggi, J.M., Park, C., et al., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907—915.

Kozlowski, G., Betrisey, S., Song, Y.G., et al., 2018. Wingnuts (Pterocarya) & walnut family: relict trees: linking the past, present and future. Fribourg : Natural History Museum.

Kremer, A., Abbott, A.G., Carlson, J.E., et al., 2012. Genomics of Fagaceae. Tree Genet. Genomes 8, 583—610.

Kremer, A., Hipp, A.L., 2020. Oaks: an evolutionary success story. New Phytol. 226, 987—1011.

Kumar, S., Stecher, G., Suleski, M., et al., 2017. TimeTree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. 34, 1812—1819.

Larson-Johnson, K., 2016. Phylogenetic investigation of the complex evolutionary history of dispersal mode and diversification rates across living and fossil Fagales. New Phytol. 209, 418—435.

Leroy, T., Louvet, J.M., Lalanne, C., et al., 2020. Adaptive introgression as a driver of local adaptation to climate in European white oaks. New Phytol. 226, 1171—1182.

Li, Y., Sun, P., Lu, Z., et al., 2021. The *Corylus mandshurica* genome provides insights into the evolution of Betulaceae genomes and hazelnut breeding. Hortic. Res. 8, 54.

Liu, B., Shi, Y., Yuan, J., et al., 2013. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects (arXiv: Genomics).

Majoros, W.H., Pertea, M., Salzberg, S.L., 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878—2879.

Martins, K., Gugger, P.F., Llanderal-Mendoza, J., et al., 2018. Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. Evol. Appl. 11, 1842–1858.

Meng, H.H., Zhou, S.S., Jiang, X.L., et al., 2019. Are mountaintops climate refugia for plants under global warming? A lesson from high-mountain oaks in tropical rainforest. Alpine Bot. 129, 175–183.

Mishra, B., Gupta, D.K., Pfenninger, M., et al., 2018. A reference genome of the European beech (*Fagus sylvatica* L.). GigaScience 7, giy063.

Moses, T., Pollier, J., Thevelein, J.M., et al., 2013. Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology *in vivo* and *in vitro*. New Phytol. 200, 27–43.

Neilson, E.H., Goodger, J.Q.D., Woodrow, I.E., et al., 2013. Plant chemical defense: at what cost? Trends Plant Sci. 18, 250–258.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., et al., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Pertea, M., Pertea, G.M., Antonescu, C.M., et al., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290–295.

Phillips, D.R., Rasbery, J.M., Bartel, B., et al., 2006. Biosynthetic diversity in plant triterpene cyclization. Curr. Opin. Plant Biol. 9, 305–314.

Picard, M., Oulieu, C., Nonier, M.F., et al., 2023. The role of oak wood in the mint and floral notes of whisky: identification of common terpenoids by aromatic fractionation. J. Inst. Brew. 129, 62–79.

Plaxton, W.C., 1996. The organization and regulation of plant glycolysis. Annu. Rev. Plant Biol. 47, 185–214.

Plomion, C., Aury, J.-M., Amselem, J., et al., 2018. Oak genome reveals facets of long lifespan. Nat. Plants 4, 440–452.

Pollard, M., Beisson, F., Li, Y., et al., 2008. Building lipid barriers: biosynthesis of cutin and suberin. Trends Plant Sci. 13, 236–246.

Puttick, M.N., 2019. MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics 35, 5321–5322.

Rellstab, C., Zoller, S., Walthert, L., et al., 2016. Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. Mol. Ecol. 25, 5907–5924.

Richardson, A.C., Walton, E.F., Meekings, J.S., et al., 2010. Carbohydrate changes in kiwifruit buds during the onset and release from dormancy. Sci. Hortic. 124, 463–468.

Rinne, P.L.H., Welling, A., Vahala, J., et al., 2011. Chilling of dormant buds hyper-induces FLOWERING LOCUS T and recruits GA-Inducible 1,3-β-Glucanases to reopen signal conduits and release dormancy in *Populus*. Plant Cell 23, 130–146.

Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668.

Shao, Y., Zhou, L., Li, F., et al., 2023. Phylogenomic analyses provide insights into primate evolution. Science 380, 913–924.

Shumate, A., Wong, B., Pertea, G., et al., 2022. Improved transcriptome assembly using a hybrid of long and short reads with StringTie. PLoS Comput. Biol. 18, e1009730.

Sigalova, O.M., Chaplin, A.V., Bochkareva, O.O., et al., 2019. Chlamydia pan-genomic analysis reveals balance between host adaptation and selective pressure to genome reduction. BMC Genomics 20, 1–17.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., et al., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Simões, R., Rodrigues, A., Ferreira-Dias, S., et al., 2020. Chemical composition of cuticular waxes and pigments and morphology of leaves of *Quercus suber* trees of different provenance. Plants 9, 1165.

Singer, S.D., Chen, G., Mietkiewska, E., et al., 2016. *Arabidopsis* GPAT9 contributes to synthesis of intracellular glycerolipids but not surface lipids. J. Exp. Bot. 67, 4627–4638.

Sork, V.L., Cokus, S.J., Fitz-Gibbon, S.T., et al., 2022. High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. Nat. Commun. 13, 2047.

Sork, V.L., Squire, K., Gugger, P.F., et al., 2016. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. Am. J. Bot. 103, 33–46.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

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

Sun, P.C., Jiao, B., Yang, Y., et al., 2022b. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Mol. Plant 15, 1841–1851.

Sun, Y., Shang, L., Zhu, Q.H., et al., 2022a. Twenty years of plant genome sequencing: achievements and challenges. Trends Plant Sci. 27, 391–401.

Suyama, M., Torrents, D., Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34, W609–W612.

Tang, D., Jia, Y., Zhang, J., et al., 2022. Genome evolution and diversity of wild and cultivated potatoes. Nature 606, 535–541.

Tian, Y., Zeng, Y., Zhang, J., et al., 2015. High quality reference genome of drumstick tree (*Moringa oleifera* Lam.), a potential perennial crop. Sci. China Life Sci. 58, 627–638.

Ueno, S., Klopp, C., Leplé, J.C., et al., 2013. Transcriptional profiling of bud dormancy induction and release in oak by next-generation sequencing. BMC Genomics 14, 236.

Vasilchenko, A.S., Poshvina, D.V., Sidorov, R.Y., et al., 2022. Oak bark (*Quercus* sp. cortex) protects plants through the inhibition of quorum sensing mediated virulence of *Pectobacterium carotovorum*. World J. Microbiol. Biotechnol. 38, 184.

Waldvogel, A.M., Feldmeyer, B., Rolshausen, G., et al., 2020. Evolutionary genomics can improve prediction of species' responses to climate change. Evol. Lett. 4, 4–18.

Walker, B.J., Abeel, T., Shea, T., et al., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9, e112963.

Walther, G.R., Post, E., Convey, P., et al., 2002. Ecological responses to recent climate change. Nature 416, 389–395.

Wang, H., Moore, M.J., Soltis, P.S., et al., 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. Proc. Natl. Acad. Sci. U.S.A. 106, 3853–3858.

Wang, Y., Tang, H., DeBarry, J.D., et al., 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40, e49.

Wang, Z.Y., Tobin, E.M., 1998. Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. Cell 93, 1207–1217.

Wingett, S., Ewels, P., Furlan-Magaril, M., et al., 2015. HiCUP: pipeline for mapping and processing Hi-C data. F1000Research 4, 1310.

Wu, T., Hu, E., Xu, S., et al., 2021. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation 2, 100141.

Xiang, X.G., Wang, W., Li, R.Q., et al., 2014. Large-scale phylogenetic analyses reveal fagalean diversification promoted by the interplay of diaspores and environments in the Paleogene. Perspect. Plant Ecol. 16, 101–110.

Xie, C., Mao, X., Huang, J., et al., 2011. Kobas 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 39, W316–W322.

Xing, Y., Onstein, R.E., Carter, R.J., et al., 2014. Fossils and a large molecular phylogeny show that the evolution of species richness, generic diversity, and turnover rates are disconnected. Evolution 68, 2821–2832.

Xu, R., Fazio, G.C., Matsuda, S.P.T., 2004. On the origins of triterpenoid skeletal diversity. Phytochemistry 65, 261–291.

Yang, Z., 2007. Paml 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Yeats, T.H., Rose, J.K.C., 2013. The Formation and function of plant cuticles. Plant Physiol. 163, 5–20.

Zachos, J., Pagani, M., Sloan, L., et al., 2001. Trends, rhythms, and aberrations in global climate 65 ma to present. Science 292, 686–693.

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19, 15–30.

Zhang, J., Zhang, W., Ji, F., et al., 2020. A high-quality walnut genome assembly reveals extensive gene expression divergences after whole-genome duplication. Plant Biotechnol. J. 18, 1848–1850.

Zhang, Y., Fernie, A.R., 2018. On the role of the tricarboxylic acid cycle in plant productivity. J. Integr. Plant Biol. 60, 1199–1216.

Zhang, Z., Xiao, J., Wu, J., et al., 2012. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. Biochem. Biophys. Res. Commun. 419, 779–781.

Zhang, Z., 2021. KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. Dev. Reprod. Biol. 20, 536–540.

Zhou, B.F., Yuan, S., Crowl, A.A., et al., 2022b. Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. Nat. Commun. 13, 1320.

Zhou, X., Liu, N., Jiang, X., et al., 2022a. A chromosome-scale genome assembly of *Quercus gilva*: insights into the evolution of *Quercus* section *Cyclobalanopsis* (Fagaceae). Front. Plant Sci. 13, 1012277.