Wang Wen-Bo (Orcid ID: 0000-0002-2559-9326)
Ma Bo (Orcid ID: 0000-0002-5867-1718)
Mao Jian-Feng (Orcid ID: 0000-0001-9735-8516)

Chromosome-scale genome assembly and insights into the metabolome and gene regulation of leaf color transition in an important oak species, *Quercus dentata*

Wen-Bo Wang[1,2 †], Xiang-Feng He[1, †], Xue-Mei Yan[2, †], Bo Ma[1,2], Cun-Fu Lu[2], Jing Wu[1], Yi Zheng[1], Wen-He Wang[1], Wen-Bo Xue[3], Yousry A. El-Kassaby[4], Ilga Porth[5], Ping-Sheng Leng[1*], Zeng-Hui Hu[1*], Jian-Feng Mao[2,6*]

[1] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Engineering Research Center for Ancient Tree Health and Ancient Tree Culture of National Forestry and Grassland Administration, College of Landscape Architecture, Bioinformatics Center, Beijing University of Agriculture, Beijing 102206, China

[2] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, National Engineering Laboratory for Tree Breeding, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, Ministry of Education, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China

[3] BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China

[4] Department of Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

[5] Département des Sciences du Bois et de la Forêt, Faculté de Foresterie, de Géographie et Géomatique, Université Laval Québec, QC, G1V 0A6, Canada

[6] Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, Umeå 90187, Sweden

[*]Corresponding to: lengpsh@tom.com (Ping-Sheng Leng)
or: buahuzenghui@163.com (Zeng-Hui Hu)
or: jianfeng.mao@bjfu.edu.cn (Jian-Feng Mao)

[†]These authors contributed equally to this work.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/nph.18814

Received: 17 May 2022

Accepted: 7 February 2023

ORCID details:

Wen-Bo Wang: https://orcid.org/0000-0002-2559-9326

Ping-Sheng Leng: https://orcid.org/0000-0002-2003-1926

Zeng-Hui Hu: https://orcid.org/0000-0002-8432-6766

Jian-Feng Mao: https://orcid.org/0000-0001-9735-8516

Summary

- *Quercus dentata* Thunb., a dominant forest tree species in northern China, has significant ecological and ornamental value due to its adaptability and beautiful autumn coloration, with color changes from green to yellow into red resulting from the autumnal shifts in leaf pigmentation. However, the key genes and molecular regulatory mechanisms for leaf color transition remain to be investigated.
- First, we presented a high-quality chromosome-scale assembly for *Q. dentata*. This 893.54 Mb sized genome (contig N50=4.21 Mb, scaffold N50=75.55 Mb; 2n=24) harbors 31,584 protein-coding genes. Second, our metabolome analyses uncovered pelargonidin-3-O-glucoside, cyanidin-3-O-arabinoside, and cyanidin-3-O-glucoside as the main pigments involved in leaf color transition. Third, gene co-expression further identified the MYB-bHLH-WD40 (MBW) transcription activation complex as central to anthocyanin biosynthesis regulation.
- Notably, transcription factor (TF) *QdNAC* (*QD08G038820*) was highly co-expressed with this MBW complex and may regulate anthocyanin accumulation and chlorophyll degradation during leaf senescence through direct interaction with another TF, *QdMYB* (*QD01G020890*), as revealed by our further protein-protein and DNA-protein interaction assays.
- Our high-quality genome assembly, metabolome and transcriptome resources further enrich *Quercus* genomics, and will facilitate upcoming exploration of ornamental values and environmental adaptability in this important genus.

Key words: *Quercus*, flavonoids, leaf color transition, co-expression network, gene regulation

Introduction

Oaks (*Quercus* L.), the largest genus within the Fagaceae family comprising between 400-500 species (Manos & Stanford, 2001), are one of the most important trees of temperate forests across the Northern Hemisphere (Nixon, 2006; Denk *et al.*, 2017). China is one of the diversity centers of this genus with more than 60 species, which are the main components of subtropical evergreen broadleaved forests and temperate deciduous forest. *Quercus dentata* Thunb., belonging to the section *Mesobalanus* of *Quercus* genus, is a typical multifunctional ecological ornamental tree thanks to its red autumn foliage (Fig. 1a) and tall umbrella-shaped canopy. *Quercus* species have been widely planted as afforestation and landscape trees due to their longevity, high ornamental value, and high stress resistance (Escandón *et al.*, 2021), as well as their considerable economic value for the use of leaves, bark, acorns, and wood (Del Alamo-Sanza & Nevares, 2018; Morales, 2021).

Q. dentata has been cultivated in China for a long time as its leaves and bark are rich in bioactive compounds and tannin, acorns contain abundant amounts of starch, and the tree trunk provides high-quality wood (Liu *et al.*, 2018). *Q. dentata* leaves are also good fodder to rear the *Antheraea pernyi* silkworm for the production of valuable silk fiber (Hu *et al.*, 2018). In addition, the leaves are often used in Chinese herbal medicine to relieve dysentery, sores and urinary calculi, and most of the pharmacological effects may be attributed to the large amounts of tannins or flavonoids (Şöhretoğlu & Renda, 2020). Moreover, due to ubiquitous natural hybridization within the genus (Ishida *et al.*, 2003; Du *et al.*, 2022), *Q. dentata* is the parent species in many hybrid complexes with other *Quercus* species. Despite its immense potential uses and significant values, the genetic basis of most key traits and the genetic diversity of *Q. dentata* remain to be explored, and high-quality genomic resources are essential for such an endeavor.

The bright red autumn leaves are the most notable ornamental trait of *Q. dentata* as they turn from green to red with temperature decreasing, which is usually associated

with the degradation of chlorophyll controlled by the pheophorbide-α-oxygenase (PAO)/phenyllobilin pathway (Woo *et al.*, 2019) and the accumulation of pigments. Flavonoids have been reported to be among the most important pigments responsible for a variety of colors in higher plants (Fang, 2014; Chen *et al.*, 2019). Among flavonoids, the synthesis and accumulation of anthocyanins confers the orange, red, purple, and yellow colors to the leaves of various plants. Therefore, the accumulation of anthocyanins may play a central role in changing the leaf color of *Q. dentata*. Although the mechanisms of autumn leaf formation have been studied in some trees (Gao *et al.*, 2021) (Renner & Zohner, 2019), the pigments that determine autumn leaf color and regulation in deciduous *Quercus* are still unclear.

Changes in temperature and sunshine duration in autumn are thought to be important in triggering leaf senescence and anthocyanin synthesis in deciduous trees (Wu *et al.*, 2018). Anthocyanin biosynthesis regulated by MYBs has been shown to play a key role in leaf coloration and is associated with senescence (Wen *et al.*, 2021). Dynamic activation of transcription factors (TFs) is thought to play a key role in regulating chlorophyll degradation and anthocyanin accumulation. When plants are exposed to sudden cold spells and strong solar radiation, anthocyanins are highly beneficial for aging leaves, reducing the risk of photooxidative damage to leaf cells and mitigating oxidative damage (Juvany *et al.*, 2013; Gould *et al.*, 2018). In oak species, anthocyanin accumulation increases cold tolerance in species with limited frost tolerance under cold stress (Ramírez-Valiente *et al.*, 2015). Whether the mechanisms regulating leaf senescence and flavonoid synthesis are highly coordinated under the influence of multiple environmental signals remains to be investigated. Therefore, uncovering the biosynthetic and regulatory mechanisms of flavonoids during autumn leaf senescence may lead to a deeper understanding of red leaf formation and bioactive compound production in *Q. dentata*.

As a *Quercus* species native to Asia and a representative species of the section *Mesobalanus* of *Quercus*, the genome sequence of *Q. dentata* is therefore essential not

only for studying the molecular mechanism of formation of outstanding traits including leaf coloration, but also for exploring the evolutionary and phylogenetic position of the section *Mesobalanus* in *Quercus*. Here, a high-quality chromosomal reference genome of *Q. dentata* was generated using long and short read sequencing and Hi-C technologies. The final genome composition is presented in high quality with 893.54 Mb in size (contig N50=4.21 Mb, scaffold N50=75.55 Mb) and 12 chromosome-scale scaffolds and 31,584 annotated protein-coding genes. In addition, we performed a comprehensive metabolome and transcriptome assessment of the leaf color transition. Furthermore, a weight gene co-expression network analysis (WGCNA) was performed to gain insight into the major genes and gene regulation associated with the metabolic changes of flavonoids contributing to leaf coloration. The major metabolic compounds, enzymatic and regulatory genes and their potential regulatory networks were revealed, and the key gene regulations were revealed by yeast one-hybrid (Y1H) and yeast two-hybrid (Y2H) assays. By integrating comprehensive metabolome and transcriptome data, our study contributes to a high-quality reference genome sequence and an in-depth exploration of molecular mechanisms during leaf coloration and senescence. These resources provide the basis for further genetic, genomic, and functional studies in *Quercus* species.

Materials and Methods

Plant materials

A healthy, big in size and fruitful mature oak individual (*Q. dentata*) was selected from naturally regenerated forest at Yinshan Talin Scenic Area, Changping district (116°32′ 74″ E, 40°32′ 02″ N), Beijing, China. The young and healthy leaves from the selected individual were used for genome sequencing and assembly. Different plant tissues including leaves, stems, flowers, roots, ectomycorrhiza and acorns of three development stages (May, July, September) were also sampled from the same tree for RNA-sequencing in support of genome annotation. In addition, leaves from five

different stages (May, July, September, October, November, named as S1-S5) were collected for transcriptome sequencing. And flavonoid metabolic assays were carried out using last three-color change periods (S3-S5) of leaf. All samples were immediately flash frozen using liquid nitrogen and stored at -80°C for subsequent nucleic acid extraction.

Genome sequencing and library construction

The genomic DNA was extracted from the leaves with QIAGEN DNeasy Plant Mini Kit (QIAGEN, Darmstadt, Germany) (Methods S1 for details). For long-read sequencing, the sequencing library was constructed using Ligation Sequencing Kit (SQK-LSK109) from Oxford Nanopore Technologies (ONT). For short-read sequencing, 350-bp paired-end (PE) library was constructed and sequenced on BGI sequencing platform DNBSEQ^{TM}. Short reads were processed with fastp (version 0.23.0) (Chen *et al.*, 2018) and SOAPnuke (version 1.5.6). For Oxford Nanopore PromethION sequencing, we constructed two sequencing libraries with 20-kb DNA insert size and run on the GridION X5 sequencer platform (Oxford Nanopore Technologies, UK). One Hi-C library was prepared (Methods S2 for details) and then sequenced on a BGISEQ platform.

Genome size and heterozygosity estimation

We firstly used Jellyfish v2.1.4 (Marçais & Kingsford, 2011) to count the depth distribution of *K*-mer = 17-31, and then the spectrum of *K*-mer was fitted by GenomeScope v1.0.0 (Vurture *et al.*, 2017) to predict genomic features, such as genome size and heterozygosity.

Genome assembly and quality assessment

The long-read sequences generated by the ONT sequencing platform were assembled by the following steps. First, we used NECAT (https://github.com/xiaochuanle/NECAT) to assemble long-read sequences into contigs and subsequently corrected assembled contigs using Racon (version 0.3.0) (Vaser *et al.*, 2017). Then, short paired-end sequencing reads were aligned to primary contigs with BWA-MEM (version 0.7.17) (Li

& Durbin, 2009) and filtered for properly paired primary alignments using SAMtools (version 1.10). Finally, the assembly was further polished with the second round of pilon using cleaned short reads (Walker *et al.*, 2014) and Unified Genotyper module of the GATK (version 3.5) package (McKenna *et al.*, 2010) was used to obtain SNP and Indel loci.

Chromosome-scale assembly with Hi-C

Hi-C data was used for chromosome-scale scaffolding. First, we used Hic-pro (version 2.11.4) to filter the Hi-C data digested by MboI to obtain clean Hi-C data (Servant *et al.*, 2015). Then, the clean Hi-C data were mapped to the assembled genome using Juicer (version 1.5.6) (Durand *et al.*, 2016). 3D-DNA (version 180922) was used for chromosome-level scaffolding by clustering, sorting, and orientation components based genomic proximity information between Hi-C read pairs (Dudchenko *et al.*, 2017). Finally, the Hi-C cross-linking signal was redrawn by HicPlotter (version 0.7.3), and the adjacent anchored scaffolds were connected forming 12 super scaffolds corresponding to 12 pseudochromosomes. We evaluated the completeness of the assembled genome through Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis using gene content from the embryophyta_odb10 (Seppey *et al.*, 2019).

Gene prediction and annotation

The repetitive elements were identified by a combination of homology-based and *de novo* strategies. For the homology-based repeat library, we used RepeatMasker (version 4.0.7) to identify repeats base on the known consensus sequences of RepBase (version 21.12) (Jurka *et al.*, 2005). *De novo* repeat library was built using RepeatModeler (version 1.0.10). Then, we searched the tandem repeat from *Q. dentata* genome using Tandem Repeats Finder (version 4.09) (Benson, 1999) and LTR_FINDER (version 1.06). Furthermore, we used LTR_FINDER to identify and classify the candidate LTR-retrotransposons (Xu & Wang, 2007).

We performed three strategies to predict and annotate genes in the *Q. dentata* genome including *de novo*, homolog-based and transcriptome-based approaches. For

the *de novo* prediction, we predicted genes by Glimmer HMM (version 3.0.4) (Majoros *et al.*, 2004), Augustus (version 3.0.2) (Stanke *et al.*, 2006), Genscan (version 1.0) (Li *et al.*, 2010). For the homolog-based prediction, the protein sequences of *Oryza sativa*, *Arabidopsis thaliana*, *Vitis vinifera*, *C. mollissima*, *F. sylvatica* and *Q. robur* were aligned against the *Q. dentata* genome using tblastn with E-value < 1E-5 (Altschul *et al.*, 1997). Genewise (version 2.4.1) was employed to annotate genes based on alignments (Birney *et al.*, 2004). For transcriptome-based prediction, RNA-seq data were aligned to the genome using HISAT2 (version 2.2.1) (Kim *et al.*, 2015) and assembled to gene models by Cufflinks (version 2.2.1) (Trapnell *et al.*, 2012). Finally, with the help of Evidence Modeler (EVM) (version 1.11) (Haas *et al.*, 2008), all the genes predicted by various methods were integrated into a non-redundant and complete gene set. The functional annotation of protein-coding genes was carried out by running eggNOG-mapper (version 2.1.6) and BLAST (version 2.2.3.1) (E-value $\leq 1e^{-5}$) against NCBI non-redundant (NR) protein databases, Swiss-Prot (Bairoch & Apweiler, 2000), GO database (Ashburner *et al.*, 2000), pfam (Mistry *et al.*, 2021), tair (Poole, 2007), COG (Galperin *et al.*, 2019), and KEGG (Kanehisa & Goto, 2000).

Genome comparison and phylogenomics

We performed comparative genome analysis among *Quercus robur* (Plomion *et al.*, 2018), *Quercus suber* (Ramos *et al.*, 2018), *Quercus lobata* (Sork *et al.*, 2016), *Q. mongolica* (Ai *et al.*, 2022), *Q. acutissima* (Fu *et al.*, 2022), *Q. variabilis* (Han *et al.*, 2022), *Fagus sylvatica* (Mishra et al., 2018), *Castanea mollissima* (Xing et al., 2019), *Morella rubra* (Jia et al., 2019), *Carpinus fangiana* (Yang, X *et al.*, 2020), *Carya cathayensis* (Huang et al., 2019), *Solanum tuberosum* (Potato Genome Sequencing et al., 2011), *Solanum lycopersicum* (Tomato Genome, 2012), *Arabidopsis thaliana* (Cheng et al., 2017), *Vitis vinifera* (Jaillon et al., 2007), *Oryza sativa* (Yu et al., 2002), *Carica papaya* (Ming et al., 2008), and *Q. dentata* genome. We obtained protein sequences from each of the above-mentioned genome data to construct protein gene sets of multiple species and used OrthoFinder (version 2.2.6) to cluster the selected

protein sequences and screen low-copy genes as orthologous genes (Emms & Kelly, 2019). The concatenated amino acid sequences alignment was created by MUSCLE (version 3.8.31) (Edgar, 2004) and the phylogenetic tree was constructed using RAxML (version 8.2.4) (Stamatakis, 2014). CODEML program in the PAML (version 4.5) software package was used to estimate the molecular clock (divergence time and replacement rate), and the MCMCTree program was used to calculate the differentiation time between species. We queried the Timetree (http://www.timetree.org/) database to correct the species divergence time (Yang, 2007). CAFÉ (version 4.1) was used to analyze the expansion/contraction of gene families based on the chronogram of these 18 species and filter out gene families whose total number of genes in the family is greater than 200 and the number of genes in any species is 0 (De Bie *et al.*, 2006).

Genome synteny analysis

We performed synteny searches to identify syntenic blocks within *Q. dentata* and between *Q. dentata* and *Q. robur*, *Q. lobata*, *Q. variabilis*, *Q. acutissima*, *Q. mongolica*, *Q. suber*, *F. sylvatica*, and *V. vinifera* using MCScanX (Wang *et al.*, 2012). Subsequently, *K*s (number of substitutions per synonymous site) substitution rates of gene pairs across syntenic blocks were calculated to identify WGD (whole genome duplication) events. The dot plots between *Q. dentata* and *Q. dentata* as well as *V. vinifera* genome was generated with Quota synteny alignment software (Tang *et al.*, 2011).

Transcriptome sequencing and analysis of DEGs

The samples from leaves at five different color change stages, stems, flowers, roots, ectomycorrhiza and acorns with three biological replicates were harvested from the same tree and processed for RNA extraction and cDNA library construction. The filtered clean reads were mapped to the *Q. dentata* genome using HISAT2 (version 2.0.4) (Kim *et al.*, 2015). Bowtie2 (version 2.2.5) was applied to align the clean reads to the reference coding gene set (Langmead & Salzberg, 2012), then the number of reads

mapped to the assembled transcriptome was input to RSEM (version 1.2.12) to obtain read count for each gene. (Li & Dewey, 2011). We performed DESeq2 package to analyze differential expression genes (DEGs) and define genes with log2 fold change (FC) >1, P-value ≤ 0.001 and false discovery rate (FDR) < 0.05 as significant (Love *et al.*, 2014).

Flavonoids-Targeted metabolic analysis

Leaves of S3-S5 stages were collected for metabolomic analysis and analyzed using UPLC and MS/MS system (as described in Methods S3). The difference of among three leaf color change groups for all 19 metabolic variables were evaluated using the nonparametric Kruskal–Wallis multiple-range test (De Mendiburu, 2014). Principal component analysis (PCA) was implemented to the data of all 19 metabolic variables using ade4 (Dray & Dufour, 2007), and the relative contribution of each metabolic variables to the two dimensional spaces defined by the first two principal components(PCs) was then illustrated in a PCA distance biplot.

Analysis of genes involved in flavonoid biosynthesis during leaf color change

The flavonoid biosynthesis genes and chlorophyll catabolic genes (CCGs) in *Q. dentata* were identified based on genome annotation and combining local blast and phylogenetic analysis. Hidden Markov Model (HMM) profiles from Pfam database were used to search against proteome using HMMER (version 3.1) with E-value<0.01 (Wheeler & Eddy, 2013). We identified transcription factors (TFs) in the *Q. dentata* genome using PlantRegMap (Tian *et al.*, 2020). The 2000-bp sequences upstream of crucial genes were defined as the promoter region, from which the cis-acting regulatory elements were identified with PlantCARE (Lescot *et al.*, 2002) and PlantTFDB (version 5.0) (Tian *et al.*, 2020).

Weighted correlation network analysis

After discarding relative low expression genes (the FPKM was less than 0.5 in more than 9 samples), the R package WGCNA (Langfelder & Horvath, 2008) was used to identify modules of highly correlated genes attributable to flavonoids base on the

filtered FPKM data. The co-expression modules were obtained using automatic network construction function (blockwiseModules) with power =15, minModuleSize =100, TOMtype was signed, modules whose eigengenes were highly correlated (correlation > 0.8) were merged. Eigengene value was calculated for each module based on pearson correlation, which was used to search the modules association with leaf color change products (flavonoids). The networks were visualized by Cytoscape (v.3.8.2) (Shannon *et al.*, 2003).

Yeast one-hybrid and Yeast two-hybrid assays

For the yeast one-hybrid (Y1H) assay, the CDS fragments of *QdNAC* (*QD08G038820*) and *QdMYB* (*QD01G020890*) were inserted into the pGADT7 vector. In addition, the upstream 2 kb promoter fragment of *QdSGR* and *QdCHS* were inserted into the pAbai vector and transformed into the Y1HGold yeast strain. Then, the *QdNAC*-AD and *QdMYB*-AD plasmid were transformed into the recombinant yeast strain and empty pGADT7 plasmid was transformed into the recombinant yeast as a control. The recombinant plasmids with the empty pGADT7 and promoter fragments of *QdSGR*/*QdCHS* were spread on different concentrations of SD/-Leu/AbA medium to screen for inhibition of self-activation concentration. Finally, the transformants were spread SD/−Leu/AbA (aureobasidin A) culture medium.

For the yeast two-hybrid (Y2H) assay, the CDS fragments of *QdNAC* (*QD08G038820*) and *QdMYB* (*QD01G020890*) were inserted into the pGBKT7 and pGADT7 vectors to construct pGBKT7-*QDNAC* and pGADT7-*QdMYB* vectors, respectively. Then, AD and BK recombinant plasmids were co-transformed into Y2H and empty pGADT7 and pGBKT7-*QDNAC* were co-transformed into Y2H as a control, which was spread on SD-Leu/Trp, SD-Trp/Leu/His + X-α-gal and SD-Leu/Trp/His/Ade + X-α-gal.

Results

***Q. dentata* genome sequencing and assembly**

Based on *K*-mer analysis, *Q. dentata* genome size was estimated ~841 Mb in size with

a repeat rate of 47.02% and a heterozygosity rate of 1.53% (Fig. S1). We produced 141.71 Gb (169-fold coverage of the estimated genome size) Oxford Nanopore long reads (Fig. S2), 57.89 Gb short reads, and 154.37 Gb (184-fold coverage) Hi-C paired-end reads (Table S1). We first used Nanopore long-reads to assemble a 1.08 Gb genome (contig N50 4.21 Mb). The assembled genome had 893.54 Mb total size, composed of 312 contigs, with contig N50 of 4.8 Mb. Then, we used Hi-C data to assist the assembly correction and anchored 312 contigs into 12 chromosome-level supper scaffolds with scaffold N50 of 75 Mb (Fig. 1b; Table S2). The Hi-C interaction matrix displayed a strong intra-chromosomal interactive signal along the diagonal (Fig. S3). We evaluated the integrity of gene assembly through BUSCO analysis, and 95.7% of complete eukaryotic conserved genes were identified (Table S3). In addition, 30 million Nanopore reads were aligned to the *Q. dentata* genome and yielded a 99.97% mapping in the whole genome, with an average coverage depth of ~161X (Table S4; Fig. S4, S5). The above results confirmed that the *Q. dentata* genome assembly has a high degree of completeness and accuracy at the chromosome level.

***Q. dentata* genome annotation**

A total of 667,446,084 bp repeat elements were identified, which accounted for 74.7% of the assembled *Q. dentata* genome. Transposable elements (TEs) were dominant, accounting for about 53.57% (478.7 Mb) of the genome. Further analysis showed significantly higher proportion of retrotransposons compared to DNA transposons. Similar to other plants, LTRs (309.2 Mb, 34.6%) were the predominant retrotransposons with Gypsy (18.22%) the most abundant, followed by Copia (15.35%) (Table S5). And the LTR Assembly Index (LAI) for *Q. dentata* is 8.29.

We applied a combination of *de novo*, homology and RNAseq-based strategies to annotate the protein coding genes in *Q. dentata*. A total of 42 RNA-sequencing libraries from leaves, roots, stems, and acorns were used for annotation. Finally, 31,584 genes were predicted, and the average length of total gene regions, coding sequences (CDS), intron sequences, and exon sequences were 5,243, 1,293, 711, and 297 bp, respectively

(Table S6). We also identified 2,058 non-coding RNAs, including 251 miRNAs, 788 transfer RNAs (tRNA), 177 ribosomal RNAs (rRNA), and 797 small nuclear RNAs (snRNA) (Table S7). Furthermore, the chromosomal locations of ribosomal RNA genes were visualized (Table S8), and the number of rRNA genes in other sequenced *Quercus* species was also counted (Table S9). Among the predicted protein coding genes, 30,793 (97.5%) genes could be annotated to at least one of the following databases: NCBI non-redundant protein database (NR) (97.48%), RDP (88.21%), TAIR (89.29%), the Swiss-Prot protein database (79.33%), Gene Ontology (GO) (35.63%), KEGG PATHWAY (44.06%), COG (86.51%), EggNOG (86.51%), and the protein families database (Pfam) (77.28%) (Table S10).

In addition, we identified 1,920 genes encoding transcription factors (TFs) belonging to 92 diverse TF families, which accounted for 6.07% of all protein-coding genes. We identified 149 MYB TFs as the largest family, followed by AP2-ERF (102), bHLH (101), C2H2 (100), NAC (92), and WRKY (68) families.

Comparative and evolutionary genomics of *Q. dentata*

To investigate the phylogenetic relationship and evolutionary status of *Q. dentata*, the dated phylogenetic tree was reconstructed for 18 plant species including nine Fagaceae species (*Quercus dentata*, *Quercus robur*, *Quercus lobata*, *Quercus suber, Quercus. variabilis*, *Quercus acutissima*, *Quercus mongolica*, *Castanea mollissima*, and *Fagus sylvatica*) and nine other sequenced plants. A total of 17,144 gene families (one gene family is just one orthogroup inferred by OrthoFinder2) encompassing 30,798 genes were identified and 134 species-specific gene families were obtained for *Q. dentata* (Table S11). We used 1053 low-copy orthogroups to reconstruct the phylogenetic relationship between *Q. dentata* and the other 17 plant species. Results indicate the closest relationship between *Q. dentata* and the six other *Quercus* species (Fig. 2a). Among them, *Q. dentata* from the section *Mesobalanus* clusters with *Q. mongolica*, *Q. robur* and *Q. lobata* from the section *Quercus* in one clade, which is making a sister group to *Q. acutissima*, *Q. variablils* and *Q. suber* from the section *Cerris*. The

divergence between *Q. dentata* and *Q. mongolica* and *Q. robur* occurred approximately 8.7 and 10.9 million years ago (Mya), respectively. Moreover, the seven *Quercus* species diverged from *C. mollissima* and *F. sylvatica* around 23.6 and 50.8 Mya, respectively, following their divergence from the common ancestor of Fagaceae species. In their common order Fagales, Fagaceae diverged from other distantly related families such as the Myricaceae around 75 Mya (Fig. 2a).

Comparative analysis of gene families revealed 7,928 shared gene families among eight Fagaceae species (*Q. robur*, *Q. suber*, *Q. lobata*, *Q. mongolica*, *Q. variabilis*, *Q. acutissima*, *F. sylvatica* and *Q. dentata*) and 257 gene families unique to *Q. dentata* (Fig. 2d). In comparison with gene families of other 17 plant species, 2,405 gene families were found expanded, and 3,716 gene families contracted in *Q. dentata* (Fig. 2a). GO functional enrichment analysis revealed the expanded gene families were significantly enriched in biological processes related to cellular response to hydrogen peroxide and defense response to insect (Fig S6). KEGG functional enrichment analysis of the expanded gene families showed significant enrichment in plant-pathogen interaction (Fig. S7), with a significant expansion of 1551 members of receptor-like kinase (RLK)-encoding gene (Fig. S8).

The *K*s age distributions and synteny analyses were used to unveil WGD events. Four *Quercus* species (*Q. robur*, *Q. suber*, *Q. lobata*, *Q. dentata*) and *Vitis vinifer* present the same signature of *K*s peaks only at about 0.15, indicating that none of the *Quercus* species underwent WGD events (Fig. 2b) since the gamma triploidization. Moreover, we identified 268 syntenic blocks comprising 3579 collinear genes in *Q. dentata* through intra-genomic synteny analyses (Fig. S9; Table S12). We also built a gene homology dot plot and calculated *K*s values of one-versus-one orthologs between *V. vinifera* and *Q. dentata* (Fig. S10). The chromosome collinearity between *Q. dentata* and *V. vinifera* showed a 1:1 correspondence, and a total of 15,942 gene pairs and 740 blocks were obtained (Fig. S11a). These results were consistent with the *K*s distribution, suggesting that *Q. dentata* lacks independent WGD in its evolutionary history. In

addition, chromosome collinearity analysis between *Q. dentata* and *Q. robur*, *Q. suber*, *Q. lobata*, *Q. variabilis*, *Q. acutissima*, *Q. mongolica*, and *F. sylvatica* showed 1:1 syntenic patterns suggesting a close evolutionary relationship (Fig. 2c; Fig. S11b; Table S13).

Flavonoid metabolic profiling during leaf color change of *Q. dentata*

We divided the process of color change in *Q. dentata* into five distinct development stages (S1-S5), which involved a gradual darkening of the leaf's green during April to September, its color change to orange in October, and its color transformation to red in November (Fig. 3a). To investigate the metabolic differences during this color change, we carried out a targeted metabolomic approach to determine the flavonoid contents in the three stages S3-S4-S5. In total, 175 flavonoid metabolites were detected and identified by UPLC-MS/MS, including anthocyanins, flavanols, dihydroflavonols, flavonols, and chalcone (Table S14). By calculating the fold-change of metabolite content among these three stages, 19 flavonoid metabolites exhibited significant differences. Along with the color change in *Q. dentata* leaves, anthocyanins content (cyanidin-3-O-glucoside, cyanidin-3-O-arabinoside, delphinidin-3-O-glucoside and pelargonidin-3-O-glucoside) increased the most. Dihydroflavonols (taxifolin, dihydrokaempferol), and chalcone (naringenin chalcone, naringenin) increased to varying degrees compared to the green-leaf stage, but the content of flavanols (catechin) (epicatechin, catechin, gallocatechin, epigallocatechin and epiafzelechin) declined (Fig. 3b).

PCA analysis of the main 19 metabolites from the flavonoid synthesis pathway identified the first two components (with eigenvalues > 1) cumulatively accounted for 89.2% of the total variation, of which PC1 accounted for 61.6% (Fig. 3c). The variance of different leaf color development stages can be clearly distinguished with PC1, and the three distinct groups related to the time series of leaf discoloration. As revealed in the PCA biplot, anthocyanins (cyanidin-3-O-glucoside, cyanidin-3-O-arabinoside, delphinidin-3-O-glucoside, pelargonidin-3-O-glucoside), dihydroflavonols (taxifolin,

dihydrokaempferol), chalcones (naringenin chalcone), and flavonols (quercetin, kaempferol) were positively associated with the color change from green to red. Flavanols (epicatechin, catechin, gallocatechin, epigallocatechin, epiafzelechin) and dihydroflavonols (leucocianidol) were negatively associated with leaf color. KEGG pathway enrichment analysis among three leaf color change stages showed that anthocyanin/flavonoid biosynthesis was significantly enriched (Fig. S12). The main metabolic changes contributing to the transition from S3 to S5 development stages involved relatively high content of anthocyanins (pelargonidin-3-O-glucoside, cyanidin-3-O-arabinoside, and cyanidin-3-O-glucoside) with significant differences among the three stages.

Flavonoid gene expression during leaf color change in *Q. dentata*

To explore the molecular regulatory mechanisms of autumn leaf color formation, RNA sequencing was performed on leaves for the five development stages. Combined with our high-quality *Q. dentata* genome, we reconstructed the flavonoid synthesis and carotenoid synthesis pathways for leaf coloration (Fig. 4; Fig. S13). An average output of 6.35 Gb clean reads per sample was obtained (Table S15), with an average mapping rate of 79.90%, and a total of 33,872 DEGs were identified, among which the highest number of DEGs was S1 vs S4 with 22,869 (Fig. S14). The cluster dendrogram of transcriptomes showed that the gene expression patterns were divided into two groups, the green leaf group (S1-S3 stages) and the autumn leaf group (S4-S5 stages) (Fig. 3d). In agreement, both *Q. dentata* metabolome and transcriptome, revealed significant specificity during the progression of leaf coloration. We annotated 51 enzymatic genes in the flavonoid biosynthesis pathway and visualized their expression patterns at the five leaf color change stages (Fig. 4). By reconstructing the flavonoid biosynthesis pathway for the *Q. dentata* leaf, we found *Qd4CL* (QD09G051600), *QdCHS* (QD06G044950), *QdF3'H* (QD10G034000), *QdANS* (QD12G029520), and *QdA3oGT* (QD09G052790) genes expressed at relatively high levels in the autumn leaf group S4-S5, coinciding with significant accumulation of the main coloration metabolites

pelargonidin-3-O-glucoside, cyanidin-3-O-glucoside, cyanidin-3-O-arabinoside, cyanidin-3,5-O-diglucoside, and delphinidin-3-O-glucoside (Fig. 3b; Fig. 4).

Gene co-expression network associated with leaf coloration in *Q. dentata*

To obtain further insights into the regulation of flavonoid related metabolic changes during leaf coloration, we performed weight gene co-expression network analysis (WGCNA) analysis. We retained 21,651 differentially expressed genes (DEGs) with Fragments Per Kilobase Million (FPKM) greater than 0.5 and a significant difference among samples at various stages. A total of 11 co-expression modules were identified (Fig. 5a, color-coded). Among those sub-networks of gene co-expression (Fig. 5b), some strongly correlated with the five anthocyanins naringenin, naringenin chalcone, taxifolin, dihydrokaempferol, and kaempferol and their downstream metabolites. Among those, the three main leaf color-rendering anthocyanins cyanidin-3-O-glucoside, cyanidin-3-O-arabinoside, and pelargonidin-3-O-glucoside showed the highest correlation coefficients with the brown-coded module, that is 0.99, 0.98 and 0.98, respectively. Gene expression patterns for the brown-, magenta-, and red-coded modules all showed up-regulation from stages S3 to S5 (Fig. 5c). GO enrichment analysis exhibited that genes in the brown module were significantly enriched for the organonitrogen compound catabolic and the protein catabolic processes within biological processes (Fig. S15).

We identified structural genes of the *Quercus* flavonoid metabolic pathway and CCGs for the brown-coded module and examined a total of 198 related TFs for this module, mostly members of MYB, bHLH, WD40, NAC, WRKY, MADS, bZIP, and AP2-EREBP families. There, we identified six enzymatic genes for flavonoid synthesis and five chlorophyll catabolic genes, including *4CL* (*QD09G051600*, *QD01G027240*, *QD02G142950*), *CHS* (*QD06G044950*), *F3'H* (*QD10G034000*), *ANS* (*QD12G029520*), and *NYC1* (*QD02G167200*), *SGR* (*QD12G029700*), *PPH* (*QD04G017270*), *PAO* (*QD01G025900*), *RCCR* (*QD06G080940*), respectively, whose gene expression was highly correlated with the accumulation of five anthocyanins at the autumn leaf stages

S4-S5 (Table S16). The potential regulatory networks for flavonoid metabolite synthesis and chlorophyll degradation were also resolved (Fig. 6b). Based on gene connectivity pattern within the brown-coded module, 18 genes were hub genes, two of which are TF genes *MYB* (*QD06G024550*) and *bHLH* (*QD06G032240*).

To further screen the TFs for potential binding affinity to flavonoid metabolizing structural genes and chlorophyll catabolic genes, we predicted the DNA binding sites of promoter regions located 2.0 kb upstream of *CHS*, *ANS, F3'H*, *PAO*, *NYC1,* and *SGR* genes. All three flavonoid metabolizing structural genes had DNA binding sites for MYB, bHLH, and bZIP, while the three CCGs rather displayed NAC binding sites (Fig. 6a).

Identification of key TFs and potential regulatory network associated with leaf color transition in *Q. dentata*

To address the regulatory roles of key TFs in flavonoid metabolism during leaf color transition, we focused our analysis on the potential regulatory network of *ANS* and *CHS*. *CHS* and *ANS* represent the key biosynthetic enzymes in the early and late stages of flavonols/anthocyanins synthesis, respectively, and the corresponding genes were highly expressed at stages S4 and S5. In addition, we analyzed the regulatory mechanisms of three CCGs (*PAO*, *NYC1,* and *SGR*) that may control chlorophyll degradation during leaf senescence.

By correlating TFs binding sites and expression patterns of key TFs, we identified 48 important TFs from the brown-coded co-expression module, including MYB, bHLH, NAC, WRKY, WD40, and bZIP (Fig. 6c). Further, two sub-networks were inferred for the brown-coded module with TFs as potential regulators of chlorophyll catabolic genes (Fig. 7a) and flavonoid biosynthesis genes (Fig. 7b). Among those TFs, most members (11) belonged to the *MYB* family in the sub-network of *ANS* and *CHS*, and 11 members of the NAC family were found in the CCGs sub-network. Most of these TFs were barely expressed at green leaf stages S1-S3, but highly expressed at autumn leaf stages S4-S5. These TF expression patterns were consistent with those of key structural genes in

flavonoid metabolism and the chlorophyll catabolic pathway (Fig. 4). We note that TF expression at S5 already declined compared to S4.

We found seven key TFs that may simultaneously regulate anthocyanin accumulation and chlorophyll degradation, with *CHS* and *ANS* genes as the hubs within these potential regulatory relationships (Fig. 7d). *Cis*-acting element analysis and expression patterns supported the regulatory network of the related TFs, among which MYB, WD40, bHLH, NAC, WRKY, and bZIP may play important roles (Fig. 7c). The promoter regions of *CHS* and *ANS* genes both carry MYB binding sites. We could infer that *CHS* and *ANS* may be directly regulated by the MYB-bHLH-WD40 (MBW) transcriptional activator complex derived from TF genes *MYB QD01G020890*, *bHLH QD06G032240*, and *WD40 QD03G039610* (Fig. 7d). TFs bZIP, WRKY, and NAC may participate as secondary regulators. The *cis*-acting element analysis showed that the *MYB* gene *QD01G020890* may be regulated by other upstream TFs such as HY5, WRKY, and NAC (Fig. 7c). Concurrently, *NAC QD08G038820* and *MYC QD08G028710* TF genes of the bHLH family may act as positive regulators of leaf senescence, whereby their gene products directly bind to the promoters of three CCGs and therefore further regulate chlorophyll degradation (Fig. 7d). In addition, *HY5 QD02G034660* (one bZIP member) encoding a light-induced TF may regulate anthocyanin synthesis and accumulation by activating NAC and MYB TFs. Furthermore, we found CRT/DREB response elements in the promoters of *MYB* and *NAC*, suggesting that these genes may be regulated by CBF TFs (Fig. 7c).

QdNAC binds to the promoter of _QdSGR_ and interacts with _QdMYB_

To further verify regulatory relationships among key TFs and structural genes involved in leaf color transition, we performed yeast one-hybrid (Y1H) and yeast two-hybrid (Y2H) assays. Y1H assays showed that the *QdMYB QD01G020890* gene product could bind to *cis* elements in the promoter region of gene *QdCHS QD06G044950* to regulate its expression. The Y1H results also confirmed the binding of the *QdNAC QD08G038820* gene product to the promoter region of gene *QdSGR QD12G029700*

(Fig. 7f). The Y2H result showed that gene products *QdNAC QD08G038820* and *QdMYB QD01G020890* can specifically interact (Fig. 7e).

Through comprehensive analysis of potential hierarchical regulation of TFs, genes *QdMYB QD01G020890* and *QdNAC QD08G038820* were identified as important upstream regulators that may respond to environmental cues (photoperiod and temperature) and further modulate the flavonoid metabolic shifting during autumn leaf coloration and senescence.

Discussion

Here we present a high-quality chromosome-scale assembly for *Q. dentata*, a genome of an oak native to Asia. *Ks* distribution and collinearity analyses revealed that there were no recent WGD events in *Q. dentata* or other *Quercus* species, confirming that a recent burst of local gene duplications in *Q. robur* and *Q. lobata* occurred mainly after the ancient γ whole genome hexaploid duplication event and by tandem duplication (Sork *et al.*, 2022). A comparison of gene families from 18 species revealed expanded gene families in *Q. dentata* that are highly enriched in plant-pathogen interaction. Previous studies of *Q. robur* have also shown that gene duplication mainly caused by TD events which amplified families of genes involved in defense against pathogens such as NB-LRR and RLK (Plomion *et al.*, 2018). In addition, we found that the tandemly duplicated genes of *Q. dentata* largely contributed to the proportion of enzyme genes within the anthocyanin biosynthetic pathway, which is similar to the results of azalea flower color study (Yang *et al.*, 2020). All the above results suggest that TD events may have played an important role in the accumulation of anthocyanins during leaf color change and maintaining the immune system in oak trees.

Anthocyanin pigments and autumn leaf color transition

Autumn leaf coloration is not only an ornamental trait, but foremost a strategy to perform physiological functions in response to environmental stress. Eastern North America and Asia experience more severe temperature fluctuations and strong solar irradiation in autumn, which result in a higher frequency of autumn leaf colored species

(Renner & Zohner, 2019). During autumn leaf senescence, the decomposition process of chlorophyll (Chl) induced by low temperature and decreasing sunshine duration is usually accompanied by the synthesis of anthocyanin pigments in vacuoles (Gould *et al.*, 2018). In northern China autumn, *Q. dentata* leaves quickly turn red within two weeks as temperatures drop. We identified 19, 45, and 43 significantly different flavonoid metabolites between leaf color stages S3/S4, S3/S5, S4/S5, respectively (Table S17). The three anthocyanins pelargonidin-3-O-glucoside, cyanidin-3-O-arabinoside, and cyanidin-3-O-glucoside were the most important compounds in terms of abundance and composition of differentially accumulating metabolites in leaf color change in autumn. As a major class of flavonoids, anthocyanins are beneficial to autumn leaves, protecting them against high light intensities and serving as antioxidants in free radical scavenging (Pourcel *et al.*, 2007; Xu & Rothstein, 2018). Our results also suggest that adjusting the accumulation of certain anthocyanins in autumn may help *Q. dentata* cope with low temperatures and high light exposure.

Gene co-expression and potential regulatory network of flavonoid biosynthesis

Based on the chromosome-level genome sequence, in combination with gene co-expression and flavonoid metabolome analyses, we first reconstructed the entire anthocyanin/flavonoid metabolic pathway and its regulatory networks in *Q. dentata*. The increasing expression of structural genes, such as *CHS*, *F3'H*, *ANS,* and *A3oGT* encoding key enzymes within the flavonoid metabolic pathway, paralleled the accumulation of anthocyanins in autumn. It has been reported that *CHS*, *ANS*, and *UGT* are involved in the low-temperature induced anthocyanin accumulation (Ubi *et al.*, 2006; Li *et al.*, 2017). Our WGCNA analysis identified *QdCHS*, *QdF3'H*, *Qd4CL,* and *QdANS* genes encoding key enzymes of the anthocyanin biosynthetic pathway within a gene co-expression module (the brown-coded one) with the highest correlation coefficient with anthocyanins and flavonols content. This module also contained five key enzyme encoding genes (*QdNYC1*, *QdSGR*, *QdPPH*, *QdPAO,* and *QdRCCR*) for chlorophyll degradation, indicating that the synthesis of anthocyanins is synchronized

with chlorophyll degradation. Nowadays, the regulation of MYB-bHLH-WD (MBW) transcriptional activation complex in the process of flavonoid synthesis is well described in many plant species (Xu *et al.*, 2015), and some studies have confirmed the role of MYB TFs in the regulation of *CHS* and *ANS* genes (Vimolmangkang *et al.*, 2013; Xu *et al.*, 2020). From our gene co-expression networks, it can be inferred that the MBW complex may be able to directly regulate *CHS* and *ANS* genes to promote the accumulation of anthocyanins and flavonols in the leaf coloration process of *Q. dentata*. In our study, we confirmed that *QdMYB* is able to bind to the promoter of *QdCHS* and activate gene transcription. The co-expression networks showed that WRKY, NAC, and bZIP TFs may also play important roles in regulating anthocyanin accumulation and chlorophyll degradation. They could act as secondary regulators to induce transcriptional activation of MYB. Oxidative stress caused by environmental cues, such as low temperature and change in photoperiod, could regulate the expression of R2R3-MYB, leading to an overproduction of flavonoids (Zheng *et al.*, 2016). HY5, a photosensitive protein of the bZIP family in plants, was reported to regulate anthocyanin biosynthesis through certain TFs (Shin *et al.*, 2013). By studying the coexpression of the *Q. dentata HY5* gene and anthocyanin biosynthesis genes, we concluded that the *HY5 QD02G034660* may respond to excess light after chloroplast degradation in autumn and regulate anthocyanin biosynthesis by further activating MYB or NAC.

WRKY and NAC are known to be responsible for biotic/abiotic stress responses (Rushton *et al.*, 2010), and TF NAC may also regulate plant senescence (Li *et al.*, 2018). However, how NACs are involved in both the regulation of anthocyanin biosynthesis and chlorophyll degradation during autumn leaf coloration is still unknown. We found several NAC TFs with high expression during leaf color transition in *Q. dentata* (Fig. 6c), indicating that they may act as upstream regulators for anthocyanin accumulation. Previous studies have shown that NAC can interact with MYB to regulate anthocyanin accumulation in red-fleshed apple (Zhang *et al.*, 2020). Interestingly, CRT/DREB

response elements are present in promoters of both NAC and MYB. Therefore, we hypothesized that roles for *QdMYB* (*QD01G020890*) and *QdNAC* (*QD08G038820*) genes in the stress response to low temperature were dependent on the CBF/DREB pathway, which is activated by a sharp drop in temperature. When *Q. dentata* senses low temperatures and decreasing sunshine duration in fall, *QdNAC* could directly activate the expression of *QdSGR* and regulate chlorophyll degradation. Our study showed that *QdNAC* interacts with *QdMYB*, suggesting that *QdNAC* is not only responsible for regulating autumn leaf senescence but may also be involved in regulating anthocyanin synthesis, which plays an important role in autumn coloration of *Q. dentata*. We suggest that *QdNAC* (*QD08G038820*) acts as a potential target for HY5 and promotes anthocyanin accumulation by regulating MYB via interactions.

In sum, the high-quality reference genome of *Q. dentata* presented in this study provided resources for studying evolution and diversity in *Quercus*. Moreover, the comprehensive transcriptome and flavonoid metabolic profiling provided novel insights into leaf color change during autumn leaf senescence. Furthermore, our data represent crucial resources for functional genomics and genetic studies in *Quercus*.

Acknowledgements

This study was supported by the National Key R&D Program of China (2018YFD1000406), the Joint Project of Beijing Municipal Education Commission and Beijing Municipal Natural Science Foundation (KZ201810020028), the Beijing Municipal Education Commission for their financial support through the Innovative Transdisciplinary Program of Ecological Restoration Engineering, and the grant for Beijing Forestry University Outstanding Postgraduate Mentoring Team Building (YJSY-DSTD2022005).

Author contributions

P-S.L., X-F.H., Z-H.H., and J-F.M. conceived and designed the study; W-B.W., X-F.H., B.M., J.W., Y.Z., and W-B.X. prepared the materials, conducted the experiments, W-

B.W., X-M.Y., analyzed the data and prepared the results; W-B.W. wrote the manuscript; J-F.M., P-S.L., Z-H.H., X-F.H., W-H.W., C-F.L., I.P. and Y.A.E-K. edited and improved the manuscript; all authors approved the final manuscript. W-B.W., X-F.H. and X-M.Y. contributed equally to this work.

Competing interests

None declared.

Data availability

Raw data (both of DNA and RNA sequencing) generated in the present study have been deposited under the NCBI Bio-Project: PRJNA781072. The assembled genome were uploaded in NCBI with accession number GCA_028216015.1 that are available at (https://www.ncbi.nlm.nih.gov/assembly/GCA_028216015.1/). The *Q. dentata* genomic data are also publicly available in the National Genomics Data Center under the accession number GWHBRAD00000000 with the BioProject number PRJCA013491 (https://ngdc.cncb.ac.cn/search/?dbId=gwh&q=GWHBRAD00000000). Genome assembly, gene structure annotation, mRNA, CDS and protein sequences profiles are also available at FigShare (https://doi.org/10.6084/m9.figshare.21624159.v1).

References

Ai W, Liu Y, Mei M, Zhang X, Tan E, Liu H, Han X, Zhan H, Lu X. 2022. A chromosome-scale genome assembly of the Mongolian oak (*Quercus mongolica*). *Mol Ecol Resour* **22**(6): 2396-2410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17): 3389-3402.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**(1): 25-29.

Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **28**(1): 45-48.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**(2): 573-580.

Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. *Genome Res* **14**(5): 988-995.

Chen L, Hu B, Qin Y, Hu G, Zhao J. 2019. Advance of the negative regulation of anthocyanin biosynthesis by MYB transcription factors. *Plant Physiol Biochem* **136**: 178-187.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**(17): i884-i890.

Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. 2017. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J* **89**(4): 789-804.

De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**(10): 1269-1271.

De Mendiburu FJRpv. 2014. Agricolae: statistical procedures for agricultural research. **1**(1): 1-4.

Del Alamo-Sanza M, Nevares I. 2018. Oak wine barrel as an active vessel: A critical review of past and current knowledge. *Crit Rev Food Sci Nutr* **58**(16): 2711-2726.

Denk T, Grimm GW, Manos PS, Deng M, Hipp AL 2017. An Updated Infrageneric Classification of the Oaks: Review of Previous Taxonomic Schemes and Synthesis of Evolutionary Patterns. In: Gil-Pelegrín E, Peguero-Pina JJ, Sancho-Knapik D eds. *Oaks Physiological Ecology. Exploring the Functional Diversity of Genus Quercus L.* Cham: Springer International Publishing, 13-38.

Dray S, Dufour A-B. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software* **22**(4): 1 - 20.

Du FK, Qi M, Zhang YY, Petit RJ. 2022. Asymmetric character displacement in mixed oak stands. *New Phytol* **236**(3):1212-1224.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim

MS, Machol I, Lander ES, Aiden AP, et al. 2017. *De novo* assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**(6333): 92-95.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3**(1): 95-98.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**(5): 1792-1797.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**(1): 238.

Escandón M, Castillejo MÁ, Jorrín-Novo JV, Rey M-D. 2021. Molecular research on stress responses in *Quercus* spp.: from classical biochemistry to systems biology through Omics Analysis. *Forests* **12**(3): 364.

Fang J. 2014. Bioavailability of anthocyanins. *Drug Metabolism Reviews* **46**(4): 508-520.

Fu R, Zhu Y, Liu Y, Feng Y, Lu RS, Li Y, Li P, Kremer A, Lascoux M, Chen J. 2022. Genome-wide analyses of introgression between two sympatric Asian oak species. *Nat Ecol Evol* **6**(7): 924-935.

Galperin MY, Kristensen DM, Makarova KS, Wolf YI, Koonin EV. 2019. Microbial genome analysis: the COG approach. *Brief Bioinform* **20**(4): 1063-1070.

Gao YF, Zhao DH, Zhang JQ, Chen JS, Li JL, Weng Z, Rong LP. 2021. *De novo* transcriptome sequencing and anthocyanin metabolite analysis reveals leaf color of *Acer pseudosieboldianum* in autumn. *BMC Genomics* **22**(1): 383.

Gould KS, Jay-Allemand C, Logan BA, Baissac Y, Bidel LPR. 2018. When are foliar anthocyanins useful to plants? Re-evaluation of the photoprotection hypothesis using *Arabidopsis thaliana* mutants that differ in anthocyanin accumulation. *Environmental and Experimental Botany* **154**: 11-22.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**(1): R7.

Han B, Wang L, Xian Y, Xie XM, Li WQ, Zhao Y, Zhang RG, Qin X, Li DZ, Jia KH. 2022. A chromosome-level genome assembly of the Chinese cork oak (*Quercus variabilis*). *Front Plant Sci* **13**: 1001583.

Hu H-L, Zhang J-Y, Li Y-P, Xie L, Chen D-B, Li Q, Liu Y-Q, Hui S-R, Qin L. 2018. The complete chloroplast genome of the daimyo oak, *Quercus dentata* Thunb. *Conservation Genetics Resources* **11**(4): 409-411.

Huang Y, Xiao L, Zhang Z, Zhang R, Wang Z, Huang C, Huang R, Luan Y, Fan T, Wang J, et al. 2019. The genomes of pecan and Chinese hickory provide insights into Carya evolution and nut nutrition. *Gigascience* **8**(5): giz036.

Ishida TA, Hattori K, Sato H, Kimura MT. 2003. Differentiation and hybridization between *Quercus crispula* and *Q. dentata* (Fagaceae): insights from

morphological traits, amplified fragment length polymorphism markers, and leafminer composition. *Am J Bot* **90**(5): 769-776.

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**(7161): 463-467.

Jia HM, Jia HJ, Cai QL, Wang Y, Zhao HB, Yang WF, Wang GY, Li YH, Zhan DL, Shen YT, et al. 2019. The red bayberry genome and genetic basis of sex determination. *Plant Biotechnol J* **17**(2): 397-409.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**(1-4): 462-467.

Juvany M, Muller M, Munne-Bosch S. 2013. Photo-oxidative stress in emerging and senescing leaves: a mirror image? *J Exp Bot* **64**(11): 3087-3098.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**(1): 27-30.

Kim HS, Lee BY, Won EJ, Han J, Hwang DS, Park HG, Lee JS. 2015. Identification of xenobiotic biodegradation and metabolism-related genes in the copepod Tigriopus japonicus whole transcriptome analysis. *Mar Genomics* **24 Pt 3**: 207-208.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**(1): 559.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**(4): 357-359.

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* **30**(1): 325-327.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**(1): 323.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Li P, Li YJ, Zhang FJ, Zhang GZ, Jiang XY, Yu HM, Hou BK. 2017. The Arabidopsis UDP-glycosyltransferases *UGT79B2* and *UGT79B3*, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. *Plant J* **89**(1): 85-103.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, et al. 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res* **20**(2): 265-272.

Li W, Li X, Chao J, Zhang Z, Wang W, Guo Y. 2018. NAC family transcription factors in tobacco and their potential role in regulating leaf senescence. *Front Plant Sci* **9**: 1900.

Liu Y, Li Y, Song J, Zhang R, Yan Y, Wang Y, Du FK. 2018. Geometric morphometric analyses of leaf shapes in two sympatric Chinese oaks: *Quercus dentata* Thunberg and *Quercus aliena* Blume (Fagaceae). *Annals of Forest Science* **75**(4): 90.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**(12): 550.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**(16): 2878-2879.

Manos PS, Stanford AM. 2001. The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere. **162**(S6): S77-S93.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**(6): 764-770.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**(9): 1297-1303.

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, et al. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **452**(7190): 991-996.

Mishra B, Gupta DK, Pfenninger M, Hickler T, Langer E, Nam B, Paule J, Sharma R, Ulaszewski B, Warmbier J, et al. 2018. A reference genome of the European beech (*Fagus sylvatica* L.). *Gigascience* **7**(6).

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**(D1): D412-d419.

Morales D. 2021. Oak trees (*Quercus* spp.) as a source of extracts with biological activities: A narrative review. *Trends in Food Science & Technology* **109**: 116-125.

Nixon KC 2006. Global and Neotropical Distribution and Diversity of Oak (genus *Quercus*) and Oak Forests. In: Kappelle M ed. *Ecology and Conservation of Neotropical Montane Oak Forests*. Berlin, Heidelberg: Springer Berlin Heidelberg, 3-13.

Plomion C, Aury JM, Amselem J, Leroy T, Murat F, Duplessis S, Faye S, Francillonne N, Labadie K, Le Provost G, et al. 2018. Oak genome reveals facets of long lifespan. *Nat Plants* **4**(7): 440-452.

Poole RL. 2007. The TAIR database. *Methods Mol Biol* **406**: 179-212.

Potato Genome Sequencing C, Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, et al. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* **475**(7355): 189-195.

Pourcel L, Routaboul JM, Cheynier V, Lepiniec L, Debeaujon I. 2007. Flavonoid oxidation in plants: from biochemical properties to physiological functions.

Trends Plant Sci **12**(1): 29-36.

Ramírez-Valiente JA, Koehler K, Cavender-Bares J. 2015. Climatic origins predict variation in photoprotective leaf pigments in response to drought and low temperatures in live oaks (*Quercus series* Virentes). *Tree Physiol* **35**(5): 521-534.

Ramos AM, Usié A, Barbosa P, Barros PM, Capote T, Chaves I, Simões F, Abreu I, Carrasquinho I, Faro C, et al. 2018. The draft genome sequence of cork oak. *Scientific Data* **5**(1): 180069.

Renner SS, Zohner CM. 2019. The occurrence of red and yellow autumn leaves explained by regional differences in insolation and temperature. *New Phytol* **224**(4): 1464-1471.

Rushton PJ, Somssich IE, Ringler P, Shen QJ. 2010. WRKY transcription factors. *Trends Plant Sci* **15**(5): 247-258.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol* **1962**: 227-245.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**(1): 259.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**(11): 2498-2504.

Shin DH, Choi M, Kim K, Bang G, Cho M, Choi SB, Choi G, Park YI. 2013. HY5 regulates anthocyanin biosynthesis by inducing the transcriptional activation of the *MYB75/PAP1* transcription factor in Arabidopsis. *FEBS Lett* **587**(10): 1543-1547.

Şöhretoğlu D, Renda G. 2020. The polyphenolic profile of Oak (*Quercus*) species: a phytochemical and pharmacological overview. *Phytochemistry Reviews* **19**(6): 1379-1426.

Sork VL, Cokus SJ, Fitz-Gibbon ST, Zimin AV, Puiu D, Garcia JA, Gugger PF, Henriquez CL, Zhen Y, Lohmueller KE, et al. 2022. High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. *Nat Commun* **13**(1): 2047.

Sork VL, Fitz-Gibbon ST, Puiu D, Crepeau M, Gugger PF, Sherman R, Stevens K, Langley CH, Pellegrini M, Salzberg SL. 2016. First Draft Assembly and Annotation of the Genome of a California Endemic Oak *Quercus lobata* Née (Fagaceae). *G3 (Bethesda)* **6**(11): 3485-3495.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9): 1312-1313.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34**(Web Server issue): W435-439.

Tang H, Lyons E, Pedersen B, Schnable JC, Paterson AH, Freeling M. 2011. Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Bioinformatics* **12**(1): 102.

Tian F, Yang DC, Meng YQ, Jin J, Gao G. 2020. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res* **48**(D1): D1104-D1113.

Tomato Genome C. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**(7400): 635-641.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**(3): 562-578.

Ubi BE, Honda C, Bessho H, Kondo S, Wada M, Kobayashi S, Moriguchi T. 2006. Expression analysis of anthocyanin biosynthetic genes in apple skin: Effect of UV-B and temperature. *Plant Science* **170**(3): 571-578.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**(5): 737-746.

Vimolmangkang S, Han Y, Wei G, Korban SS. 2013. An apple MYB transcription factor, *MdMYB3*, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biology* **13**(1): 176.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**(14): 2202-2204.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**(11): e112963.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, 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**(7): e49.

Wen CH, Tsao NW, Wang SY, Chu FH. 2021. Color variation in young and senescent leaves of Formosan sweet gum (*Liquidambar formosana*) by the gene regulation of anthocyanidin biosynthesis. *Physiol Plant* **172**(3): 1750-1763.

Wheeler TJ, Eddy SR. 2013. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**(19): 2487-2489.

Woo HR, Kim HJ, Lim PO, Nam HG. 2019. Leaf Senescence: Systems and Dynamics Aspects. *Annu Rev Plant Biol* **70**: 347-376.

Wu C, Wang X, Wang H, Ciais P, Peñuelas J, Myneni RB, Desai AR, Gough CM, Gonsamo A, Black AT, et al. 2018. Contrasting responses of autumn-leaf senescence to daytime and night-time warming. *Nature Climate Change* **8**(12): 1092-1096.

Xing Y, Liu Y, Zhang Q, Nie X, Sun Y, Zhang Z, Li H, Fang K, Wang G, Huang H, et al. 2019. Hybrid de novo genome assembly of Chinese chestnut (*Castanea*

mollissima). *Gigascience* **8**(9): giz112.

Xu H, Zou Q, Yang G, Jiang S, Fang H, Wang Y, Zhang J, Zhang Z, Wang N, Chen X. 2020. *MdMYB6* regulates anthocyanin formation in apple both through direct inhibition of the biosynthesis pathway and through substrate removal. *Hortic Res* **7**(1): 72.

Xu W, Dubos C, Lepiniec L. 2015. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci* **20**(3): 176-185.

Xu Z, Rothstein SJ. 2018. ROS-Induced anthocyanin production provides feedback protection by scavenging ROS and maintaining photosynthetic capacity in Arabidopsis. *Plant Signal Behav* **13**(3): e1451708.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**(Web Server issue): W265-268.

Yang FS, Nie S, Liu H, Shi TL, Tian XC, Zhou SS, Bao YT, Jia KH, Guo JF, Zhao W, et al. 2020. Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nat Commun* **11**(1): 5269.

Yang X, Wang Z, Zhang L, Hao G, Liu J, Yang Y. 2020. A chromosome-level reference genome of the hornbeam, *Carpinus fangiana*. *Sci Data* **7**(1): 24.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**(8): 1586-1591.

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**(5565): 79-92.

Zhang S, Chen Y, Zhao L, Li C, Yu J, Li T, Yang W, Zhang S, Su H, Wang L. 2020. A novel NAC transcription factor, *MdNAC42*, regulates anthocyanin accumulation in red-fleshed apple by interacting with *MdMYB10*. *Tree Physiol* **40**(3): 413-423.

Zheng C, Wang Y, Ding Z, Zhao L. 2016. Global transcriptional analysis reveals the complex relationship between tea quality, leaf senescence and the responses to cold-drought combined stress in *Camellia sinensis*. *Frontiers in Plant Science* **7**: 1858.

Figure Legends

Fig. 1 Photograph and genome characterization of *Q. dentata*. **a** Photo of autumn leaves. **b** Synteny and distribution of genomic features. Concentric circles - from outermost to innermost - show, Assembled chromosomes, Gene density (green), Transposon density (red), Tandem repeat density (blue), GC content (pink). Lines in the center of the circle represent chromosome collinearities.

Fig. 2 Phylogenomics and genome evolution. **a** Phylogenetic tree, divergence times and gene family expansion and contraction among 18 plant species. Pie charts show the percentage of gene families that underwent expansion or contraction. The numbers above the node indicate the range of species divergence times. The nodes for dating calibration are noted with red color. **b** *K*s distribution of *C. fangiana*, *C. mollissima*, *F. sylvatica*, *Q. dentata-V. vinifera*, *Q. dentata*, *Q. lobata, Q. robur*, *V. vinifera*. **c** Chromosome collinearity analysis among *Q. dentata*, *Q. robur*, *Q. lobata*, *Q. variabilis*, *Q. acutissima*, *Q. mongolica*, and *F. sylvatica.* **d** Venn diagram showing the shared and unique gene families among *Q. dentata*, *Q. robur*, *Q. suber*, *Q. lobata*, *Q. variabilis*, *Q. acutissima*, *Q. mongolica*, and *F. sylvatica*.

Fig. 3 Flavonoid metabolic composition and gene expression during the leaf color changes of *Q. dentata*. **a** Leaf color change in *Q. dentata* (S1-S5) bar=1cm. **b** Relative contents of anthocyanins (Pelargonidin-3-O-glucoside, Cyanidin-3-O-glucoside, Cyanidin-3-O-arabinoside, Cyanidin-3,5-O-diglucoside, Delphinidin-3-O-glucoside), flavonols (Naringenin, Naringenin chalcone, Taxifolin, Dihydrokaempferol and Kaempferol), flavanols (Epicatechin, Catechin, Gallocatechin, Epigallocatechin, Epiafzelechin) at three leaf color stages. Data are presented as mean ± s.d. (Student's t-test, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$). **c** PCA distance biplot. The PCA biplot shows both, the PC scores of samples (here: three sampled leaves for each of the three leaf color change stages shown as dots) and the loadings of variables (here: the individual metabolite variables as vectors) in 2D space defined by PC1 and PC2. The PCA biplot shows both the PC scores of metabolites as dots and variable loadings as vectors. when

two vectors form a large angle, then the variables these vectors represent are negatively correlated; when they form a small angle, the two variables they represent are positively correlated. **d** Cluster dendrogramm of transcriptomes. Anthocyanins are noted in magenta color.

Fig. 4 Flavonoids biosynthetic pathways and gene expression in *Q. dentata.* Gene expression levels (log_{10}(FPKM+1)) at different leaf color change stages (S1-S5) are represented by color grading. Gene IDs are shown on the right side of each heatmap. Genes identified in tandem clusters are marked in turquois color. The leaf anthocyanins and flavonoid glycosides are marked and framed. PAL, phenylalanine ammonia-lyase, C4H, cinnamate-4-hydroxylase, 4CL, 4-coumarate CoA ligase 4, CHS, chalcone synthase, CHI, chalcone isomerase, F3H, flavanone 3-hydroxylase, F3'H, flavonoid 3'-hydroxylase, F3'5'H, flavonoid 3'5'-hydroxylase, FLS, flavonol synthase, DFR, dihydroflavonol 4-reductase, ANR, anthocyanidin reductase, LAR, leucoanthocyanidin reductase, ANS, anthocyanidin synthase, A3oGT, anthocyanidin-3-O-glucosyl transferase.

Fig. 5 Transcriptomic and metabolic correlation analysis across leaf color change stages (S1-S5) in *Q. dentata.* **a** Dendrogram showing co-expression modules identified by WGCNA across leaf color changes. **b** Heat map exposing the correlations between gene expression modules and flavonoid metabolites. Each of the 11 rows corresponds to a specific module (ME) indicated by a different color. Each of the 19 columns corresponds to a flavonoid compound. Red color indicates a positive correlation. Green color indicates a negative correlation. **c** Expression trend of each module across the five time stages (S1-S5). Anthocyanins are noted in magenta color.

Fig. 6 Identification of key TFs and transcriptional regulatory network associated with leaf color transition. **a** DNA binding sites of promoter sequences within 2.0 kb upstream of *PAO*, *NYC1*, *SGR*, *CHS*, *ANS* and *F3'H* genes. **b** The resolved gene regulatory network of key flavonoid biosynthesis and chlorophyll degradation in *Q. dentata.* Red and green hexagons represent structural genes involved in flavonoid

synthesis and chlorophyll degradation, respectively. Blue squares represent hub genes involved in flavonoid metabolism and chlorophyll degradation within the brown module. Colored circles represent different families of transcription factors within the same module. **c** Gene expression pattern (log10(FPKM+1)) heat map of significantly differentially expressed genes of different TFs classes during leaf color change (S1-S5).

Fig. 7 Sub-network of chlorophyll degradation and flavonoids biosynthesis pathways. **a** Sub-network of *QdNYC1*, *QdSGR* and *QdPAO*. **b** Sub-network of *QdANS* and *QdCHS* genes. **c** DNA binding sites of promoter sequences located 2.0 kb upstream of *NAC*, *MYB* and *WRKY*. **d** Resolved hierarchical regulation for *CHS*, *ANS* and chlorophyll catabolic genes (CCGs). **e** Yeast two-hybrid assay showing interaction between pGADT7- *NAC QD08G038820* and pGBKT7-*MYB QD01G020890*. **f** Yeast one-hybrid assays verifying interactions between pGADT7-*NAC QD08G038820* and upstream 2.0 kb promoter fragment of *SGR QD12G029700*; pGADT7-*MYB QD01G020890* and upstream 2.0 kb promoter fragment of *CHS QD06G044950*.

Supporting Information

Table S2 Statistics of the final assembly.

Table S3 Summary of BUSCO evaluation results for the final assembly.

Table S4 Statistics of ONT long reads and quality control of the genome assembly of *Q. dentata*.

Table S5 Summary of the annotated repeat elements in the genome of *Q. dentata*.

Table S6 Statistics of the predicted genes in the genome of *Q. dentata*.

Table S7 Summary of noncoding RNA genes annotated in the *Q. dentata* genome.

Table S8 Location and sequence of rRNA genes annotated in the genome of *Q. dentata*.

Table S9 Number of ribosomal RNA genes annotated in the *Quercus species*.

Table S10. Statistics of annotated coding genes.

Table S11 Homologous gene families identified in 7 *Quercus* species.

Table S12 Summary of intra- and inter- genomic collinearity among oak species.

Table S13 List of orthologous gene pairs identified among *Quercus* and *Fagus* species.

Table S14 Summary of 188 flavonoid metabolites detected based on UPLC-MS/MS.

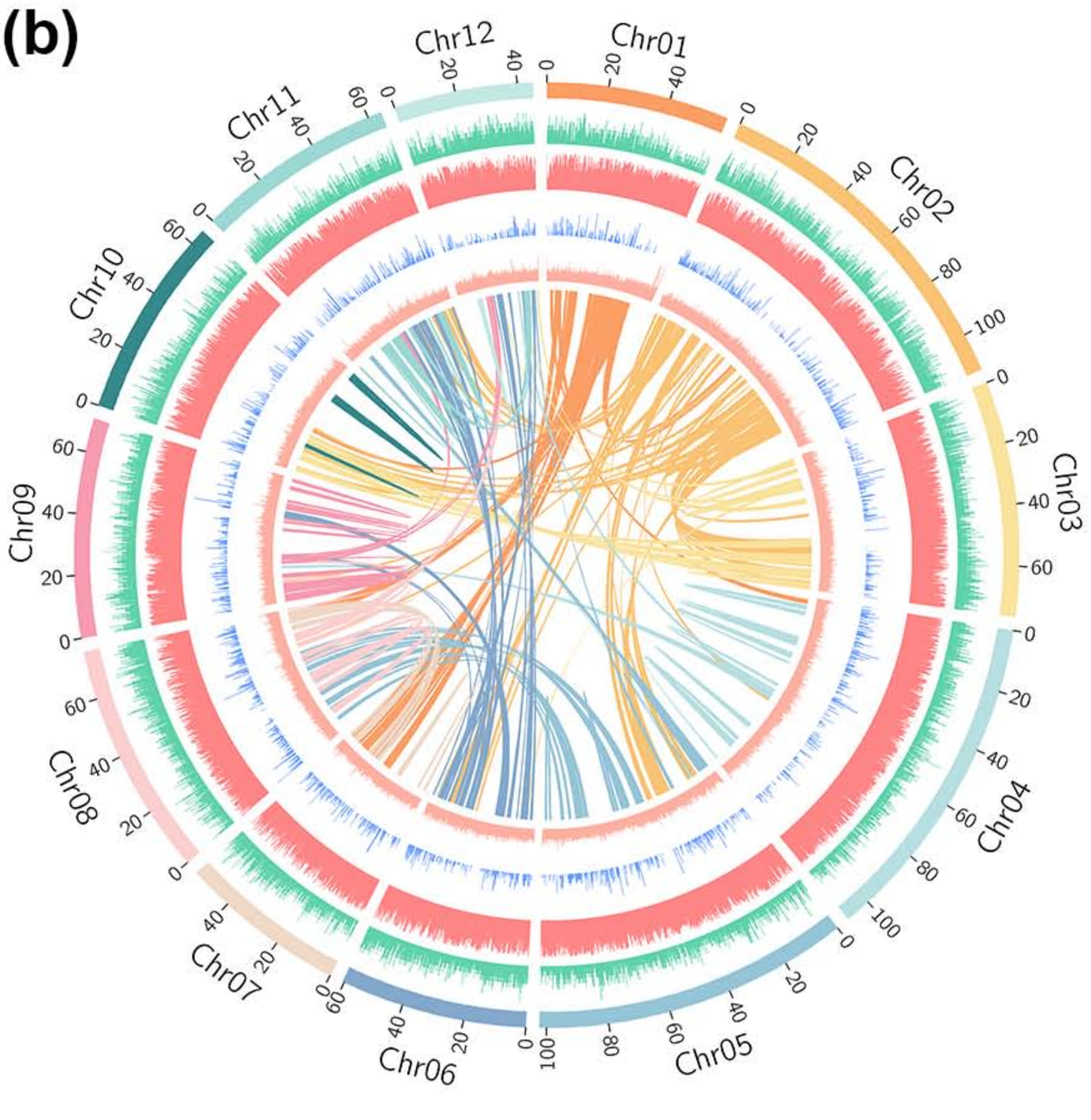
Table S15 Statistics of qualities of RNA-seq data.

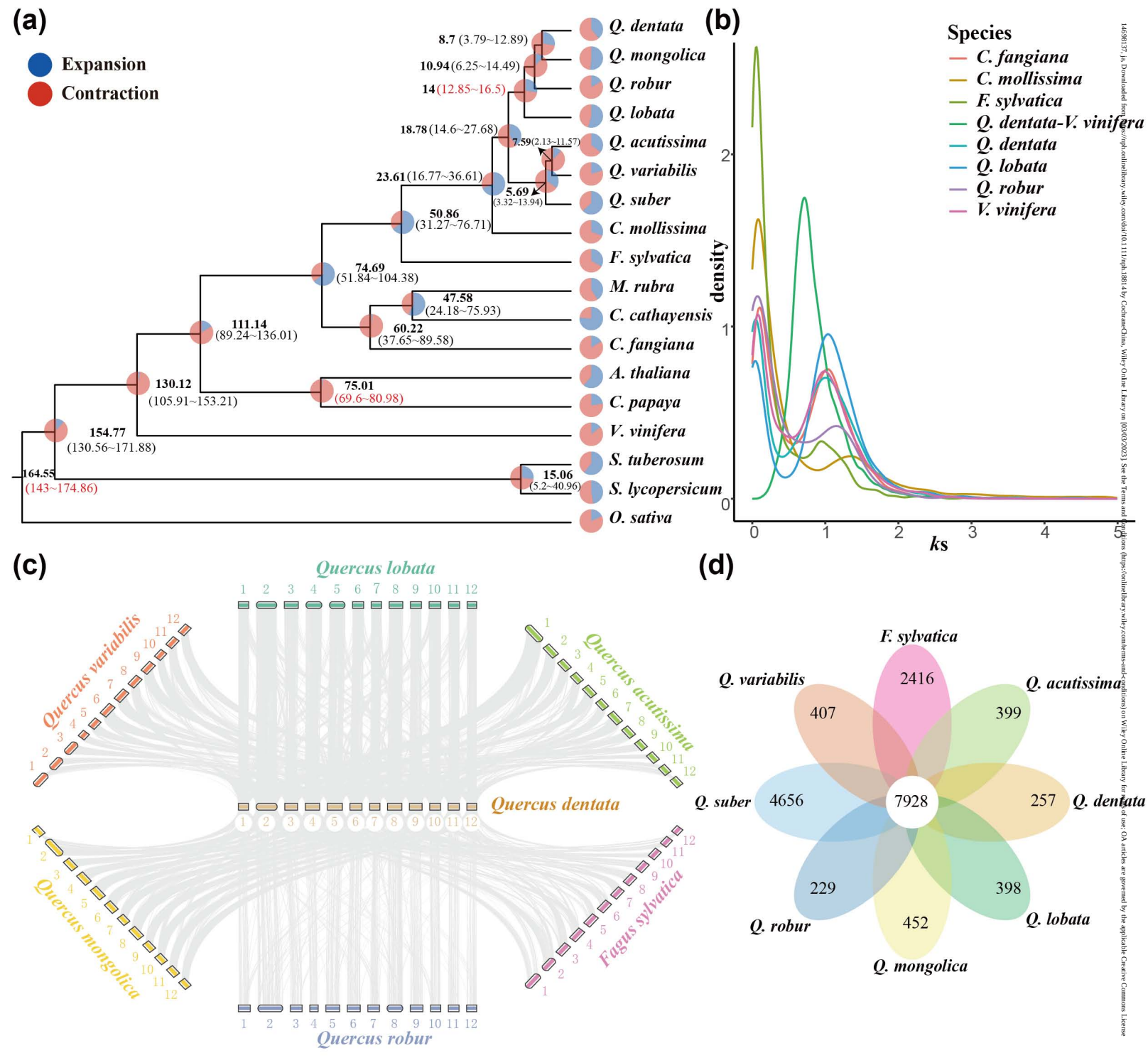
Table S16 Expression levels of chlorophyll catabolic genes (CCGs) and flavonoid biosynthesis genes.

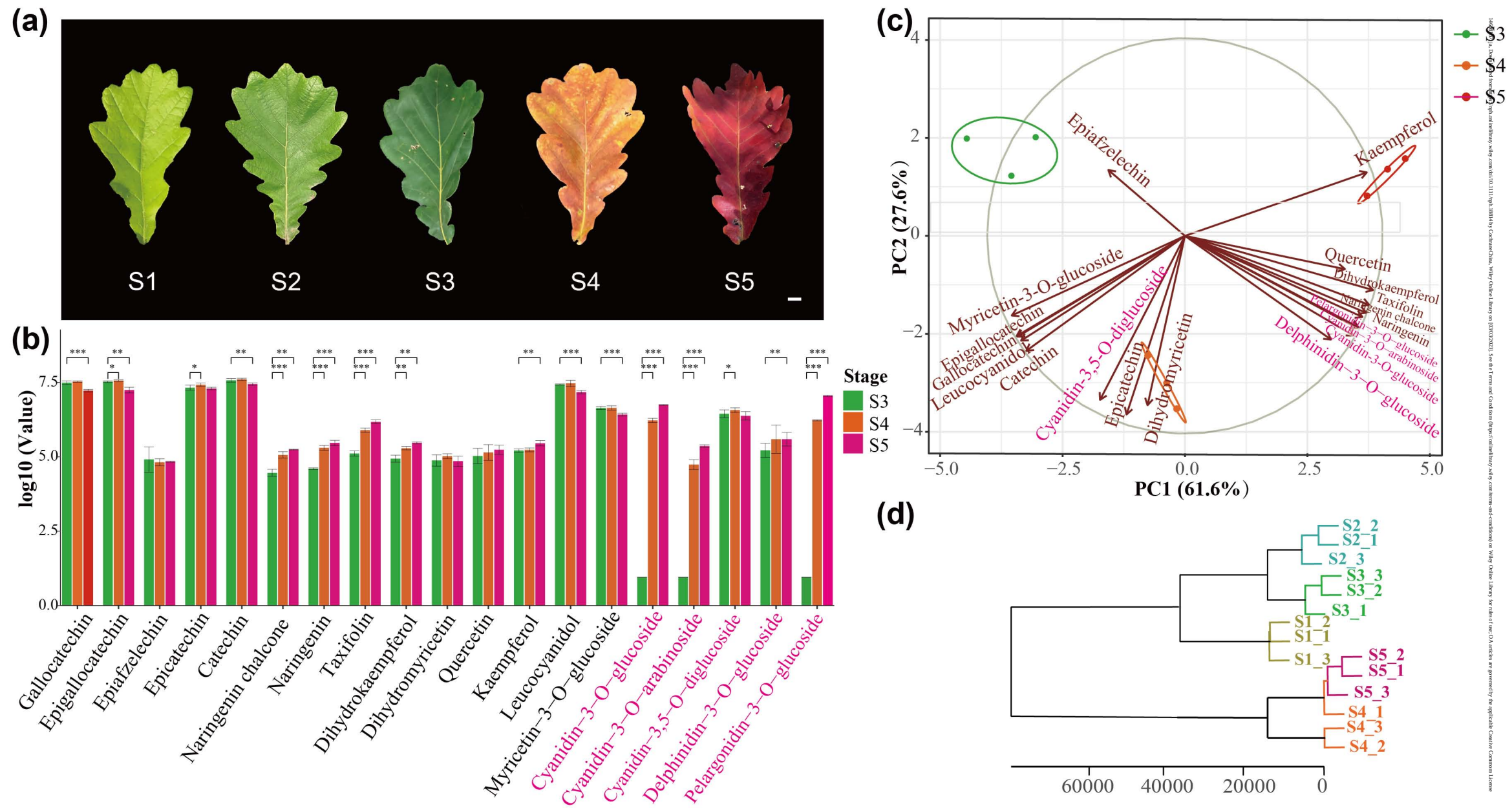
Table S17 Statistics of significantly differential metabolites among three stages.

(a)

(b)






(a)
S1
S2
S3
S4
S5
(b)
log10 (Value)
Stage
S3
S4
S5
Gallocatechin
Epigallocatechin
Epiafzelechin
Epicatechin
Catechin
Naringenin chalcone
Naringenin
Taxifolin
Dihydrokaempferol
Dihydromyricetin
Quercetin
Kaempferol
Leucocyanidol
Myricetin−3−O−glucoside
Cyanidin−3−O−glucoside
Cyanidin−3−O−arabinoside
Cyanidin−3,5−O−diglucoside
Delphinidin−3−O−glucoside
Pelargonidin−3−O−glucoside
(c)
PC2 (27.6%)
PC1 (61.6%)
Epiafzelechin
Kaempferol
Myricetin-3-O-glucoside
Quercetin
Dihydrokaempferol
Taxifolin
Naringenin chalcone
Naringenin
Pelargonidin-3-O-glucoside
Cyanidin-3-O-arabinoside
Cyanidin-3-O-glucoside
Delphinidin-3-O-glucoside
Epigallocatechin
Gallocatechin
Leucocyanidol
Catechin
Cyanidin-3,5-O-diglucoside
Epicatechin
Dihydromyricetin
(d)
S2_2
S2_1
S2_3
S3_3
S3_2
S3_1
S1_2
S1_1
S1_3
S5_2
S5_1
S5_3
S4_1
S4_3
S4_2
60000
40000
20000
0