

Chromosome-level genome assembly of *Quercus variabilis* provides insights into the molecular mechanism of cork thickness

Ermei Chang, Wei Guo, Jiahui Chen, Jin Zhang, Zirui Jia, Timothy J. Tschaplinski, Xiaohan Yang, Zeping Jiang, Jianfeng Liu



PII: S0168-9452(23)00291-1

DOI: <https://doi.org/10.1016/j.plantsci.2023.111874>

Reference: PSL111874

To appear in: *Plant Science*

Received date: 10 July 2023

Revised date: 3 September 2023

Accepted date: 18 September 2023

Please cite this article as: Ermei Chang, Wei Guo, Jiahui Chen, Jin Zhang, Zirui Jia, Timothy J. Tschaplinski, Xiaohan Yang, Zeping Jiang and Jianfeng Liu, Chromosome-level genome assembly of *Quercus variabilis* provides insights into the molecular mechanism of cork thickness, *Plant Science*, (2023)
doi:<https://doi.org/10.1016/j.plantsci.2023.111874>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Published by Elsevier.

Special Issue ‘Tree Biology’

Chromosome-level genome assembly of *Quercus variabilis* provides insights into the molecular mechanism of cork thickness

Ermei Chang¹, Wei Guo², Jiahui Chen³, Jin Zhang⁴, Zirui Jia¹, Timothy J. Tschaplinski^{5,6}, Xiaohan Yang^{5,6}, Zeping Jiang^{7*} and Jianfeng Liu^{1*1}

¹ State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of the National Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 10091, China

² Taishan Academy of Forestry Sciences, Taian, Shandong 271000, China

³ CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

⁴ State Key Laboratory of Subtropical Silviculture, College of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou, Zhejiang 311300, China

⁵ Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

⁶ The Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

⁷ Key Laboratory of Forest Ecology of National Forestry and Grassland Administration, Environment and Protection, Research Institute of Forest Ecology, Environment and Protection, Chinese Academy of Forestry, Beijing 100091, China

ABSTRACT

Quercus variabilis is a deciduous woody species with high ecological and economic value, and is a major source of cork in East Asia. Cork from thick softwood sheets have higher commercial value than those from thin sheets. It is extremely difficult to genetically improve *Q. variabilis* to produce high quality softwood due to the lack of genomic information. Here, we present a high-quality chromosomal genome assembly for *Q. variabilis* with length of 791,89 Mb and 54,606 predicted genes. Comparative analysis of protein sequences of *Q. variabilis* with 11 other species revealed that specific and expanded gene families were significantly enriched in the "fatty acid biosynthesis" pathway in *Q. variabilis*,

*Corresponding authors.

E-mail addresses: liujf@caf.ac.cn (J. Liu); jiangzp@caf.ac.cn (Z. Jiang)

¹ These authors contributed equally to this work.

which may contribute to the formation of its unique cork. Based on weighted correlation network analysis of time-course (i.e., five important developmental ages) gene expression data in thick-cork versus thin-cork genotypes of *Q. variabilis*, we identified one co-expression gene module associated with the thick-cork trait. Within this co-expression gene module, 10 hub genes were associated with suberin biosynthesis. Furthermore, we identified a total of 198 suberin biosynthesis-related new candidate genes that were up-regulated in trees with a thick cork layer relative to those with a thin cork layer. Also, we found that some genes related to cell expansion and cell division were highly expressed in trees with a thick cork layer. Collectively, our results revealed that two metabolic pathways (i.e., suberin biosynthesis, fatty acid biosynthesis), along with other genes involved in cell expansion, cell division, and transcriptional regulation, were associated with the thick-cork trait in *Q. variabilis*, providing insights into the molecular basis of cork development and knowledge for informing genetic improvement of cork thickness in *Q. variabilis* and closely related species.

Keywords: *Quercus variabilis*, chromosome-level genome, cork thickness, suberin biosynthesis

1. Introduction

Over 450 oak species (*Quercus* spp.; Fagaceae) are distributed worldwide, with most inhabiting the Northern Hemisphere. Oak species generally act as dominant species in temperate, subtropical, and tropical forests [1], and have important economic, ecological and cultural value, e.g., timber production, biodiversity conservation, serve as a major carbon sink, and in forest recreation. China is one of the major zones of distribution of oak species around the world, with more than 100 species occurring from northern to southern China (Flora of China, 1999). Among all oak species in China, *Q. variabilis* is one of the most abundant tree species (~1.428 billion trees) and constitutes 0.75% of all tree species [2,3]. *Q. variabilis* is renowned for its high-quality wood and high-value cork material [4]. Therefore, an accurate and complete chromosome-scale *Q. variabilis* genome sequence will further enhance the conservation and genetic improvement of this critical tree species.

Cork is a renewable continuous layer of suberized cells in the external bark of the periderm of tree stems and branches. It is mainly supplied by two cork oak species, *Quercus suber* (evergreen) in the Mediterranean region and *Q. variabilis* (deciduous) in East Asia [5]. As a nonconductive and chemically stable material, cork has been widely used in wine stoppers, flooring, decorative wall-paper, and as insulation against fire and cold in space rockets [6]. The thickness of cork is a key feature in determining its value [6]. Extensive studies have been conducted on the physical structure, chemical composition, thickness, and the molecular mechanism of periderm biosynthesis in the cork oak species *Q. suber* [6,7], which has a long history of breeding and cultivation. However, there have only a few studies on the basic physical or chemical properties in the cork from *Q. variabilis* [4], let alone large-scale directive breeding, cultivation and utilization. Consequently, in comparison with the mass of cork obtained from *Q. suber*, the commercial deployment of *Q. variabilis* cork is limited due to its low quality. Therefore, genetic improvement of *Q. variabilis* is critically required to obtain high quality cork.

Cork (phellem) cells are predominantly composed of suberin, lignin, and polysaccharides [8]. The suberin content determines cork qualities, such as permeability and elasticity [9]. Suberin is a heteropolymer formed by acyl-lipid, phenylpropanoids, isopropanoids and favonoids compounds, and has hydrophobic properties and high auto-ignition resistance [7]. Meanwhile, suberin act as a

waterproof barrier to prevent nonstomatal water loss, making the trees more resistant against drought stress and helping the trees maintain moisture and prevent fires [6]. Suberin has antibacterial, antioxidant, and stress resistant effects, which can help trees resist external wounding and other environmental pressure [6]. Recent studies revealed that CYP86A1 (cytochrome p450 86A1), GPAT (glycerol-3-phosphate acyltransferase), KCS (3-ketoacyl-CoA synthase), and ABCG (ABC transporter G family) encode key enzymes involved in the biosynthesis of suberin monomers in the root and seed coat of *Arabidopsis thaliana* and in the periderm of *Solanum tuberosum* [11–15]. Phytohormones and transcription factors (TFs) also play key roles in suberin biosynthesis, including QsMYB1 in *Q. suber*, AtWRKY9 in *A. thaliana*, and StANAC103 in *S. tuberosum* [16–19]. In the thick cork of *Q. suber*, genes involved in the suberin monomers biosynthesis pathway, including CYP86A1, KCS, and GPAT were highly expressed [6,7,20]. In contrast, genes involved in phenol biosynthesis and biotic and abiotic resistance were highly expressed in thin cork layers [6,7,20]. Suberin monomer content and composition in cell walls vary among species, tissues, and stages of plant growth [21,22]. To produce more cork with the highest economic value, determination of the genetic basis of suberin biosynthesis and its manipulation in *Q. variabilis* is of critical importance.

Q. variabilis is a good ornamental tree species for greening, and is suitable for reforestation of barren hills and wastelands, and also known as Chinese cork oak. Improving cork quality by increasing its suberin content and thickness has always been an important goal of *Q. variabilis* breeding programs. A fully sequenced and annotated genome is essential for clarifying the genetic basis underlying key traits in *Quercus* species, including cork characteristics [23,24], wood properties [25–27], and life expectancy [28]. Improving cork quality by increasing its suberin content and thickness has always been an important goal of *Q. variabilis* and *Q. suber* breeding programs. In this study, we assembled a high-quality chromosome-level genome of *Q. variabilis* to be deployed as a reference in the identification and characterization of key gene families involved in the biosynthesis of cork and their manipulation to improve cork quality.

2. Materials and methods

2.1. Estimation of genome size

Cork cambium of *Q. variabilis* was collected from a 9-year-old tree located at the Chinese Academy of Forestry (CAF), Beijing, China, and immediately preserved in liquid nitrogen. Total genomic DNA

was extracted using the CTAB method [29]. Small DNA fragments (~ 350 bp) were obtained by ultrasonication and further processed for DNA end-repair, A-tailing, adapter ligation, and PCR amplification to construct the genomic library, which was fixed on a microarray chip by bridge PCR amplification. Libraries were 150 bp paired-end sequencing on an Illumina NovaSeq 6000 system. Genome size, heterozygosity, and repetitive sequence rate were calculated using a distribution map created on the basis of the library data (k -mer length of 21). The best k -mer length was used for the k -mer count analysis, which was performed using Jellyfish 2.1.4 (<https://github.com/gmarcais/Jellyfish>). After converting the k -mer counts into a histogram format, the k -mer distribution was analyzed to estimate the genome size and heterozygosity.

2.2. Genome sequencing and de novo assembly

The *Q. variabilis* genome was sequenced via a third-generation sequencing platform PacBio Sequel II. DNA fragments, obtained using g-TUBE, were used to construct a PacBio library in the following steps: DNA damage repair, DNA end repair, dumbbell-type adapter ligation, exonuclease digestion, target DNA fragment screening using BluePippin, DNA end repair, A-tailing, adapter ligation, and PCR amplification. Reads obtained from multiple sequencing cycles in a ZWM well were self-corrected to obtain highly accurate CCS reads. Genome sequences were assembled using the hifiasm 0.12 [30]. CEGMA v2.5 [31] and BUSCO v5 [32] were used to evaluate the completeness of the genome assembly.

2.3. Hi-C sequencing and assembly

Young leaves from the same *Q. variabilis* tree were sampled, treated with formaldehyde, and the resulting cross-linked nuclear DNA was extracted, digested with *Hind*III, purified, and sheared into 300–700 bp fragments via ultrasonication. After end repair, the DNA fragments were used to construct a Hi-C library. Clean data were obtained by eliminating adapters and low-quality paired-end reads. HiC-Pro 2.10.0 was used to identify and keep valid interacting pairs [33]. Correction of scaffold errors was performed prior to chromosomal assembly by dividing into segments of ~ 50 kb and mapped to the segments by using BWA 0.7.10-r789 software. The uniquely mapped data were retained to assemble the genome using LACHESIS software by adjusting the following parameters; CLUSTER_MIN_RE_SITES=100; CLUSTER_MAX_LINK_DENSITY=2;

ORDER_MIN_N_RES_IN_TRUNK=107; and ORDER_MIN_N_RES_IN_SHREDS=105. Finally, a chromosome-level genome was assembled after manual inspection and mapping.

2.4. Transcriptome sequencing

Seeds, roots, terminal buds, young leaves, and young stems were collected from 5-year-old trees, whereas vascular cambium and phellogen were collected from 9-, 17-, 29-, 38-, and 51-year-old *Q. variabilis* trees for RNA sequencing (RNA-seq) in support of genome annotation. Total RNA was extracted using an RNAPrep Pure Plant Kit (Tiangen, Beijing, China). RNA-seq libraries were pair-edited and sequenced on the Illumina NovaSeq® 6000 platform. Raw reads were removed and only high-quality reads were further aligned against the reference genome by deploying TopHat2 2.0.7 software [34]. Additionally, high-quality reads were also mapped against the *de-novo* *Q. variabilis* transcriptome assembly using Cufflinks (version 1.3.0) [35].

2.5. Repeat annotation

Repeat gene families found in the genome assemblies of *Q. variabilis* were identified *de novo* using the software package RepeatModeler (version 2.0.1) [36]. The repeating sequences were classified using RepeatClassifier [36] and other databases, including Repbase 19.06, REXdb 3.0, and Dfam 3.2. LTR_retriever 2.8 [37], were used to predict *de-novo* long terminal sequence repeats on the basis of the predication of LTRharvest 1.5.9 [38] and LTR_FINDER 1.1 [39]. To predict TE sequences, RepeatMasker 4.1.0 [40] was used to identify homologous sequences in a nonredundant species-specific TE library comprised of both predicted and only nonredundant sequences.

2.6. Protein-coding gene prediction and functional annotation

Annotation of protein-coding sequences was conducted through a combination of *de novo* prediction, homology-based prediction, and transcript assembly prediction. Augustus (v 2.4) [41] and SNAP (2006-07-28) [42] were used for *de novo* gene model prediction. Homology (*Q. suber*, *Q. robur*, *Q. lobata*, *P. trichocarpa*, and *A. thaliana*) was predicted *via* GeMoMa 1.7. RNA-seq data were mapped to the reference genome using HISAT 2.0.4 [43] and assembled using StringTie 1.2.3 [44]. The gene model was subsequently predicted using GeneMarkS-T 5.1 [45] according to the assembled transcripts. PASA 2.0.2 was employed to predict genes based on the unigenes that were assembled by Trinity 2.11. Finally, predicted gene models were combined using EVM 1.1.1 [46], and further updated using PASA

and functionally annotated using the following databases; National Center for Biotechnology Information (NCBI-NR), SWISS-PROT, Gene ontology (GO), Cluster of Orthologous Group (COG), Eukaryotic Orthologous Group (KOG), Evolutionary Genealogy of Genes: Non-supervised Orthologous Group (eggNOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG).

2.7. Noncoding RNA prediction

Putative pseudogenes were retrieved *via* a whole-genome exploration using GenBlastA 1.0.4 [47] after the predicted functional genes were masked. The pseudogenes were further analyzed by searching for immature and frame-shift mutations using GeneWise 2.4.1 [48]. tRNA sequences were identified using tRNAscan-SE (v1.3.1) [49], and rRNA sequences were identified using barnap 0.9 (<https://github.com/tseemann/barnap>) by screening the Rfam 12.0 database [50,51]. miRNA sequences were identified using miRBase databases [52], and snoRNAs and snRNAs were identified using Infernal 1.1 [50] by exploring the Rfam 12.0 database.

2.8. Gene family expansion and contraction analysis

To identify orthologs, protein sequences of *Q. variabilis* and the other 11 plant species, including *Q. suber*, *B. pendula*, *Cinnamomum micranthum*, *Eucommia ulmoides*, *Punica granatum*, *Vitis vinifera*, *Populus trichocarpa*, *Amborella trichopoda*, *A. thaliana*, *Q. robur*, and *Q. lobata*, were analyzed by using Orthofinder (v2.4) with the diamond alignment method with an *e*-value cutoff at 0.001 [53]. All identified orthogroups were segregated into different families by exploring the PANTHER (v15) database (Table S1). Expansions and contractions of gene families in all 12 species were analyzed by using CAFE (v4.2) [54] and functionally characterized by KEGG pathway and GO enrichment analyses at a significance level of $P < 0.05$.

2.9. Phylogenetic analysis and divergence time estimation

For phylogenetic analysis, JTT+F+I+G4 was selected as an optimal sequence evolution model using ModelFinder [55] and the phylogenetic tree was constructed by the maximum likelihood method using IQ-TREE (v1.6.11) at 1000 bootstrap values [56]. Divergence times were calibrated between different pairs of plant species using the TimeTree database and then calculated using MCMCTree in PAML (v4.9i) [57] as *A. trichopoda* and *Q. robur* (173–199 Mya), *C. micranthum* and *V. vinifera* (115–308 Mya), *A. thaliana* and *B. pendula* (98–117 Mya), *P. trichocarpa* and *E. ulmoides* (111–131 Mya), and

Q. robur and *B. pendula* (51–87 Mya). The iteration parameters for the Markov chain were -burnin 5,000,000, -sampfreq 30, and -nsample 5,000,000. MCMCTreeR (v1.1) was used for graphical presentation of divergence.

2.10. Gene cluster analysis

Protein sequences encoded by single-copy genes in *B. pendula*, *Q. lobata*, *Q. robur*, *Q. suber*, and *Q. variabilis* were aligned using the MAFFT software (localpair-maxiterate 1000), whereas the corresponding nucleotide sequences were aligned using PAL2NAL. The CodeML program in the PAML package was used to detect genes under positive selection. Briefly, the branch-site model and multiple sequence alignment were adjusted according to the F3X4 model of codon frequencies. Similarly, the likelihood ratio was tested using the chi2 program in the PAML package by selecting Model A (foreground branch positive selection; $\omega > 1$) and the null Model at a significance threshold level of $P < 0.01$. Finally, positively selected genes during the evolutionary process were identified by the Bayesian method in PAML ($\omega > 0.95$).

2.11. Synteny analysis

To identify gene pairs between two species, the DIAMOND (v0.9.29.130) program was used to align gene sequences at a 50% identity threshold, e-value cutoff $< 1e^{-5}$ and C score > 0.5 (C score was filtered with the help of JCVI software) [58]. Similarly, adjacent gene pairs located on chromosomes were identified with deployment of MCScanX software, and finally, collinear gene blocks were retrieved.

2.12. Analyses of whole-genome duplication (WGD) events

WGD events were identified using the WGD (v1.1.1) software at a fourfold synonymous third-codon transversion rate (4DTv) and synonymous substitution rate (Ks) [59]. Subsequently, the transition-to-transversion ratio of each homolog was also calculated by 4DTv analysis using scripts (<https://github.com/JinfengChen/Scripts>) and corrected with the application of the HKY model. WGD events were assigned date of duplication with the help of the local multi-rate clock $T = Ks/2r$ ($r = 7 \times 10^{-9}$), here “r” refers to the rate of substitution of nucleotides.

2.13. Transcriptomic data analysis

Samples were collected in mid-June, 2021, when trees were at their maximum growth period [60]. A

group of cork oaks including individuals from 5 to 80 years old with a consistent growth environment were selected. Tree-ring cores of them were taken at a height of 1.3 m with an increment borer and analyzed to estimate the tree ages using dendrochronological procedures. Cork thickness was measured at a height of 1.3 m with a bark gauge (Haglöf Sweden AB, Långsele, Sweden). There are totally five age groups (9-, 17-, 29-, 38-, and 51-year-old) and each age group was divided into two groups significantly differed in cork thickness, extreme thin group and extreme thick group ($P < 0.01$). Each group was collected from three individual trees. The pale brown tissue (phellogen) was carefully harvested at a height of 1.3 m by scraping with RNase-free blades according to the previous report Teixeira et al [6].

A total of thirty RNA-seq libraries were constructed and sequenced as described above. FPKM (reads per kilobase per million) values were calculated using HTSeq (v0.6.1). Differentially expressed genes (DEGs) were identified using the R package DESeq2 at a significance level of $P < 0.05$, and functionally annotated in GO enrichment and KEGG pathways using the R package Goseq and KOBAS 2.0 web server, respectively.

For gene coexpression network construction, each gene in any functional module (Eigengenes value $E^{(q)}$) was annotated using the R package WGCNA (Weighted Correlation Network Analysis). To identify conserved functional domains, protein sequences of key gene families were blasted against different protein databases, including Pfam, SMART, and NCBI by using BLASTP and HMMER at a default cutoff e-value of $1e^{-5}$ (Table S2). Finally, a phylogenetic tree was constructed in MEGA 7.0 with the maximum likelihood method and 1000 bootstrap replicates.

2.14. Microscopic examination

Cork samples smaller than $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$ were prepared by carving the cork granules with a sharp razor. Samples were dehydrated before mounting on stubs (ProSciTech, Australia) coated with gold-palladium for 3 min at 20 mA using an E5100 sputtering system (Polaron, USA). Transverse, radial, and tangential sections were observed at $50 \sim 1000 \times$ magnification under an S-4800 scanning electron microscope (Hitachi, Tokyo, Japan). Using ImageJ software, 500 cells from cork layers that varied in terms of thickness were analyzed to measure cell wall thickness on the basis of the prism base edge and area.

2.15. Determination of physiological changes

Air dried cork samples were ground and filtered through a sieve (40 mesh). A 1.5 g sample was accurately weighed for reflux and boiling using 250 ml methanol solution containing NaOCH₃. Then, the suberin content was determined as described earlier [4]. The subsequent determinations of total lignin were made in the suberin-free material as reported previously [4]. Fatty acid content was measured with an assay kit according to the manual (Cominbio, Suzhou, China). All the chemical results were calculated and reported as percentages of the initial material.

2.16. Quantification of hormones

Fresh phellogen samples were ground in liquid nitrogen. The plant hormone IAA was extracted using isopropanol/water/hydrochloric acid and quantified using a UPLC-MS/MS platform (Agilent 1290 Infinity II UPLC system equipped with the AB Sciex QTrap 6500+ mass spectrometer), calibrated with pure reference compounds and using inositol as the internal control. All analyses were performed with three biological replicates.

3. Results

3.1. Genome sequencing, assembly, and annotation

Cork samples (Fig. 1a–c) of *Q. variabilis* were sequenced by using the next generation Illumina NovaSeq 6000 platform. Incomplete reads were discarded, and 49.30 Gb (~ 69× coverage) of clean short reads were further analyzed at a Q30 score of 91.39%, and the *k*-mer results revealed that the genome size of *Q. variabilis* was 715.17 Mb with 1.97% heterozygosity (Fig. S1 and Table S3) and 45.84% of repeated sequences (Table S4). Furthermore, the third generation long read data from the PacBio Sequel System were generated at approximately 37.73 Gb (~ 45 × coverage) CCS reads with an average read length of 13,930 bp and N50 read length of 13,814 bp (Table S5). The primary assembly with the total genome contig sequences was retrieved after splicing as 833.54 Mb, with 645 contigs and a contig N50 of 18.15 Mb (Table S6). The Hi-C seq platform generated 110.36 Gb reads, with 794.52 Mb reads distributed and anchored on 12 pseudochromosomes (Fig. 1d), and the syntenic order and direction of 791.89 Mb (99.67%) reads determined (Fig. S2–3 and Table S7). The final chromosome-level genome assembly with scaffold N50 (63.37 Mb) and contig N50 (17.10 Mb), accounts for 35.56% of the total GC content (Fig. 1d, Table 1, and S8).

To assess the quality and completeness of the above assembled genome, three of the following indexes were used. Assessment of quality completeness revealed a 99.38% mapping rate of Illumina paired-end reads to the assembled genome (Table S9). The core eukaryotic gene mapping approach (CEGMA) analysis detected 96.07% of the core eukaryotic genes in the *Q. variabilis* genome (Table S10). Finally, 91.82% of the single-copy genes were fully annotated by BUSCO analysis (Table S11). We propose the following additional criteria (Table S12): 4.09% false duplications, 79.09% kmer completeness, and 82% of transcripts from the same species mappable. Suggestions of and links to standard tools for measuring these metrics are provided on the EBP website [61]. These results indicated that a high-quality assembled genome for *Q. variabilis* was attained.

Transposable elements occupied approximately 49.27% (396.10 Mb) of the genome, with 5,836 full-length long terminal repeat retrotransposons retrieved, and the shared tandem repeats were 7.87% (63.31 Mb) of the genome (Fig. 1d and Table S13). Homology and transcriptomic analyses revealed 54,606 protein encoding genes with an average transcript size of 4,007.40 bp, coding sequence size of 1,294.18 bp, exon size of 1,489.55 bp, and intron size of 2,517.86 bp (Table 1, S14, and Fig. S4). The completeness assessment identified 1,586 (98.27%) BUSCOs in all annotated proteins. The BLASTN algorithm and tRNAscan-SE 1.3.1 program revealed 158 microRNAs, 4,215 rRNAs, 438 pseudogenes, and 819 tRNAs (Table S15). In total, 98.23% (53,639) of the genes were annotated according to at least one database, covering 93.82% in TrEMBL, 61.13% in SWISS-PROT, and 98.09% in the NCBI Non-Redundant (NR) database (Table S16).

3.2. Identification of lineage-specific gene family expansion

To investigate the relationship between gene families and specific *Q. variabilis* traits, we clustered the orthologs of *Q. variabilis* and 11 other sequenced plant species. In total, 42,301 gene families were identified, out of which 2,480 were common among all 12 species, and 1,364 were specifically expressed in *Q. variabilis* (Fig. S5). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis indicated significant enrichment in fatty acid biosynthesis among *Q. variabilis* specific gene families ($P < 0.05$) (Table S17). In the *Q. variabilis* genome, a total of 475 expanded gene families were identified with KEGG pathway enrichment in ‘monoterpenoid biosynthesis’, ‘galactose metabolism’, and ‘fatty acid biosynthesis’ (Table S18).

A phylogenetic tree comprised of 474 single-copy genes selected from 12 different species revealed that *Q. variabilis* is closely related to *Q. suber*, with a divergence time of 6–33 million years ago (Mya) (Fig. 2a). Among these gene families, 7,701 orthogroups are shared by several species, including *Q. suber*, *Q. lobata*, *Q. robur*, and *Betula pendula* closely related to *Q. variabilis*, whereas 1,657 orthogroups are unique to *Q. variabilis* (Fig. 2b).

3.3. Synteny analysis and Whole genome duplication analysis

To confirm the phylogenetic position of *Q. variabilis*, the timing of key evolutionary changes in the *Q. variabilis* genome was determined. Specifically, during the Ks distribution analysis, the peak was centered at Ks values of 0.0480 and 1.52 (Fig. 2c), indicating that the *Q. variabilis* genome was altered by a recent burst of local gene duplications and that the ancient γ whole-genome triplication event occurred approximately 1.37 and 43.4 Mya ago, respectively. These genomic changes were consistent with those in *Q. robur* [28] and *Q. mongolica* [27]. The analysis of the synteny between the *Q. variabilis* genome and the *Q. robur* and *Q. lobata* genomes revealed highly linear relationships between *Q. variabilis* and these two oak species (Fig. 2d).

3.4. Characteristics of the cell structure of thick and thin cork

Thickness of the cork is one of the main economic characteristics of *Q. variabilis*. Compared with the trees with a thin cork layer, those with a thick cork layer had thinner cell walls, higher prism heights, and larger periderm base edges (Fig. 3a–f, S7, and Table S19). Compared to younger trees, older *Q. variabilis* trees generally grow thicker cork cell walls and have shorter prism heights and periderm base edges. Our sampled trees demonstrated distinctive phenotypic variations between thick and thin cork types, providing avenues for subsequent omics research for trait improvement.

3.5. Physiological and biochemical factors affect the differences in thick and thin cork of *Q. variabilis*

To investigate the physiological differences of cork thickness, determination of their synthetic products was performed with different cork thicknesses. The results demonstrated that trees with a thick cork layer had lower lignin and higher suberin contents, respectively, than trees with a thin cork layer. Older *Q. variabilis* trees had higher lignin content when compared to younger trees (Fig. 3g–h). Notably, the IAA content in phellogen was significantly higher in trees with a thick cork layer compared to those with a thin cork layer ($P < 0.05$) (Fig. 3i). The biosynthesis pathway of suberin formation in *Q. variabilis*

was outlined providing key genetic resources for further investigation of molecular approaches to improve the thickness of cork production.

3.6. Transcriptomic analysis reveals the molecular mechanism of cork thickness

We selected five tree ages for a transcriptomic analysis of *Q. variabilis* trees with distinct differences in cork thickness. The number of differentially expressed genes (DEGs) was compared in 30 tree samples spanning five age groups (9-, 17-, 29-, 38-, and 51-year-old) with two different cork thicknesses (thin vs thick) of cork cambium from *Q. variabilis* (Table S20). A total of 54,880 DEGs were subjected to a weighted gene coexpression network analysis (WGCNA), which clustered genes into 20 modules (Fig. 4a). The yellow module contained hormone-specific signal transduction genes, including cyclin-D3-3-like (*CYCD3*), auxin-responsive protein small auxin-up RNA 50 (*SAUR50*), and suberin biosynthesis proteins glycerol-3-phosphate 2-O-acyltransferase 4 (*GPAT4*), fatty alcohol: caffeoyl-CoA acyltransferase-like (*FACT*), and caffeic acid 3-O-methyltransferase (*COMT*), expressed in 9- and 17-year-old trees with a thick cork layer (Fig. 4b, S8a–b, and Table S21). The cyan module contained genes involved in plant–pathogen interaction pathways that were predominantly expressed in trees with thick cork (Fig. S9a–b and Table S22). The blue module contained genes involved in photosynthesis-related metabolic pathways and were predominantly expressed in 9-year-old trees with thin cork (Fig. S10a–b and Table S23). In contrast, the ivory module contained genes involved in starch and sucrose metabolism, the royal-blue module contained genes involved in plant–pathogen interaction pathways, including senescence-associated carboxylesterase 101-like (*SAG101*), LRR receptor-like serine (*LRR-RLK*), and *disease resistance protein* which were highly expressed in 38- and 51-year-old trees with thick cork (Fig. S11a–b, S12a–b, and Tables S24–25). Finally, lightsteelblue1 module contained genes involved in phenylpropanoid and flavonoid biosynthesis pathways, e.g., *peroxidase 15*, *peroxidase 17* (*PER17*), dihydroflavonol 4-reductase (*DFR*), and chalcone synthase, which were overexpressed in 51-year-old trees with thin cork (Fig. S13a–b and Table S26).

To validate the accuracy of our transcriptomic data, we analyzed the corresponding transcript level patterns for eight selected genes with quantitative real-time polymerase chain reaction (qRT-PCR) assays (Table S27). The expression levels of the 8 candidate genes relative to that of the *Actin* reference

gene were compared with differentially-expressed proteins. The patterns of gene and protein expression levels were similar, corroborating our results (Fig. S14).

3.7. Hub genes could induce the thickness traits of *Q. variabilis*

To identify the key genes of suberin biosynthesis that contribute to thick cork formation, CytoHubba analysis of the yellow module genes was performed on ten hub genes known to be associated with suberin biosynthesis in *Q. variabilis* from the five age groups (9-, 17-, 29-, 38-, and 51-year-old) and two different cork thicknesses (thin vs thick) (Fig. 4c and Table S28). There were five genes involved in suberin biosynthesis, including *GPAT4*, *FATB* (palmitoyl-acyl carrier protein thioesterase), *COMT*, cytochrome P450 86A1 (*CYP86A1*), and nonspecific lipid-transfer protein *LTPG* (Fig. 4d). Among these genes, the expression of *GPAT4* and *FATB* were associated with triacylglycerol monomer and fatty acid biosynthesis regulating the differentiation of thick and thin cork of *Q. variabilis*.

3.8. *GPAT* family genes probably affect suberin biosynthesis

To infer the influence of the *GPAT* family on suberin biosynthesis in *Q. variabilis*, molecular evolutionary analysis was conducted. In total, 72 *GPAT* genes were identified among all 12 species, 13 of which were highly regulated in the thick cork of *Q. variabilis* (Fig. 4e and Table S29). *GPAT* encodes an enzyme involved in the initiation of biosynthesis of triacylglycerol [62]. For example, *GPAT4* and 5 were involved in the biosynthesis of suberin and cutin in *Q. suber* [20], and *GPAT5*, 6, 7, and 8 were involved in suberin biosynthesis in poplar [62]. Phylogenetic analyses revealed separate clustering of *GPAT* genes involved in the biosynthesis of cutin and suberin [63]. These findings led to the conclusion that the *GPAT* gene family probably influence plant cuticle or cork biosynthesis.

3.9. Fatty acid biosynthesis may affect the thickness traits of *Q. variabilis*

Fatty acid synthesis occurs in the plastid giving rise predominately to 16:0 and 18:1 acyl chain in the suberin biosynthesis pathway [21]. *FATB*, a hub gene, is important for the biosynthesis of fatty acids used for suberin production [64,65]. Interestingly, specific and expanded gene families involved in fatty acid biosynthesis were detected in *Q. variabilis*. Seventeen genes with significantly higher expression levels in thick cork than that of thin cork were identified as important for fatty acid biosynthesis (Fig. 5a). This result is consistent with the changes in fatty acid content (Fig. 5b). The

fatty acid synthesis pathway may be the crucial step contributing to the suberin biosynthesis in the thick cork of *Q. variabilis*.

3.10. Genomic insights into suberin biosynthesis and regulation in *Q. variabilis*

Comparative genomic and transcriptomic analyses were performed to identify key suberin biosynthesis-related genes in *Q. variabilis*. Based on the expression profile of suberin known genes, blast analyses with *A. thaliana* homologs revealed the presence of 66 genes in 14 suberin biosynthesis related families in *Q. variabilis* (Fig. 6 and Table S30). Furthermore, 27 DEGs were identified in thick versus thin corks of 17-year-old *Q. variabilis* at the \log_2 -fold-change ≥ 2 threshold (thick/thin). The genes of suberin polyaliphatic monomers biosynthesis included those involved in very long chain fatty acids (*LACS1* and *LACS2*), fatty acid elongation (*KCS2*, *KCS6*, and *KCS20*), oxidation (*CYP86A1* and *CYP86B1*), reduction (*FAR3* and *FAR4*). Furthermore, the genes of suberin polyphenolic monomer formation were identified, including the phenylpropanoid pathway (*ASFT* and *FACT*). Additionally, the transcription levels of suberin monomer transport to the cell wall (*ABCG6* and *ABCG20*), polymerization (*GELP38*, *GELP51* and *GELP72*), and positive regulation (*MYB93*, *MYB36*, *MYB41*, and *WRKY9*) were higher in the trees with a thick cork layer than in the trees with a thin cork layer. Therefore, these genes associated with the main suberization process may contribute to the molecular pathways leading to a relatively thick cork layer.

Further analysis to identify new candidate genes involved in suberin biosynthesis revealed 198 DEGs in thick versus thin corks of 17-year-old *Q. variabilis* at fold-change ≥ 2 and FDR < 0.01 (Table S31) involved in very long chain fatty acid formation (*LACS4*, *LACS7*, *FATB*, and *KAS2*), fatty acid elongation (*KCS4*), suberin biosynthesis (*FAR*, *CYP94A1*, and *CYP94A2*), phenylpropanoid pathway (*4CLL9*, *COMT*, *PER24*, *ASFT* and *FACT*), transport (*ABCG11* and *ABCG34*), deposition, assembly, and regulation (GDSL esterase/lipase genes, *CASPLIB2*, *CASPLIC1*, *MYB3*, *MYB308*, *NAC083*, and *WRKY56*). The transcription levels of these candidates were higher in the trees with a thick cork layer than in those with a thin cork layer, suggesting that they may contribute to suberin production.

4. Discussion

A high-fidelity reference genome is a prerequisite for genetic studies, transcriptomic analyses, and molecular assisted breeding programs [66]. *Q. variabilis* is a woody tree species that contributes

significantly to cork production in China. To instigate a molecular breeding program to improve cork yield and quality, a relatively precise *Q. variabilis* reference genome was assembled. Deployment of robust, high-fidelity third generation sequencing platforms such as PacBio CCS long-reads and Hi-C seq ensured the quality of the *Q. variabilis* genome. Collectively, our data indicated that the scaffold N50 (63.37 Mb) increased about 49-fold and 127-fold compared with that of *Q. suber* (0.5 Mb), *Q. robur* (1.3 Mb), but is slightly lower than the latest published genome assembly data of several *Quercus* genus including *Q. mongolica* (66.7 Mb), *Q. variabilis* (64.86 Mb) and *Q. dentata* (75.55 Mb), while the contig N50 (17.10 Mb) is much larger than that of *Q. suber* (0.08 Mb), *Q. mongolica* (2.64 Mb), *Q. robur* (0.07 Mb), *Q. lobata* (0.095 Mb) and *Q. dentata* (4.21 Mb), but is slightly lower than that of *Q. variabilis* (26.04 Mb) (Table S32) [23,25–29]. We have identified 54,606 protein-coding genes, which is significantly more than that of another published genome of *Q. variabilis* with 32,466 protein-coding genes. Effective gene annotation is of great significance for further investigations of gene function and identification of gene transcription regulation information in the future [67]. It is obviously the larger number of genes were helpful to identify more genes related to cork formation [24].

The genome and transcriptome data further confirmed that the obtained *Q. variabilis* genome was nearly complete. Moreover, the genes regulating the biosynthesis and controlled deposition of suberin, which affect cork thickness, were identified. Importantly, 66 DEGs involved in known suberin biosynthesis which was 27 more than identified in *Q. suber* that corresponded to *Arabidopsis* homologs [68]. These DEGs including *KCS*, *FAR*, *CYP86A1*, *GPAT*, and *ABCG*, were up-regulated in the thick cork of *Q. variabilis*, similar to the high-quality cork species *Q. suber* [6,19]. Specifically, homologs of *AtKCS2* and *StKCS6* were identified that encode rate-limiting enzymes that regulate elongation of C20 acyl chain suberin precursors in roots and potato tuber periderm, respectively [12,69]. *CYP86A1* and *B1* are involved in the oxidation of the fatty acyl group of the suberin monomer [70,71]. *FAR* genes induce the conversion of active fatty acyl-CoA to primary alcohols [11,72]. The *ABCG6*, *ABCG20*, and *LTPG* genes encode proteins that interact with embryonic precursors to pass through the plasma membrane and deposit in the cell wall [73]. Biosynthesis of suberin as well as transport of components to the cell wall play key roles in the production of cork. The crucial steps and genes of the suberin

biosynthesis pathway and the regulation of phellem formation in *Q. variabilis* require more comprehensive elucidation.

We identified 10 hub genes and 198 DEGs were highly up-regulated which are involved in regulation of phellogen development and the thickness of the cork layer. *GPAT4*, *COMT*, *CYP86A1*, and *LTPG* family genes play key roles in the formation of the plant cuticle layer and cork formation in *Q. suber* and poplar [20,62]. However, the other identified hub genes, including *FATBs*, *HHT*, *HIPP04*, *HHT1*, *NAKR2*, *PRP5*, and *PERK2*, that have not or rarely been reported to be related to cork layer formation. Specifically, *FATBs* were identified in the *Q. variabilis* genome as expanded gene families that are involved in the biosynthesis of saturated fatty acids [64], and precursors of wax production [65,74]. The further functional characterization of these candidates may clarify the molecular mechanisms for their functional roles in increasing the cork thickness of *Q. variabilis*.

Our results revealed the rapid growth and expansion of *Q. variabilis* phellogen cells in the thick cork layer during different ontogenetic stages. Exogenous application of IAA induces expression of putative suberin monomer transporters *ABCG6* and *ABCG20* [13]. Auxin-responsive *SAUR50* and *CYCD3* were upregulated in 9- and 17-year-old *Q. variabilis* trees, similar to cassava, poplar, and *Q. suber* [25,68,75]. In contrast, down-regulation of a number of genes involved in cell division and biosynthesis of IAA was observed in 38- and 51-year-old *Q. variabilis*, similar to *Ginkgo biloba* [76]. It is hypothesized that the expression of *MYB93*, *MYB36*, *MYB41*, and *WRKY9* in 9- and 16-year-old *Q. variabilis* is induced by auxin [77], and overexpression of *MYB41* further induces the expression of downstream targets, including *CYP86A1*, *KCS*, *FAR*, and *ASFT* [78,79], resulting in biosynthesis of suberin and cork thickness [78]. The complex networks regulating cell division- and suberin biosynthesis-related genes in *Q. variabilis* trees with a thick cork layer must be more precisely characterized.

Cork properties are determined by the suberin and lignin contents and their relative proportions [5]. The biosynthesis of free phenolic compounds is enhanced by the decrease in cell division as trees age, resulting in the thickening of cork cell walls [7]. However, in thin cork layers, *COMT*, *PER17*, and *DFR* expression levels are consistently up-regulated; these genes are involved in the phenylpropanoid and flavonoid pathways, which are important for regulating lignification [80]. These findings support the reported wound-induced decrease in suberin synthesis as potato plants at age [22]. The responses

of these pathways to stress were reflected by expression-level changes to the associated genes in the low-quality cork of *Q. suber* [7], *Q. ilex* [65], and *Q. cerris* × *suber* trees [81]. The *SAG101* and *LRR-RLK* expression levels in the *Q. variabilis* phellogen increased as the trees' age, which was in line with the findings of previous investigations on *Q. robur* [82] and *G. biloba* [76]. Hence, our results demonstrate that the tree age negatively regulates phellogen formation. Therefore, phellogen growth and development is affected by age, but the underlying mechanism require further study.

Conclusions

In conclusion, the newly assembled *Q. variabilis* genome is a high-fidelity reference that is required for genomic selection and molecular breeding to produce high-quality cork. Based on comparative genomic and transcriptomic analysis, we identified hub genes and key steps involved in suberin biosynthesis and cork formation. Thus, we provide a precise genomic resource for marker-assisted breeding to improve cork production and quality in *Q. variabilis*.

Author contributions

JL, ZJ and EC planned and designed the research, and managed the project. JL and EC contributed to sample preparation and sequencing. WG and JZ analyzed the data. EC and JL analyzed data and wrote the manuscript. WG, JC, AM, TJT, XY, and ZJ revised the manuscript, provided advice on the experiments and finalized the manuscript. All authors read and approved the manuscript.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the National Non-Profit Research Institutions of the Chinese Academy of Forestry [Grant numbers: CAFYBB2022ZA001] and the National Natural Science Foundation of China [Grant number 42071065]. The genome sequencing, assembly and annotation were performed with the help of Biomarker Technologies. We are particularly thankful to Dr. Quanzi Li, Dr. Yunhui Xie, and Dr. Guodong Rao for their helpful suggestions on our work.

Data availability

The raw sequence data and assembly of *Quercus variabilis* genome sequencing (PRJCA010364, CRA008412 and CRA008454) and RNA sequencing (PRJCA010362, CRA007638) have been deposited in at the National Genomics Data Center. Genome assembly, gene annotation, mRNA, protein and CDS sequences profiles are also available at FigShare (<https://doi.org/10.6084/m9.figshare.24064014>).

References

- [1] P.S. Manos, A.M. Stanford, The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere. *Int. J. Plant Sci.* 162 (2001) S77–S93.
- [2] J.F. Liu, Y.P. Deng, X.F. Wang, Y.Y. Ni, Q. Wang, W.F. Xiao, J.P. Lei, Z.P. Jiang, M.H. Li, The concentration of non-structural carbohydrates, N, and P in *Quercus variabilis* does not decline toward its northernmost distribution range along a 1500 km transect in China. *Front. Plant Sci.* 9 (2018) 1444.
- [3] X. Zhao, L. Zheng, X. Xia, W. Yin, J. Lei, S. Shi, X. Shi, H. Li, Q. Li, Y. Wei, E. Chang, Z. Jiang, J. Liu, Responses and acclimation of Chinese cork oak (*Quercus variabilis* Bl.) to metal stress: the inducible antimony tolerance in oak trees, *Environ. Sci. Pollut. Res.* 22 (2015) 11456–11466.
- [4] I. Miranda, J. Gominho, H. Pereira, Cellular structure and chemical composition of cork from the Chinese cork oak (*Quercus variabilis*), *J. Wood Sci.* 59 (2013) 1–9.
- [5] C. Leite, V. Oliveira, I. Miranda, H. Pereira, Cork oak and climate change: Disentangling drought effects on cork chemical composition, *Sci. Rep.* 10 (2020) 7800.
- [6] R.T. Teixeira, A.M. Fortes, C. Pinheiro, H. Pereira, Comparison of good- and bad-quality cork: application of high-throughput sequencing of phellogenic tissue, *J. Exp. Bot.* 65 (2014) 4887–4905.
- [7] R.T. Teixeira, A.M. Fortes, H. Bai, C. Pinheiro, H. Pereira, Transcriptional profiling of cork oak phellogenic cells isolated by laser microdissection, *Planta* 247 (2018) 317–338.
- [8] D.G. Branco, J.R. Campos, L. Cabrita, D.V. Evtuguin, Structural features of macromolecular components of cork from *Quercus suber* L., *Holzforschung* 74 (2020) 625–633.
- [9] A. Sen, M. Zhianski, M. Glushkova, K. Petkova, J. Ferreira, H. Pereira, Chemical composition and cellular structure of corks from *Quercus suber* trees planted in Bulgaria and Turkey, *Wood Sci. Technol.* 50 (2016) 1261–1276.
- [10] B. Dehane, A. Benrahou, R. Bouhraoua, F. Hamani, L. Belhoucine, Chemical composition of Algerian cork according the origin and the quality, *Int. J. Res. Envir. Studies* 1 (2014) 17–25.
- [11] F. Domergue, S.J. Vishwanath, J. Joubès, J. Ono, J.A. Lee, M. Bourdon, R. Alhattab, C. Lowe, S. Pascal, R. Lessire, O. Rowland, Three Arabidopsis fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate

- primary fatty alcohols associated with suberin deposition, *Plant Physiol.* 153 (2010) 1539–1554.
- [12] O. Serra, M. Soler, C. Hohn, R. Franke, L. Schreiber, S. Prat, M. Molinas, M. Figueras, Silencing of *StKCS6* in potato periderm leads to reduced chain lengths of suberin and wax compounds and increased peridermal transpiration, *J. Exp. Bot.* 60 (2009) 697–707.
- [13] V. Yadav, I. Molina, K. Ranathunge, I.Q. Castillo, S.J. Rothstein, J.W. Reed, J.W. 2014, ABCG transporters are required for suberin and pollen wall extracellular barriers in *Arabidopsis*, *Plant Cell* 26 (2014) 3569–3588.
- [14] G. Nomberg, O. Marinov, G.C. Arya, E. Manasherova, H. Cohen, The key enzymes in the suberin biosynthetic pathway in plants: An update, *Plants* 11 (2022) 392.
- [15] R. Franke, R. Höfer, I. Briesen, M. Emsermann, N. Efremova, A. Yephremov, L. Schreiber, The *DAISY* gene from *Arabidopsis* encodes a fatty acid elongase condensing enzyme involved in the biosynthesis of aliphatic suberin in roots and the chalaza-micropyle region of seeds, *Plant J.* 57 (2009) 80–95.
- [16] T. Capote, P. Barbosa, A. Usié, A.M. Ramos, V. Inácio, R. Ordás, S. Gonçalves, L. Morais-Cecílio, ChIP-Seq reveals that QsMYB1 directly targets genes involved in lignin and suberin biosynthesis pathways in cork oak (*Quercus suber*), *BMC Plant Biol.* 18 (2018) 198.
- [17] P. Krishnamurthy, B. Vishal, A. Bhal, P.P. Kumar, WRKY9 transcription factor regulates cytochrome P450 genes *CYP94B3* and *CYP86B1*, leading to increased root suberin and salt tolerance in *Arabidopsis*, *Physiol. Plantarum* 172 (2021) 1673–1687.
- [18] R. Verdaguer, M. Soler, O. Serra, A. Garrote, S. Fernández, D. Company-Arumí, E. Anticó, M. Molinas, M. Figueras, Silencing of the potato *StNAC103* gene enhances the accumulation of suberin polyester and associated wax in tuber skin, *J. Exp. Bot.* 67 (2016) 5415–5427.
- [19] M. Zhou, N. Chen, Y. Zou, P. Zhang, J. He, X. Xu, Comparative analysis of periderm suberin in stems and roots of *Tetraena mongolica* Maxim and *Zygophyllum xanthoxylum* (Bunge) Engl, *Trees* 36 (2022) 325–339.
- [20] M. Soler, O. Serra, M. Molinas, G. Hugué, S. Fluch, M. Figueras, A genomic approach to suberin biosynthesis and cork differentiation. *Plant Physiol.* 144 (2007) 419–431.
- [21] S.J. Vishwanath, C. Delude, F. Domergue, O. Rowland, Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier, *Plant Cell Rep.* 34 (2015) 573–586.
- [22] G.N.M. Kumar, E.C. Lulai, J.C. Suttle, N.R. Knowles, Age-induced loss of wound-healing ability in potato tubers is partly regulated by ABA, *Planta* 232 (2010) 1433–1445.
- [23] A.M. Ramos, A. Usié, P. Barbosa, P.M. Barros, T. Capote, I. Chaves, F. Simões, I. Abreu, I. Carrasquinho, C. Faro, J.B. Guimarães, D. Mendonça, F. Nóbrega, L. Rodrigues, N.J.M. Saibo, M.C. Varela, C. Egas, J. Matos, C.M. Miguel, M.M. Oliveira, C.P. Ricardo, S. Gonçalves, The draft genome sequence of cork oak, *Sci. Data* 5 (2018) 180069.
- [24] B. Han, L. Wang, Y. Xian, X.M. Xie, W.Q. Li, Y. Zhao, R.G. Zhang, X. Qin, D.Z. Li, K.H. Jia, A chromosome-level genome assembly of the Chinese cork oak (*Quercus variabilis*), *Front. Plant Sci.* 13 (2022).
- [25] X. Fu, H. Su, S. Liu, X. Du, C. Xu, K. Luo, Cytokinin signaling localized in phloem noncell-autonomously regulates cambial activity during secondary growth of *Populus* stems, *New Phytol.* 230 (2021) 1476–1488.
- [26] V.L. Sork, S.T. Fitz-Gibbon, D. Puiu, M. Crepeau, P.F. Gugger, R. Sherman, K. Stevens, C.H. Langley, M.

- Pellegrini, S.L. Salzberg, First draft assembly and annotation of the genome of a California endemic oak *Quercus lobata* Née (Fagaceae), *G3-Genes Genom. Genet.* 6 (2016) 3485–3495.
- [27] W. Ai, Y. Liu, M. Mei, X. Zhang, E. Tan, H. Liu, X. Han, H. Zhan, X. Lu, A chromosome-scale genome assembly of the Mongolian oak (*Quercus mongolica*), *Mol. Ecol. Resour.* 22 (2022) 2396–2410.
- [28] C. Plomion, J.M. Aury, J. Amselem, T. Leroy, F. Murat, S. Duplessis, S. Faye, N. Francillonne, K. Labadie, G. Le Provost, I. Lesur, J. Bartholomé, P. Faivre-Rampant, A. Kohler, J.C. Leplé, N. Chantret, J. Chen, A. Diévert, T. Alaeitabar, V. Barbe, C. Belser, H. Bergès, C. Bodénès, M.B. Bogeat-Triboulot, M.L. Bouffaud, B. Brachi, E. Chancerel, D. Cohen, A. Couloux, C. Da Silva, C. Dossat, F. Ehrenmann, C. Gaspin, J. Grima-Pettenati, E. Guichoux, A. Hecker, S. Herrmann, P. Hugueney, I. Hummel, C. Klopp, C. Lalanne, M. Lascoux, E. Lasserre, A. Lemainque, M.L. Desprez-Loustau, I. Luyten, M.A. Madoui, S. Mangenot, C. Marchal, F. Maumus, J. Mercier, C. Michotey, O. Panaud, N. Picault, N. Rouhier, O. Rué, C. Rustenholz, F. Salin, M. Soler, M. Tarkka, A. Velt, A.E. Zanne, F. Martin, P. Wincker, H. Quesneville, A. Kremer, J. Salse, Oak genome reveals facets of long lifespan, *Nat. Plants* 4 (2018) 440–452.
- [29] W.B. Wang, X.F. He, X.M. Yan, B. Ma, C.F. Lu, J. Wu, Y. Zheng, W.H. Wang, W.B. Xue, X.C. Tian, Y.A. El-Kassaby, I. Porth, S.P. Leng, Z.H. Hu, J.F. Mao, (2023). Chromosome-scale genome assembly and insights into the metabolome and gene regulation of leaf color transition in an important oak species, *Quercus dentata*. *New Phytol.* (2023) 238: 2016–2032
- [30] H. Cheng, G.T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm, *Nat. Methods.* 18 (2021) 170–175.
- [31] G. Parra, K. Bradnam, I. Korf, CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes, *Bioinformatics* 23 (2007) 1061–1067.
- [32] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics* 31 (2015) 3210–3212.
- [33] N. Servant, N. Varoquaux, B.R. Lajoie, E. Viara, C.J. Chen, J.P. Vert, E. Heard, J. Dekker, E. Barillot, HiC-Pro: an optimized and flexible pipeline for Hi-C data processing, *Genome Biol.* 16 (2015) 259.
- [34] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, *Genome Biol.* 14 (2013) R36.
- [35] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, *Nat. Protoc.* 7 (2012) 562–578.
- [36] J.M. Flynn, R. Hubley, C. Goubert, J. Rosen, A.G. Clark, C. Feschotte, A.F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families, *Proc. Natl. Acad. Sci. U.S.A.* 117 (2020) 9451–9457.
- [37] S. Ou, N. Jiang, LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons, *Plant Physiol.* 176 (2018) 1410–1422.
- [38] D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons, *BMC Bioinformatics* 9 (2008) 18.
- [39] Z. Xu, H. Wang, LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons, *Nucleic Acids Res.* 35 (2007) W265–W268.

- [40] N. Chen, Using Repeat Masker to identify repetitive elements in genomic sequences, *Curr. Protoc. Bioinformatics* 5 (2004) 4–10.
- [41] M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding, *Bioinformatics* 24 (2008) 637–644.
- [42] I. Korf, Gene finding in novel genomes, *BMC Bioinformatics* 5 (2004) 1–9.
- [43] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, *Nat. Methods* 12 (2015) 357–360.
- [44] M. Pertea, G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, *Nat. Biotechnol.* 33 (2015) 290–295.
- [45] S. Tang, A. Lomsadze, M. Borodovsky, Identification of protein coding regions in RNA transcripts, *Nucleic Acids Res.* 43 (2015) e78–e78.
- [46] B.J. Haas, S.L. Salzberg, W. Zhu, M. Pertea, J.E. Allen, J. Orvis, O. White, C.R. Buell, J.R. Wortman, Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments, *Genome Biol.* 9 (2008) R7.
- [47] R. She, J.S.C. Chu, K. Wang, J. Pei, N. Chen, genBlastA: Enabling BLAST to identify homologous gene sequences, *Genome Res.* 19 (2009) 143–149.
- [48] E. Birney, M. Clamp, R. Durbin, GeneWise and genomewise, *Genome Res.* 14 (2004) 988–995.
- [49] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.* 25 (1997) 955–964.
- [50] E.P. Nawrocki, S.R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches, *Bioinformatics* 29 (2013) 2933–2935.
- [51] S. Griffiths-Jones, Rfam: annotating non-coding RNAs in complete genomes, *Nucleic Acids Res.* 33 (2004) D121–D124.
- [52] S. Griffiths-Jones, miRBase: microRNA sequences, targets and gene nomenclature, *Nucleic Acids Res.* 34 (2006) D140–D144.
- [53] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, *Genome Biol.* 20 (2019) 238.
- [54] T. De Bie, N. Cristianini, J.P. Demuth, M.W. Hahn, CAFE: a computational tool for the study of gene family evolution, *Bioinformatics* 22 (2006) 1269–1271.
- [55] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermin, ModelFinder: fast model selection for accurate phylogenetic estimates, *Nat. Methods* 14 (2017) 587–589.
- [56] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, *Mol. Biol. Evol.* 32 (2015) 268–274.
- [57] Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood, *Bioinformatics* 13 (1997) 555–556.

- [58] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, *Nat. Methods*. 12 (2015) 59–60.
- [59] A. Zwaenepoel, Y. Van de Peer, WGD-simple command line tools for the analysis of ancient whole-genome duplications, *Bioinformatics* 35 (2019) 2153–2155.
- [60] S.T. Lopes, D. Sobral, B. Costa, P. Perdiguero, I. Chaves, A. Costa, C.M. Miguel, Phellem versus xylem: genome-wide transcriptomic analysis reveals novel regulators of cork formation in cork oak, *Tree Physiol.* 40 (2020) 129–141.
- [61] The Earth BioGenome Project Working Group, The Earth BioGenome Project Website, (2021), <https://www.earthbiogenome.org>.
- [62] M.K. Rains, N.D. Gardiyehewa de Silva, I. Molina, Reconstructing the suberin pathway in poplar by chemical and transcriptomic analysis of bark tissues, *Tree Physiol.* 38 (2018) 340–361.
- [63] E. Waschburger, F.R. Kulcheski, N.M. Veto, R. Margis, M. Margis-Pinheiro, A.C. Turchetto-Zolet, Genome-wide analysis of the Glycerol-3-Phosphate Acyltransferase (GPAT) gene family reveals the evolution and diversification of plant GPATs, *Genet. Mol. Biol.* 41 (2018) 355–370.
- [64] G. Bonaventure, J.J. Salas, M.R. Pollard, J.B. Ohlrogge, Disruption of the *FATB* gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth, *Plant Cell* 15 (2003) 1020–1033.
- [65] P. Boher, M. Soler, A. Sánchez, C. Hoede, C. Noirot, J.A.P. Paiva, O. Serra, M. Figueras, A comparative transcriptomic approach to understanding the formation of cork, *Plant Mol. Biol.* 96 (2018) 103–118.
- [66] G. Rao, J. Zhang, X. Liu, C. Lin, H. Xin, L. Xue, C. Wang, De novo assembly of a new *Olea europaea* genome accession using nanopore sequencing, *Hortic. Res.* 8 (2021) 64.
- [67] Y. Sun, J. Guo, X. Zeng, R. Chen, Y. Feng, S. Chen, K. Yang, Chromosome-scale genome assembly of *Castanopsis tibetana* provides a powerful comparative framework to study the evolution and adaptation of Fagaceae trees, *Mol. Ecol. Resour.* 22 (2022) 1178–1189.
- [68] S. Fernández-Piñán, P. Boher, M. Soler, M. Figueras, O. Serra, Transcriptomic analysis of cork during seasonal growth highlights regulatory and developmental processes from phellogen to phellem formation, *Sci Rep.* 11 (2021) 12053.
- [69] S.B. Lee, S.J. Jung, Y.S. Go, H.U. Kim, J.K. Kim, H.J. Cho, O.K. Park, M.C. Suh, Two *Arabidopsis* 3-ketoacyl CoA synthase genes, *KCS20* and *KCS2/DAISY*, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress, *Plant J.* 60 (2009) 462–475.
- [70] R. Höfer, I. Briesen, M. Beck, F. Pinot, L. Schreiber, R. Franke, The *Arabidopsis* cytochrome P450 *CYP86A1* encodes a fatty acid ω -hydroxylase involved in suberin monomer biosynthesis, *J. Exp. Bot.* 59 (2008) 2347–2360.
- [71] V. Compagnon, P. Diehl, I. Benveniste, D. Meyer, H. Schaller, L. Schreiber, R. Franke, F. Pinot, *CYP86B1* is required for very long chain ω -hydroxyacid and α , ω -dicarboxylic acid synthesis in root and seed suberin polyester, *Plant Physiol.* 150 (2009) 1831–1843.
- [72] S. Wicke, G.M. Schneeweiss, C.W. Depamphilis, K.F. Müller, D. Quandt, The evolution of the plastid chromosome in land plants: gene content, gene order, gene function, *Plant Mol. Biol.* 76 (2011) 273–297.

- [73] K.N. Woolfson, M. Esfandiari, M.A. Bernards, Suberin biosynthesis, assembly, and regulation, *Plants*. 11 (2022) 555.
- [74] Y. Li-Beisson, B. Shorrosh, F. Beisson, M.X. Andersson, V. Arondel, P.D. Bates, S. Baud, D. Bird, A. DeBono, T.P. Durrett, Acyl-lipid metabolism, *The Arabidopsis Book/American Society of Plant Biologists*. 11 (2013).
- [75] D. Rüscher, J.M. Corral, A.V. Carluccio, P.A.W. Klemens, A. Gisel, L. Stabolone, H.E. Neuhaus, F. Ludewig, U. Sonnewald, W. Zierer, Auxin signaling and vascular cambium formation enable storage metabolism in cassava tuberous roots, *J. Exp. Bot.* 72 (2021) 3688–3703.
- [76] L. Wang, J. Cui, B. Jin, J. Zhao, H. Xu, Z. Lu, W. Li, X. Li, L. Li, E. Liang, X. Rao, S. Wang, C. Fu, F. Cao, R.A. Dixon, J. Lin, Multifeature analyses of vascular cambial cells reveal longevity mechanisms in old *Ginkgo biloba* trees, *Proc. Natl. Acad. Sci. U.S.A.* 117 (2020) 2201–2210.
- [77] R. Ursache, C. De Jesus Vieira Teixeira, V. Déneraud Tendon, K. Gully, D. De Bellis, E. Schmid-Siegert, T. Grube Andersen, V. Shekhar, S. Calderon, S. Pradervand, C. Nawrath, N. Geldner, J.E.M. Vermeer, GDSSL-domain proteins have key roles in suberin polymerization and degradation, *Nat. Plants* 7 (2021) 353–364.
- [78] X. Wei, W. Lu, L. Mao, X. Han, X. Wei, X. Zhao, M. Xia, C. Xu, ABF2 and MYB transcription factors regulate feruloyl transferase *FHT* involved in ABA-mediated wound suberization of kiwifruit, *J. Exp. Bot.* 71 (2020) 305–317.
- [79] D.K. Kosma, J. Murmu, F.M. Razeq, P. Santos, R. Bourgault, I. Molina, O. Rowland, At MYB 41 activates ectopic suberin synthesis and assembly in multiple plant species and cell types, *Plant J.* 80 (2014) 216–229.
- [80] M. Li, X. Cui, L. Jin, M. Li, J. Wei, Bolting reduces ferulic acid and flavonoid biosynthesis and induces root lignification in *Angelica sinensis*, *Plant Physiol. Bioch.* 170 (2022) 171–179.
- [81] B. Meireles, A. Usié, P. Barbosa, A.M. Fortes, A. Folgado, I. Chaves, I. Carrasquinho, R.L. Costa, S. Gonçalves, R.T. Teixeira, A.M. Ramos, F. Nóbrega, Characterization of the cork formation and production transcriptome in *Quercus cerris* × *suber* hybrids, *Physiol. Mol. Biol. Pla.* 24 (2018) 535–549.
- [82] T. Leroy, C. Plomion, A. Kremer, Oak symbolism in the light of genomics, *New Phytol.* 226 (2020) 1012–1017.

Figure legends

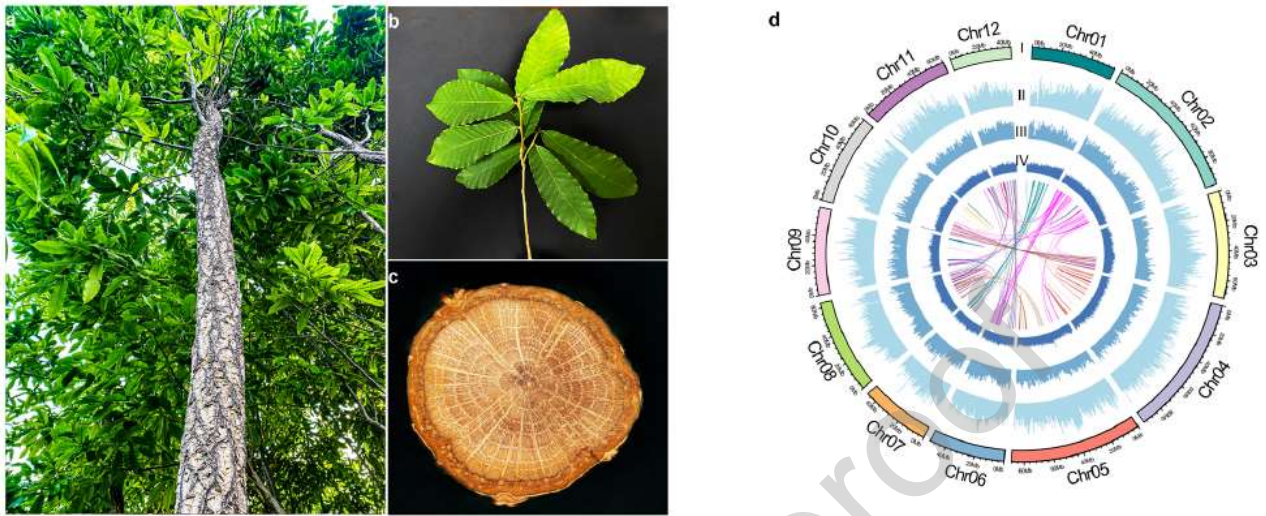


Fig. 1. Characteristics of the *Q. variabilis* tree and its genome assembly.

(a) Images of a 9-year-old *Q. variabilis* tree. (b) Leaves. (c) Stem cross-section. (d) *Q. variabilis* genome assembly and genomic features. The circles (outer to inner) represent the following: chromosome ideograms (I); TE density, sliding window = 500 kb (II); gene density, sliding window = 1 Mb (III); and GC content, sliding window = 500 kb (IV).

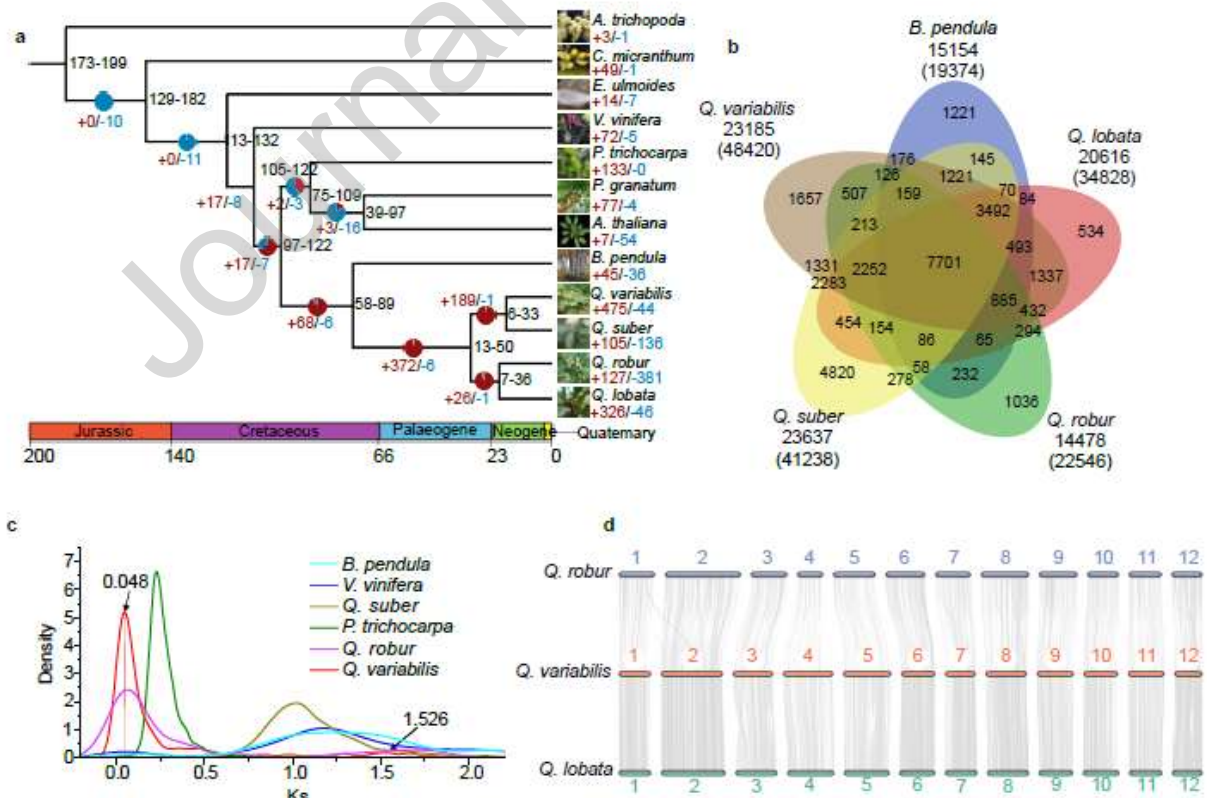


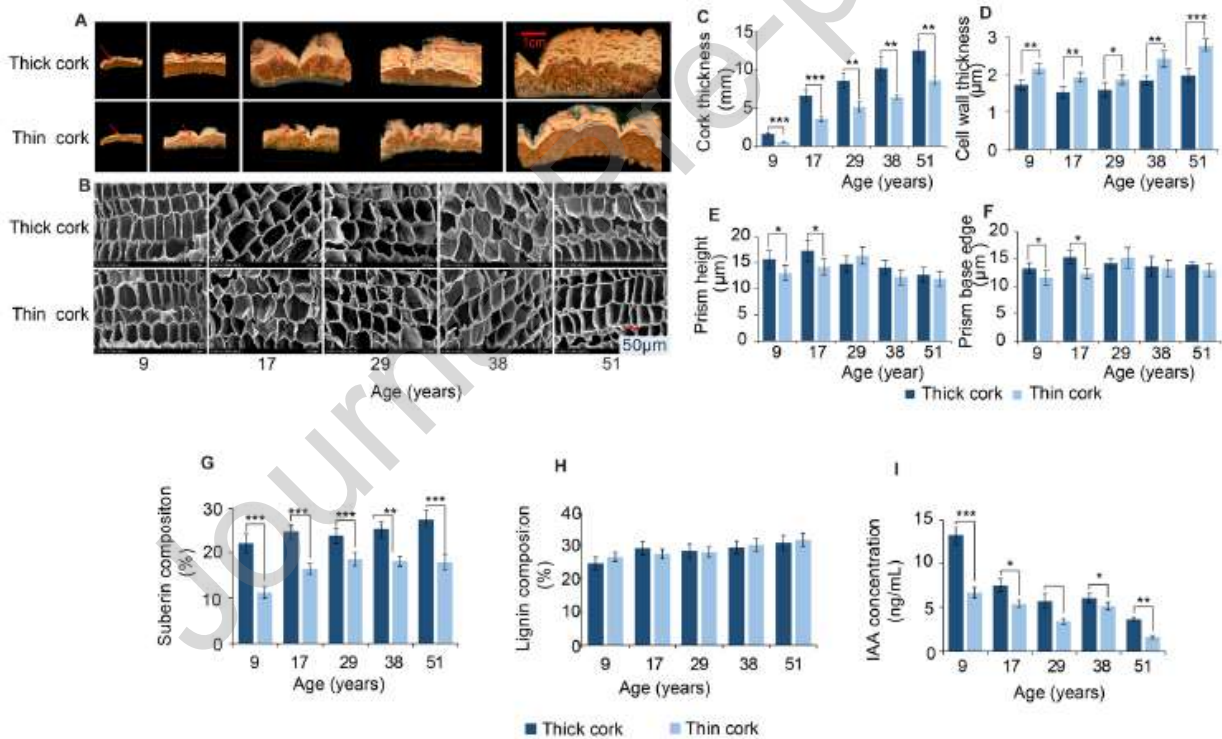
Fig. 2. Evolution and comparative analysis of the *Q. variabilis* genome.

(a) Phylogenetic tree of the expanded and contracted gene families in *Q. variabilis* and 11 other plant species (*A. trichopoda*, *C. micranthum*, *E. ulmoides*, *V. vinifera*, *P. trichocarpa*, *P. granatum*, *A. thaliana*, *B. pendula*, *Q. suber*, *Q. robur*, and *Q. lobata*). The blue and red values above each branch respectively represent the expanded and contracted gene families, respectively, after the diversification from the most common ancestor.

(b) Venn diagram of shared orthologs among *Q. variabilis* and four other plant species (*Q. suber*, *Q. lobata*, *Q. robur*, and *B. pendula*). Gene family numbers are presented in the diagram. Note: numbers below the species name and in the venn diagrams are the numbers of gene families; numbers in parentheses are the corresponding gene numbers.

(c) Synonymous substitution rate (Ks) distribution for pairs of syntenic paralogs in *Q. variabilis*, *B. pendula*, *Q. suber*, *Q. robur*, and *V. vinifera*.

(d) Synteny among *Q. variabilis*, *Q. lobata*, and *Q. robur*.

**Fig. 3.** Microscopic observations and physical and chemical analyses.

(a) Bark characteristics. (b) Scanning electron microscopy radial sections (Scale bar, 50 μm).

(c) Cork thickness, (d) Cell wall thickness, (e) Prism height, (f) Prism base edge, (g) suberin composition, (h) lignin content, and (i) IAA concentration of *Q. variabilis* trees that varied in terms of cork thickness and age. Values are means \pm SDs of three biological replicates. The asterisks indicate the significance ($***P < 0.001$, $**P < 0.01$, $*P < 0.05$, two-sided Student's *t*-test) were used to analyze the significant differences in phenotypes between the thick and thin groups.

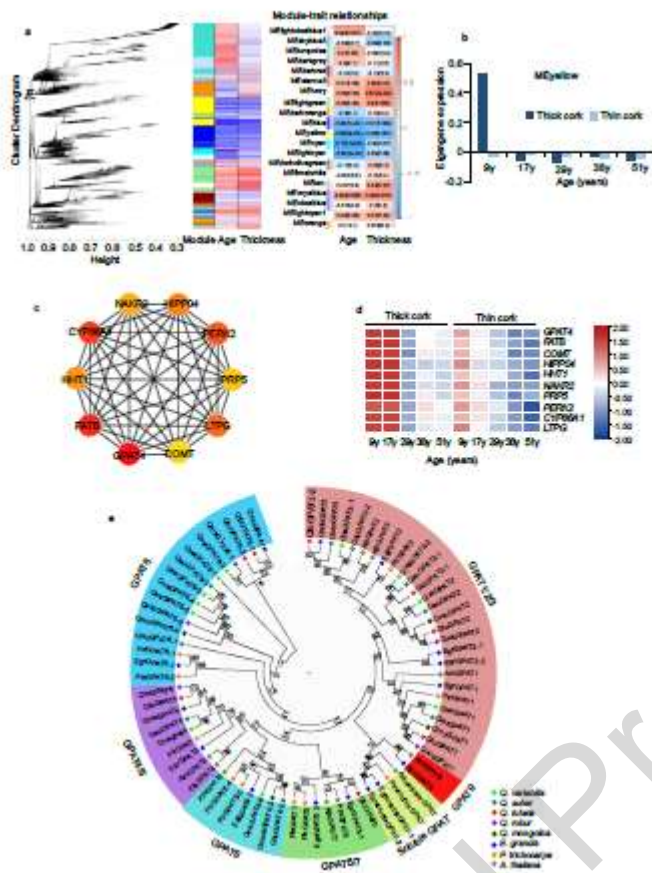


Fig. 4. Functional verification of key genes that regulate cork thickness.

(a) WGCNA dendrogram indicating the expression of different gene modules among *Q. variabilis* trees that varied in terms of cork thickness and age. For the trait and module association analyses, different colors represent different modules.

(b) Expression of the eigengenes in the yellow modules among samples.

(c) Hub genes in the intramodule regulatory networks for the yellow modules. Node color reflects the degree of connectivity. The pseudocolor scale from red to yellow represents the highest degree to lowest degree.

(d) Heatmap showing the gene expression profiles of 10 hub genes associated with suberin biosynthesis in thick cork identified in the yellow modules. The FPKM values were normalized using the LogNormalize method.

(e) Phylogenetic tree for the *GPAT* genes. The tree was constructed on the basis of protein sequences in *Q. variabilis*, *Q. suber*, *Q. lobata*, *Q. robur*, *Q. mongolica*, *E. grandis*, *P. trichocarpa*, and *A. thaliana*.

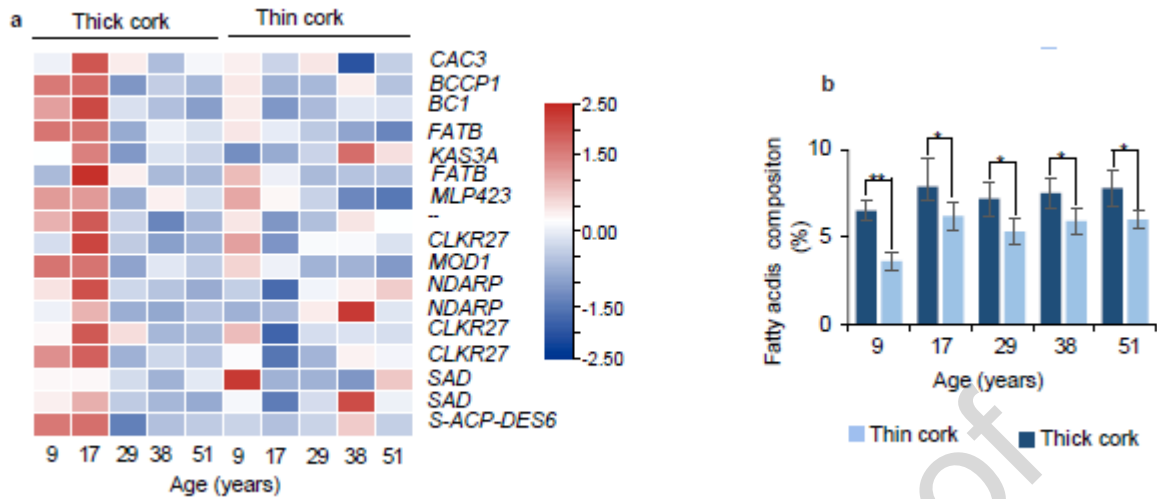


Fig. 5. (a) Heat map of expression data of candidate genes for the thickness cork of fatty acid biosynthesis. The FPKM values were normalized using the LogNormalize method. (b) The fatty acid content of the cork from the bark of *Q. variabilis* trees that varied in terms of cork thickness and age. Values are means \pm SDs of three biological replicates. The asterisks indicate the significance (** $P < 0.01$, * $P < 0.05$, two-sided Student's *t* test) and were used to analyze the significant differences in phenotypes between the thick and thin groups.

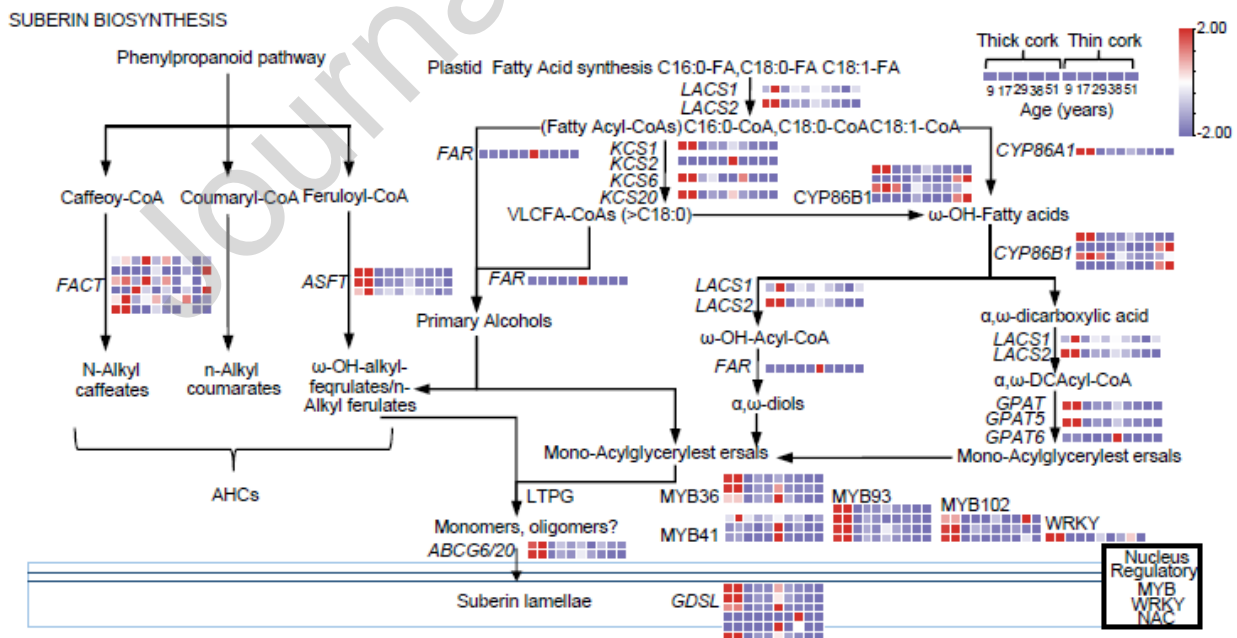


Fig. 6. Expression profiles of known genes involved in suberin-related processes in the cork of *Q. variabilis* trees. Blue and red arrows indicate established and hypothesized biosynthetic steps,

respectively. The FPKM values were normalized using the LogNormalize method. Gene abbreviations: LACS, Long chain acyl-CoA synthetase; KCS, 3-ketoacyl-CoA synthase; CYP86A1, Cytochrome P450 86A1; CYP86B1, Cytochrome P450 86B1; FAR, Fatty acyl reductases GPAT, Glycerol-3-phosphate acyltransferase; ASFT, Omega-hydroxypalmitate O-feruloyl transferase; ASFT, Aliphatic Suberin Feruloyl Transferase; FACT, Fatty alcohol:caffeoyle-CoA acyltransferase-like; ABCG, ABC transporter G family member; GELP, GDSL esterase/lipase.

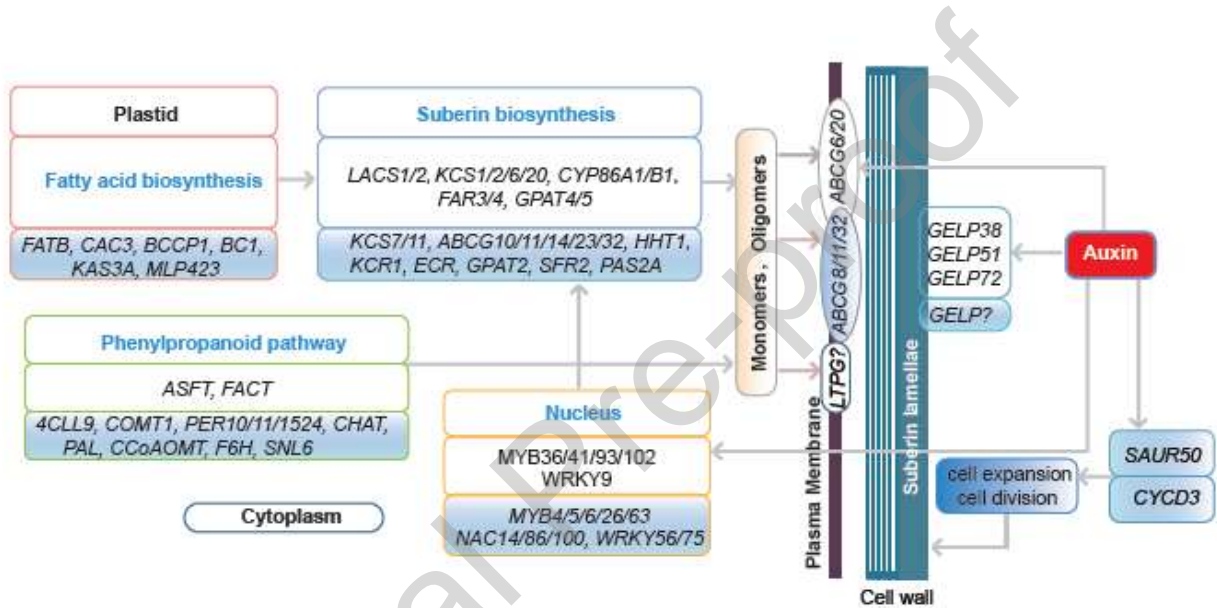


Fig. 7. Hypothetical schematic representation of the biosynthetic pathway of regulatory network affecting cork thickness of *Q. variabilis*. The scheme is proposed based on previous evidence in the literature and shows all genes listed in Table S27-S28 according to their known enzymatic reactions and the expression level of new candidate genes. The genes with names on the white background have been verified to play roles in improving suberin synthesis in plants. The genes with names on the blue background have high expression levels in thick corks of *Q. variabilis*. The direction of the arrows indicates the enhancement of the levels of gene expression in thick cork. Gene abbreviations: LACS, Long chain acyl-CoA synthetase; KCS, 3-ketoacyl-CoA synthase; CYP86A1, Cytochrome P450 86A1; CYP86B1, Cytochrome P450 86B1; FAR, Fatty acyl reductases GPAT, Glycerol-3-phosphate acyltransferase; ASFT, Omega-hydroxypalmitate O-feruloyl transferase; ASFT, Aliphatic Suberin Feruloyl Transferase; FACT, Fatty alcohol:caffeoyle-CoA acyltransferase-like; ABCG, ABC transporter G family member; GELP, GDSL esterase/lipase; FATB, Palmitoyl-acyl carrier protein

thioesterase; CAC3, Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha; BCCP1, Biotin carboxyl carrier protein of acetyl-CoA carboxylase 1; BC1, Biotin carboxylase 1; KAS3A, 3-oxoacyl-[acyl-carrier-protein] synthase 3 A; MLP423, MLP-like protein 423; 4CL1, 4-coumarate--CoA ligase 1; COMT1, Caffeic acid 3-*O*-methyltransferase 1; PER, Peroxidase; CHAT, (Z)-3-hexen-1-ol acetyltransferase; PAL, Phenylalanine ammonia-lyase; CCoAOMT, Caffeoyl-CoA *O*-methyltransferase; F6H, Feruloyl CoA ortho-hydroxylase; SNL, Cinnamoyl-CoA reductase-like; CYCD3, cyclin-D3-3-like; SAUR50, small auxin-up RNA 50.

Table 1 Summary of the *Q. variabilis* genome assembly and annotation

Assembly and annotation features	Value
Estimated genome size (Mb)	715.17
The assembled genome (Mb)	791.89
Number of scaffolds	270
Scaffold N50 (Mb)	63.37
Number of contigs	347
Contig N50 size (Mb)	17.10
GC content (%)	35.56
Pseudochromosomes	12
Repeat sequences	49.27%
BUSCO complete percentage in assembly	98.02%
Number of gene models	54,606

Compliance with ethics requirement

There was no use of any animals or human patients.

Declaration of Interests

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

Graphical abstract

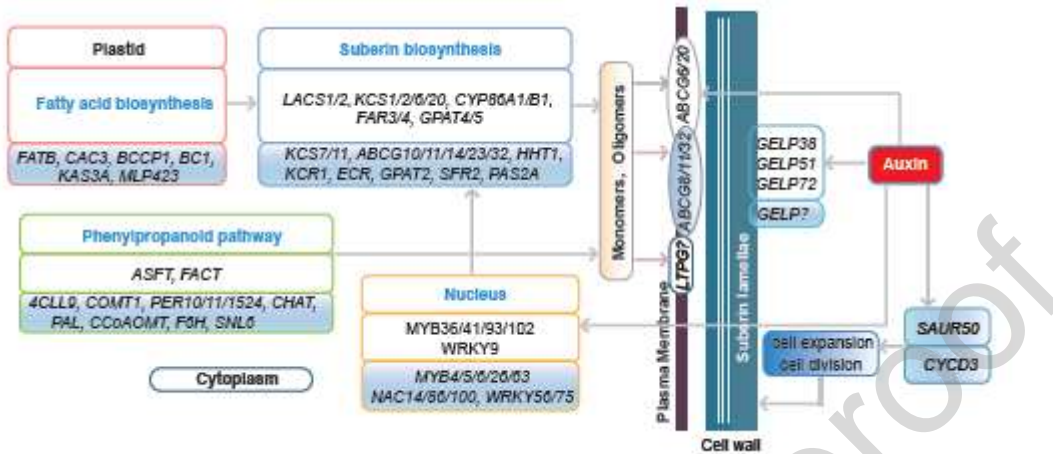


Fig. 7. Hypothetical schematic representation of the biosynthetic pathway of regulatory network affecting cork thickness of *Q. variabilis*. The scheme is proposed based on previous evidence in the literature and shows all genes listed in Table S27-S28 according to their known enzymatic reactions and the expression level of new candidate genes. The genes with names on the white background have been verified to play roles in improving suberin synthesis in plants. The genes with names on the blue background have high expression levels in thick corks of *Q. variabilis*. The direction of the arrows indicates the enhancement of the levels of gene expression in thick cork. Gene abbreviations: LACS, Long chain acyl-CoA synthetase; KCS, 3-ketoacyl-CoA synthase; CYP86A1, Cytochrome P450 86A1; CYP86B1, Cytochrome P450 86B1; FAR, Fatty acyl reductases GPAT, Glycerol-3-phosphate acyltransferase; ASFT, Omega-hydroxypalmitate O-feruloyl transferase; ASFT, Aliphatic Suberin Feruloyl Transferase; FACT, Fatty alcohol:caffeoyl-CoA acyltransferase-like; ABCG, ABC transporter G family member; GELP, GDSL esterase/lipase; FATB, Palmitoyl-acyl carrier protein thioesterase; CAC3, Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha; BCCP1, Biotin carboxyl carrier protein of acetyl-CoA carboxylase 1; BC1, Biotin carboxylase 1; KAS3A, 3-oxoacyl-[acyl-carrier-protein] synthase 3 A; MLP423, MLP-like protein 423; 4CL1, 4-coumarate--CoA ligase 1; COMT1, Caffeic acid 3-O-methyltransferase 1; PER, Peroxidase; CHAT, (Z)-3-hexen-1-ol acetyltransferase; PAL, Phenylalanine ammonia-lyase; CCoAOMT, Caffeoyl-CoA O-methyltransferase; F6H, Feruloyl CoA ortho-hydroxylase; SNL, Cinnamoyl-CoA reductase-like; CYCD3, cyclin-D3-3-like; SAUR50, small auxin-up RNA 50.

Suberin lamellae

Highlights

- High-quality chromosomal genome was assembled for *Q. variabilis* with length of 791,89 Mb and 54,606 predicted genes.
- Cell expansion, cell division, and suberin biosynthesis related genes contribute the cork thickness.
- A total of 198 suberin biosynthesis-related new candidate genes were identified, which contribute to the thick cork layer formation.