



OPEN ACCESS

EDITED BY

Veena Prahlad,
The University of Iowa, United States

REVIEWED BY

Kristina Gagalova,
Canada's Michael Smith Genome
Sciences Centre, Canada
Steven M. Johnson,
Brigham Young University, United States

*CORRESPONDENCE

María-Dolores Rey,
✉ b52resam@uco.es
Jesús V. Jorrín-Novo,
✉ bf1jonoj@uco.es

†PRESENT ADDRESS

Víctor M. Guerrero-Sánchez,
Cardiovascular Proteomics Laboratory,
Centro Nacional de Investigaciones
Cardiovasculares (CNIC), Madrid, Spain

RECEIVED 20 June 2023

ACCEPTED 20 September 2023

PUBLISHED 12 October 2023

CITATION

Rey M-D, Labella-Ortega M,
Guerrero-Sánchez VM, Carleial R,
Castillejo MÁ, Ruggieri V and
Jorrín-Novo JV (2023), A first draft
genome of holm oak (*Quercus ilex*
subsp. *ballota*), the most representative
species of the Mediterranean forest and
the Spanish agrosylvopastoral
ecosystem "dehesa".
Front. Mol. Biosci. 10:1242943.
doi: 10.3389/fmolb.2023.1242943

COPYRIGHT

© 2023 Rey, Labella-Ortega, Guerrero-Sánchez, Carleial, Castillejo, Ruggieri and Jorrín-Novo. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A first draft genome of holm oak (*Quercus ilex* subsp. *ballota*), the most representative species of the Mediterranean forest and the Spanish agrosylvopastoral ecosystem "dehesa"

María-Dolores Rey^{1*}, Mónica Labella-Ortega¹,
Víctor M. Guerrero-Sánchez^{1†}, Rômulo Carleial²,
María Ángeles Castillejo¹, Valentino Ruggieri³ and
Jesús V. Jorrín-Novo^{1*}

¹Agroforestry and Plant Biochemistry, Proteomics and Systems Biology, Department of Biochemistry and Molecular Biology, University of Cordoba, UCO-CeiA3, Cordoba, Spain, ²Royal Botanic Gardens, Kew, Richmond, United Kingdom, ³Biomeets Consulting ITNIG—Carrer d' Alaba 61 08005 Catalonia, Barcelona, Spain

The holm oak (*Quercus ilex* subsp. *ballota*) is the most representative species of the Mediterranean Basin and the agrosylvopastoral Spanish "dehesa" ecosystem. Being part of our life, culture, and subsistence since ancient times, it has significant environmental and economic importance. More recently, there has been a renewed interest in using the *Q. ilex* acorn as a functional food due to its nutritional and nutraceutical properties. However, the holm oak and its related ecosystems are threatened by different factors, with oak decline syndrome and climate change being the most worrying in the short and medium term. Breeding programs informed by the selection of elite genotypes seem to be the most plausible biotechnological solution to rescue populations under threat. To achieve this and other downstream analyses, we need a high-quality and well-annotated *Q. ilex* reference genome. Here, we introduce the first draft genome assembly of *Q. ilex* using long-read sequencing (PacBio). The assembled nuclear haploid genome had 530 contigs totaling 842.2 Mbp (N50 = 3.3 Mbp), of which 448.7 Mb (53%) were repetitive sequences. We annotated 39,443 protein-coding genes of which 94.80% were complete and single-copy genes. Phylogenetic analyses showed no evidence of a recent whole-genome duplication, and high synteny of the 12 chromosomes between *Q. ilex* and *Quercus lobata* and between *Q. ilex* and *Quercus robur*. The chloroplast genome size was 142.3 Kbp with 149 protein-coding genes successfully annotated. This first draft should allow for the validation of omics data as well as the identification and functional annotation of genes related to phenotypes of interest such as those associated with resilience against oak decline syndrome and climate change and higher acorn productivity and nutraceutical value.

KEYWORDS

Quercus, *Quercus ilex*, long-read assembly, PacBio HiFi sequencing, decline syndrome

Introduction

The holm oak (*Quercus ilex* subsp. *ballota*) is included in the 640 recorded species in the “Plants of the World Online” database within the genus *Quercus* (de Rigo and Caudullo, 2016). Oaks is one of the most important clades of woody angiosperms in the Northern Hemisphere in terms of species diversity, ecological dominance, and economic value (Backs and Ashley, 2021). It has been part of life, culture, and subsistence since ancient times (Leroy et al., 2020). Moreover, the holm oak is a priority species for afforestation in semi-arid regions due to its high drought tolerance (Barbeta and Peñuelas, 2016) and its phenotypic plasticity in response to varying edaphic conditions (Laureano et al., 2016).

The holm oak is the dominant species in the Mediterranean forest and agrosylvopastoral ecosystem “*dehesa*,” with relevance from an environmental and economic point of view, especially in rural areas (Plieninger et al., 2021). The most remarkable economic benefit is related to its fruit, the acorn, used for feeding black Iberian pigs (*Sus scrofa domesticus* L.) (Diaz-Caro et al., 2019). Recently, there has been a renewed interest in using *Quercus* spp. as a source of functional food due to the increasing demand for alternative plant species, including forest trees, for dietary diversification, and for sustainable food production (Vinha et al., 2016; Lopez Hidalgo et al., 2021).

Currently, *Q. ilex* is seriously threatened by several factors including aging tree populations, overexploitation coupled with poor regeneration capacity, inappropriate livestock management, and biotic (e.g., *Phytophthora cinnamomi*) and abiotic stresses (e.g., extreme temperatures and extended drought periods). Together, these factors likely contribute to the so-called holm oak decline syndrome, which has increased tree mortality and seriously damaged the *dehesa* (Brasier, 1992; Camilo-Alves et al., 2013; Ruiz-Gómez et al., 2018; San-Eufrasio et al., 2021). The main priority for *Q. ilex* conservation and sustainable exploitation is the establishment of breeding programs aimed at increasing acorn production and resilience to the decline syndrome (Cortés et al., 2020). These programs will contribute to the proper and sustainable management, conservation, and exploitation of *Q. ilex* and related ecosystems (Plomión et al., 2016). A better understanding of this species at the biochemistry, molecular biology, and biotechnology levels will facilitate these breeding programs, as it has been the case with many herbaceous and other woody species (Muthuramalingam et al., 2022). As with many long-lived, non-domesticated, and allogamous species, the most plausible breeding strategy for *Q. ilex* is the use of genome-wide association studies (GWAS) to identify elite genotypes associated with phenotypes of interest (e.g., high acorn productivity, high content in phenolic compounds, and resilience to adverse biotic and abiotic stresses; Maldonado-Alconada et al., 2022).

To date, the number of publications focusing on the molecular aspects of *Q. ilex* biology has been limited due to its challenging biological characteristics (long life cycle, allogamy, and high phenotypic variability) and its recalcitrance as an experimental system (Maldonado-Alconada et al., 2022). However, despite these difficulties, our group has, to some extent, characterized the genetic structure and diversity of *Q. ilex* using DNA markers, microsatellites, transcriptomics, proteomics, and metabolomics

(Maldonado-Alconada et al., 2022). These different approaches have allowed us to study seed germination, responses to biotic and abiotic stresses and their interaction, and the integration between molecular and morphometric data (Maldonado-Alconada et al., 2022). However, our research has so far been constrained by the lack of a *Q. ilex* reference genome, so the picture provided by these previous studies is still incomplete. The sequencing of the *Q. ilex* genome will add to the list of many recently published *Quercus* reference genomes (Plomión et al., 2016; Sork et al., 2016; Plomión et al., 2018; Ramos et al., 2018; Ai et al., 2022; Han et al., 2022; Liu et al., 2022; Sork et al., 2022; Zhou et al., 2022; Wang et al., 2023) and will allow us to interpret *Q. ilex* omics data in terms of its own gene products rather than relying on orthologs from other *Quercus* species. This should open new research venues related to the genetic makeup of this species, the identification of novel genes, allelic variants, and epigenetic marks, and will allow for studies on broader topics such as phylogenetics, hybridization and introgression patterns, and breeding programs informed by GWAS (Plomión et al., 2018; Stocks et al., 2019; Sork et al., 2022).

Many recently reported *de novo* genome assemblies have been carried out using next-generation short-read sequencing technologies such as Illumina or 454 platforms (Xia et al., 2017; Zhao et al., 2017), which tend to produce very fragmented genomes. Pacific BioSciences (PacBio) has developed a third-generation sequencing technology to assemble large and complex plant genomes producing tens of thousands of long individual reads (up to ~40 kb) (Roberts et al., 2013; Paajanen et al., 2019). Here, we generate a draft genome assembly for *Q. ilex* (NCBI’s BioProject repository, ID: PRJNA687489) using the single-molecule real-time (SMRT) sequencing technology from PacBio. We reveal genomic features of *Q. ilex*, including nuclear and chloroplast sequences, repetitive sequences, and gene annotations. We also perform ortholog identifications and comparative genomics analyses to investigate possible gene duplication events, gene family expansions and contractions, and the degree of collinearity between *Q. ilex* chromosomes and other *Quercus* species. This first draft should allow us to validate the omics data previously obtained by our research group (Maldonado-Alconada et al., 2022) and facilitate the genetic enhancement of *Q. ilex* including the identification and functional annotation of genes associated with phenotypes of interest, such as resilience against oak decline syndrome and climate change or increased acorn productivity. Additionally, it will help us to understand the phylogenetic relationships among species of the genus *Quercus*.

Materials and methods

Genomic DNA extraction and PacBio sequencing

Leaves were collected from a mature, healthy *Q. ilex* tree located in Aldea de Cuenca, Fuente Obejuna, Córdoba, Andalusia, Spain (UTM 30S 276751 4245466 datum ETRS89), in November 2019 (Supplementary Figure S1). Upon arrival at the laboratory, leaves were washed with 2% sodium hypochlorite and abundantly rinsed with distilled water. For long-read sequencing, extraction of high-molecular-weight (HMW) genomic DNA from leaves was carried

out using the cetyl trimethyl ammonium bromide (CTAB) method (Murray and Thomson, 1980) with slight modifications. In brief, 4 g of fresh leaves were macerated in liquid nitrogen, and 1 g of polyvinylpyrrolidone (PVPP) was added per each 20 mL of CTAB due to the high concentration of phenolic compounds in *Q. ilex* leaves. Once extracted, DNA was treated with RNase (10 mg/mL, Thermo Fisher Scientific Inc., Waltham, MA, United States) for 1 h 30 min at 37°C, and the quality and concentration were determined by 1% agarose gel electrophoresis and using a Qubit version 3.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, United States), respectively.

To determine whether the target individual was a pure *Q. ilex*, we compared its genotype with genotypes from other nine *Q. ilex* and 10 *Quercus suber* mature individuals using 20 microsatellite markers as previously described by Fernández i Marti et al. (2018). PCR amplifications were conducted in a T100™ Thermal Cycler (Bio-Rad, Hercules, CA, United States), and PCR products were analyzed on an ABI Prism 310 capillary electrophoresis system (Applied Biosystems, Foster City, CA, United States). Size alignment and quality control were analyzed using GeneMapper version 3.7 (Applied Biosystems, Foster City, CA, United States). We performed a principal component analysis (PCA) including all 20 microsatellite markers using the R package Adegenet (Jombart, 2008).

The target DNA sample was used to construct a PacBio SMRT sequencing library at the DNA Sequencing & Genotyping Center at the University of Delaware. Purified HMW DNA was fragmented to approximately 20 kb using Megaruptor version 3.0 (Diagenode, Denville, NJ, United States). Purified HMW DNA (10 µg) was used as a template to construct a SMRTbell HiFi library using Express Template Prep version 2.0 (Pacific Biosciences, Menlo Park, CA, United States) as per the manufacturer's protocol. The quality of the HiFi library was assessed using the Qubit 3.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, United States) and the Agilent Femto Pulse System (Agilent Technologies, Santa Clara, CA, United States). The HiFi library was run on two Sequel Ii system 8M SMRT cells using sequencing chemistry version 3.0 with 4 h pre-extension and 30 h movie time. Two runs of HiFi PacBio generated reads covering 64.3X of the genome assuming an *a priori* genome size of 850 Mbp. Although reads generated by HiFi PacBio are usually of high quality, an additional rough evaluation of nucleotide (A, C, G, and T) distribution was performed to eliminate any remaining low-quality reads. Reads showing a skewed nucleotide distribution (approximately 1%) were removed.

De novo *Q. ilex* genome assembly

Filtered reads were used as input for a *de novo* assembly approach undertaken using NextDenovo version 2.5.0 software (key parameters: seed_depth = 45, nextgraph: -a1) (<https://github.com/Nextomics/NextDenovo>). The software was run in a "hifi" assembly mode, which skips the read correction step and produces a consensus of reads using a string graph algorithm. Contigs were separated into primary contigs and associated secondary contigs using the Purge_haplotigs pipeline version 1.1.1 (key parameter: -l 7, -m 45 and -h 180) (<https://bitbucket.org/mroachawri/purge:haplotigs/src/master>) (Roach et al., 2018) (Supplementary Figure S2). In regions with homologous contigs

(e.g., heterozygous regions) in the haploid assembly, reads were split between the two contigs resulting in approximately half the read depth for those contigs. Based on the plot distribution, specific cut-offs were set to separate and flag low-quality contigs, secondary/alternative contigs, and high-coverage contigs (i.e., putative organelle genomes). Contigs, showing a very high coverage and forming a circular sequence, were blasted against the GenBank Non-Redundant Protein Sequence database (NCBI-BLAST) to verify matches with organelle genomes. The completeness of the genome assembly, generated by using primary contigs, was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO) version 5.0 (key parameters: default, -l viridiplantae_odb10) (Simão et al., 2015). Using a chromosome-based genome assembly of a close and syntenic species (*Q. lobata*, valley oak NCBI ID = GCF_001633185.2, ValleyOak release 3.2, with RefSeq annotation when this work was performed), contigs were scaffolded at the chromosome level using RagTag (key parameters: scaffolding) (Alonge et al., 2019). Each *Q. ilex* contig was mapped against the chromosomes of the *Q. lobata* genome and scored for confidence according to grouping, location, and orientation.

Repetitive sequences

We generated a high-quality non-redundant transposable element (TE) library for the *Q. ilex* genome using the Extensive *de novo* TE Annotator (EDTA) pipeline (<https://github.com/oushujun/EDTA>) (Ou et al., 2019). The pipeline combines a suite of best-performing packages (LTR_FINDER, LTR_harvest, LTR_retriever, Generic Repeat Finder, TIR-Learner, HelitronScanner, and RepeatMasker), and it was designed to filter out false discoveries in raw TE candidates and generate high-quality non-redundant TE libraries. The inbuilt package RepeatModeler version 2.0.1 (key parameters: default) (Bedell et al., 2000), which identifies any remaining TEs which might have been overlooked by the EDTA algorithm (--sensitive 1), was also used in the workflow. TE identification was performed using RepeatMasker version 1.332 (key parameters: -s -nolow -norna -nois -e rmbblast -gff) (Bedell et al., 2000) with the NCBI/RMBLAST version 2.6.0+ search engine.

Genome prediction and functional annotation

The annotation pipeline integrated different tasks starting from the fetching of raw data from public and/or private repositories (e.g., expressed sequence tags (ESTs), protein collections, RNA-Seq datasets, and other relevant sequences) to sequence alignment, prediction of gene models, and their functional annotations. We used MAKER2 (key parameters: default) (Campbell et al., 2014) as the core module for structural genome annotation. The software synthesizes a final annotation based on quality evidence values. Here, evidence was based on previously published transcripts, proteins, and data generated in this study as follows: 1) RNA-Seq datasets generated for *Q. ilex* by our research group (Guerrero-Sánchez et al., 2017; Guerrero-Sánchez et al., 2019; Guerrero-

TABLE 1 Comparison of summary statistics for genome assembly among *Quercus* species.

Species	Assembled genome size (Mpb)	Contig N50 (Mbp)	Repetitive element (% of genome)	Protein-coding gene (n)	Reference
<i>Q. ilex</i>	842.2	3.30	53.3	39,443	This study
<i>Q. robur</i>	814.3	16.00	53.3	25,808	Plomi3n et al. (2016); Plomi3n et al. (2018)
<i>Q. lobata</i>	845.9	0.97	54.4	38,373	Sork et al. (2016), Sork et al. (2022)
<i>Q. suber</i>	953.5	0.08	12.0	33,658	Ramos et al. (2018)
<i>Q. variabilis</i>	787.0	64.90	67.6	32,466	Han et al. (2022)
<i>Q. gilva</i>	889.8	28.30	57.6	36,442	Zhou et al. (2022)
<i>Q. mongolica</i>	809.8	2.40	53.8	36,553	Ai et al. (2022)
<i>Q. acutissima</i>	957.1	1.20	57.1	29,889	Liu et al. (2022)
<i>Q. dentata</i>	893.5	4.21	47.0	31,584	Wang et al. (2023)

S3nchez et al., 2021) (NCBI accession codes: SRR11050903 and SRX2993508), 2) RNA-Seq datasets from *Q. ilex* seedling leaves under drought conditions generated by Madritsch et al. (2019) (NCBI accession codes: SRX4725055 and SRX4725058) using Augustus gene models produced by Braker2 (Br3na et al., 2021), 3) 22,000 reviewed proteins from UniProtKB (taxa rosids), 4) two sets of proteins from two closely related species (*Q. lobata* and *Juglans regia*) [NCBI accession codes: GCF_001633185.2 (*Q. lobata*) and GCA_001411555.2 (*J. regia*)], 5) *Quercus* EST collection [downloaded from the NCBI using “EST” and “*Quercus*” as keywords (download date: June 2022)], and 6) the curated TE library produced by the EDTA pipeline generated in this work. tRNAs were predicted using tRNAscan-SE (key parameter: default) (Lowe and Eddy, 1997). Tools from the AGAT v0.9.1 (<https://github.com/NBISweden/AGAT>) package were used to refine the annotation. The gene set annotation was checked to find cases where different gene features have coding sequences (CDS) that overlap and then merged (agat_sp_fix_overlapping_genes.pl), as well as the longest isoform was kept (agat_sp_keep_longest_isoform.pl).

To assign functional descriptions, Gene Ontology (GO) terms, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway information to the gene models, sequences (transcripts/proteins) were functionally annotated with TRINOTATE version 2.0 (key parameters: protDB = uniprot_3745 (Viridiplantae), transdecoder + diamond + pfam) (<https://github.com/Trinotate/Trinotate/wiki>).

Chloroplast genome assembly and annotation

Among all the contigs, one (ctg011470, 142.3 Kbp) matched chloroplast sequences in the NCBI (best match “*Medicago sativa* chloroplast, complete genome,” identity 99%). Chloroplast annotation was performed using the GeSeq tool (key parameters: default) (Tillich et al., 2017). GeSeq compares input sequences against reference databases using BLAT (Kent, 2002). We used additional tools (i.e., ARAGORN, tRNA-scan, and Chlo3 version

0.1.0) to annotate tRNAs, rRNAs, and CDS. Finally, OrganellarGenomeDRAW (Greiner et al., 2019) was used to convert annotations into a circular graphical map.

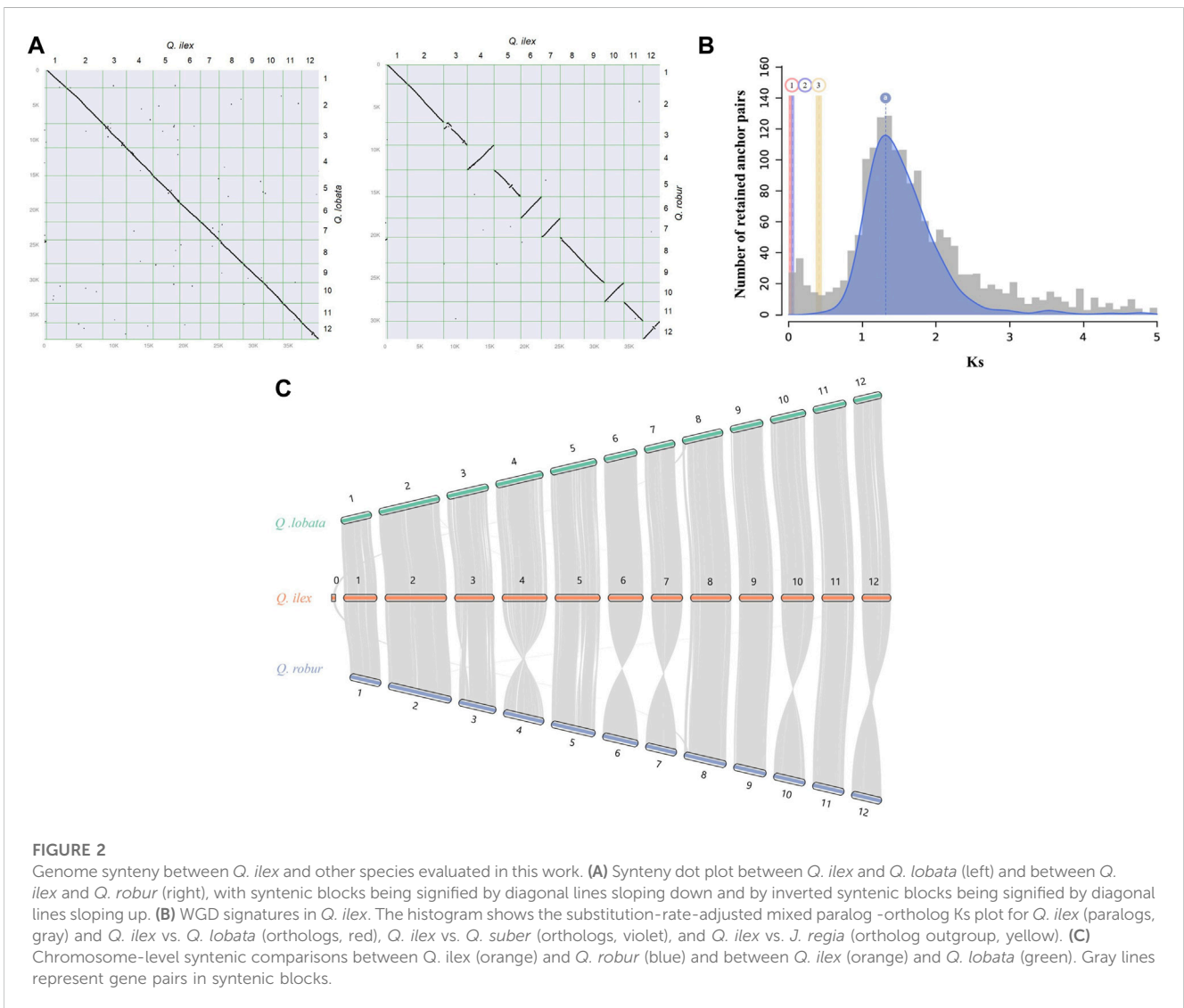
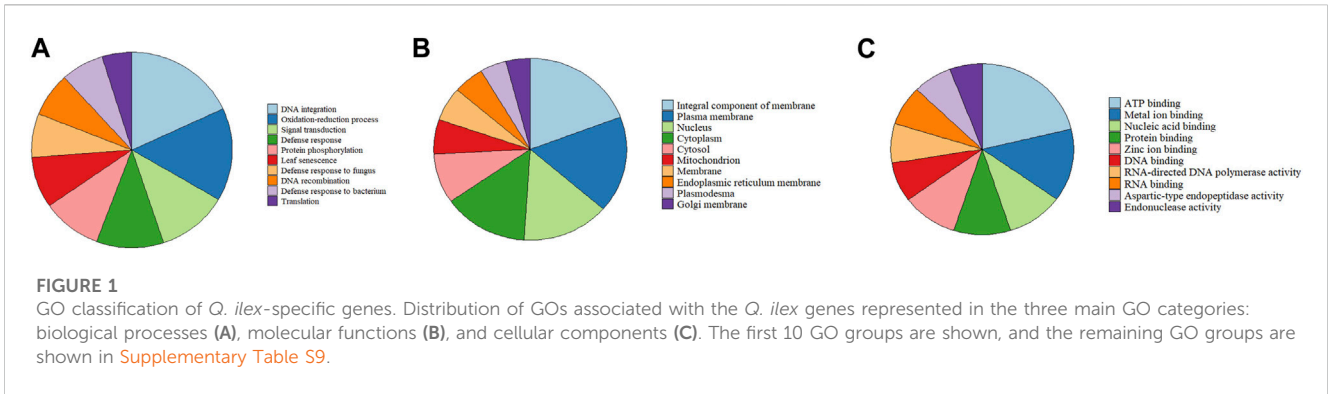
Ortholog identification

OrthoFinder version 2.5.4 (Emms and Kelly, 2015; Emms and Kelly, 2019) (key parameters: default) was used to identify orthologous and duplicated nodes across *Q. ilex* and *Q. lobata* (ValleyOak3.0), *Q. suber* (CorkOak1.0), *Q. robur* (dhQueRobu3.1), *Populus trichocarpa* (Pop_tri_v3), *J. regia* (Walnut_2.0), and *Arabidopsis thaliana* (TAIR10). Protein sequences for each species were pre-processed to extract the longest transcript per gene using an OrthoFinder script (primary_transcript.py). Trees and plots were built using the Bioconductor library “cogecq” (https://bioconductor.org/packages/release/bioc/vignettes/cogecq/inst/doc/vignette_02_assessing_orthogroup_inference.html). The tree was an output of OrthoFinder (<https://github.com/davidemms/OrthoFinder>). Following the OrthoFinder manual, species tree is a STAG species tree inferred from all orthogroups, containing STAG support values at internal nodes and rooted using STRIDE.

Genome synteny and whole-genome duplication analysis

Synteny analysis was performed between *Q. ilex* and *Q. lobata*, between *Q. ilex* and *Q. robur*, and between *Q. ilex* and *J. regia* using JCVI v1.3.3 (Tang et al., 2008) with the following parameters: -cscore = .99; -minspan = 30. Dot plots, synteny patterns, and macrosynteny visualizations were obtained using the graphics.dotplot, compara.synteny depth, and graphics.karyotype functions (<https://github.com/tanghaibao/kcvi>), respectively.

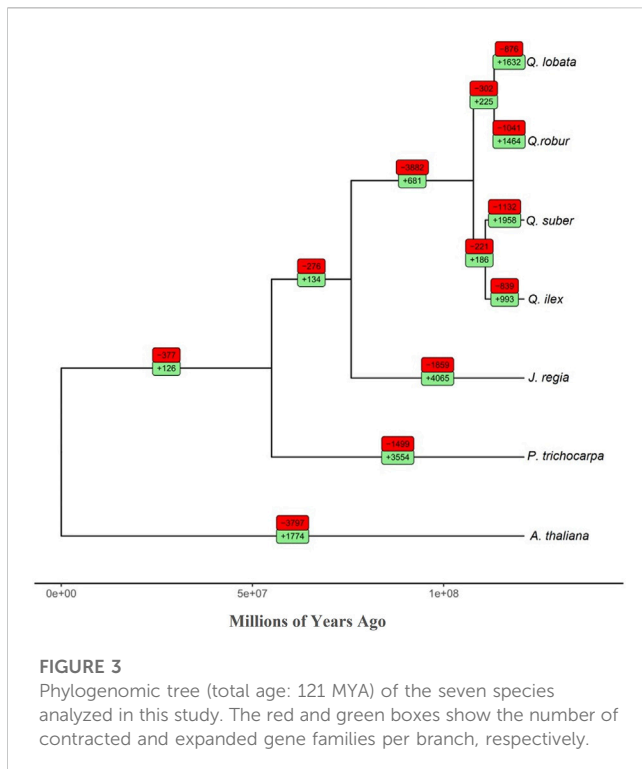
Whole-genome duplication (WGD) events were inferred across *Q. ilex* and two species of the genus *Quercus* (*Q. suber* and *Q. lobata*), with *J. regia* being used as the outgroup species. WGD analyses were performed using Ksrates v1.1.1 (key parameters: default) (Sensalari



et al., 2022). The program compares paralog and ortholog K_s distributions derived from CDS and estimates differences in synonymous substitution rates across the included species. It then generates an adjusted mixed plot of paralog and ortholog K_s distributions. This allows us to assess the relative phylogenetic positioning of presumed WGD and speciation events.

Gene family evolution analyses

In order to explore possible gene family expansions and contractions, we selected ortholog sequences from the seven species obtained through OrthoFinder (i.e., *Q. ilex*, *Q. suber*, *Q. lobata*, *Q. robur*, *J. regia*, *P. trichocarpa*, and *A. thaliana*). Since



ancestral state reconstruction requires at least two species per clade of interest, we only used gene families with gene copies present in at least two species. In addition, orthogroups with more than 100 gene copies in one or more species were removed, since big gene families cause the variance of gene copy number to be too large and lead to non-informative parameter estimates. We produced a rooted phylogenetic tree using OrthoFinder, which was used as a probabilistic model to infer family expansions and contractions. The tree was made ultrametric using an OrthoFinder accessory script and setting 121 MYA as the total age of the tree in millions of years. This value represents the median adjusted time of divergence between *A. thaliana* and the *Quercus* clade calculated from the TimeTree5 database (Kumar et al., 2022). Gene family evolution analyses were performed using the Computational Analysis of Gene Family Evolution (CAFE) algorithm (De Bie et al., 2006) with the lambda value set at $4.54495e-09$. This parameter was calculated automatically by the software as the maximum likelihood value given the gene families in the data file. CAFE computes a “family-wide *p*-value” that was used to predict contracted/expanded families.

Results and discussion

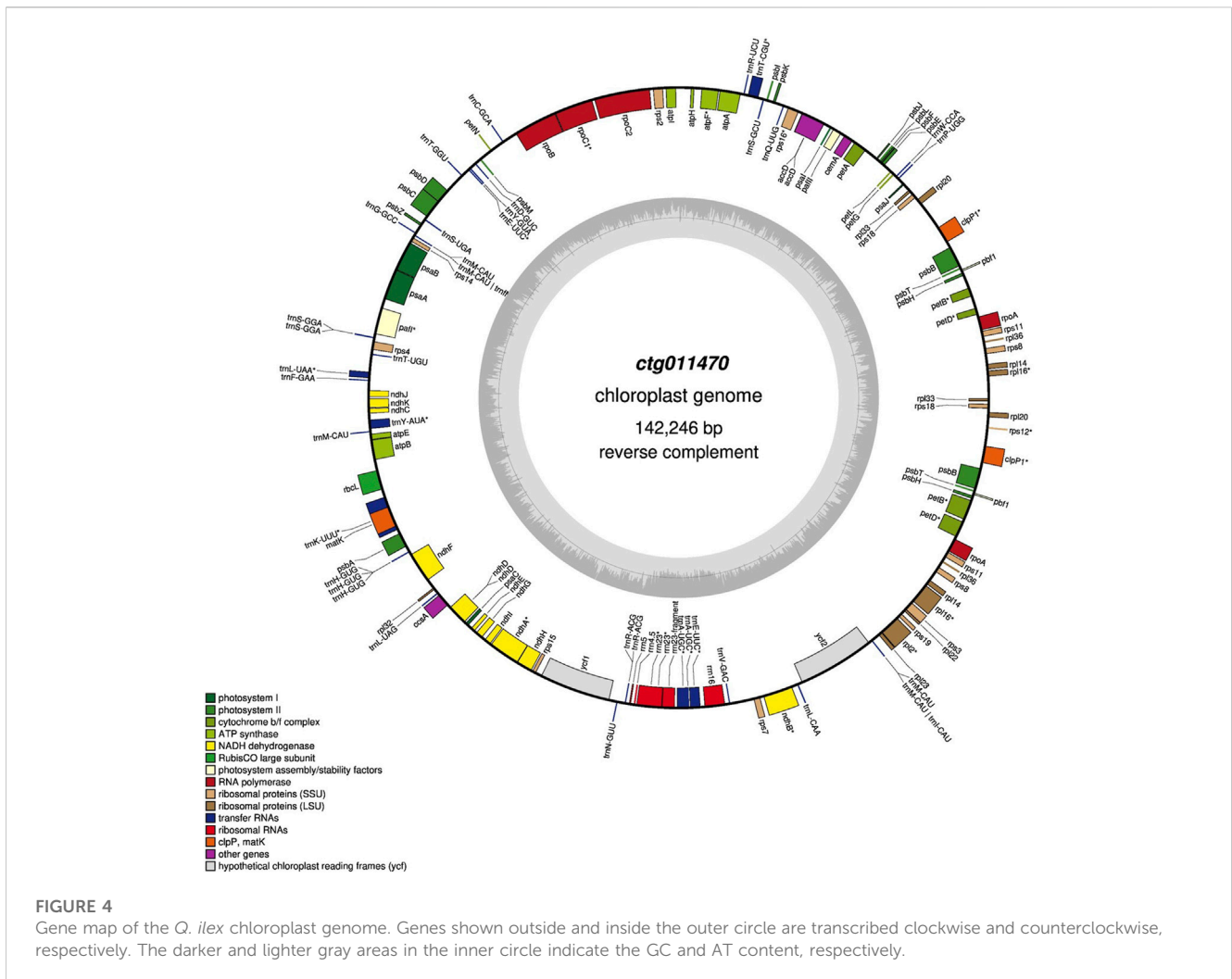
De novo *Q. ilex* genome assembly

In the genus *Quercus*, both *Q. ilex* and *Q. suber* are sympatric in the Mediterranean area, and hybrids have been identified in natural populations (Burgarella et al., 2009; López De Heredia et al., 2018). *Q. ilex* and *Q. suber* individuals clustered at different sides of the PC1 axis, indicating an allopatric speciation of this individual (Supplementary Figure S3). We employed PacBio technologies to sequence the genome of *Q. ilex* subsps. *ballota* [as defined by Flora

Ibérica (Amaral, 1983), also known as *Quercus ilex* subsp. *rotundifolia* or *Quercus rotundifolia*], resulting in the generation of a diploid reference genome. First, we assembled 1,166 contigs totaling a genome size of 1,000,594,563 bp. The N50 and the average contig sizes were 2.63 and 0.85 Mbp, respectively (Supplementary Table S1). This initial genome assembly was larger than a previous estimate using flow cytometry (935.0 Mbp) (Rey et al., 2019). This difference can be explained by regions of very high heterozygosity leading to problems during genome assembly (Kajitani et al., 2014). For example, once a pair of allelic sequences exceeded a certain threshold of nucleotide diversity, these regions were assembled as separate contigs, rather than the expected single haplotype-fused contig, resulting in a significantly larger genome size. Moreover, the initial assembly included primary, secondary, and organelle (chloroplast and mitochondria) contigs. After these were separated using Purge_haplotigs (Roach et al., 2018), we obtained a total of 530 primary contigs (total size: 842.2 Mbp) and 616 secondary contigs (total size: 157.3 Mbp), and 20 (total size: 1.1 Mbp) were flagged as artifacts (i.e., too low or too high coverage). The final assembly of the nuclear haploid genome (842.2 Mbp) (Table 1) had a contig N50 size of 3.3 Mbp, indicating a *de novo* assembly with a high level of continuity. The size of the nuclear haploid genome is similar to that of other species of the genus *Quercus* (Table 1); however, the contig N50 of *Q. ilex* is higher than that of other *Quercus* spp. such as *Q. robur* (0.07 Mb; Plomión et al., 2018), *Q. suber* (0.08 Mb; Ramos et al., 2018), *Q. acutissima* (1.44 Mb; Liu et al., 2022), *Q. lobata* (1.90 Mb; Sork et al., 2022), and *Quercus mongolica* (2.64 Mb; Ai et al., 2022). BUSCO, in the genome assembling, predicted gene models exhibiting 94.80% of complete and single-copy genes, with a low percentage of duplicated (4.20%), fragmented (0.20%), and missing (0.80%) genes (Supplementary Table S2), indicating that the assembled genome has a high quality and is nearly complete. Similar results were observed in *Q. lobata* (94%) (Sork et al., 2016), *Q. suber* (95%) (Ramos et al., 2018), and *Quercus gilva* (93.5%) (Zhou et al., 2022). A total of 493 contigs, accounting for 835 Mbp (99% of the total assembly), were found to have some correspondence on *Q. lobata* chromosomes. The relative size of each chromosome varied from 43.58 Mbp (chromosome 12) to 106.09 Mbp (chromosome 2) (Supplementary Table S3). The number of *Q. ilex* contigs assigned to each chromosome ranged from 20 (chromosome 12) to 71 (chromosome 4) (Supplementary Table S3). Less than 1% of the sequences (37 contigs, size: 6.69 Mbp) failed to be assigned, and they were merged and placed in chromosome 0 (chr 0). Arbitrary gaps of 100 bp were added between consecutive contigs (517 gap sequences, 51.7 Kbp).

Repetitive sequences

Repetitive sequences in the *Q. ilex* genome assembly accounted for 53.3% (448.68 Mb) of the total genome length (Supplementary Table S4). LTR Gypsy and Copia were the most abundant repeat types (14.37% and 11.58%, respectively). The proportion of repetitive elements in the *Q. ilex* genome was similar to that of repetitive elements in other *Quercus* genomes with 51.78% in *Q. mongolica* (Ai et al., 2022), 53.3% in *Q. robur* (Plomion et al., 2018), 53.8% in *Q. mongolica* (Ai et al., 2022), and 54.4% in *Q. lobata* (Sork



et al., 2022) (Table 1). However, this proportion was higher in the *Q. gilva* (57.57%) (Zhou et al., 2022), *Quercus variabilis* (67.7%) (Han et al., 2022), and *Quercus acutissima* (57.13%) genomes (Liu et al., 2022).

Genome prediction and functional annotation

The final gene set contained 39,443 protein-coding genes (Supplementary Tables S5, S6) and 759 tRNA genes (Supplementary Table S7), which is similar to other *Quercus* spp. (Table 1). The annotated genome showed a completeness >98% with a high degree of single-copy genes (>95%) and a low percentage of duplication (3.1%) (Supplementary Table S2). Of the total number of 39,443 genes, 35,258 (89.4%), 29,507 (74.8%), and 9,555 (24.2%) were identified with functional descriptions, GO terms, and path maps (KEGG), respectively (Supplementary Table S5). The remaining 4,185 (10.6%) genes were described without annotation. The GO representation of the whole gene set according to their biological processes, cellular components, and molecular functions is summarized in Supplementary Figure S4. Within the biological

processes, the most abundantly assigned genes were involved in protein phosphorylation, the oxidation–reduction process, and defense response. In the cellular component category, the genes fell mainly into the categories of integral components of the membrane, nucleus, and plasma membrane. In the case of the molecular functions, many genes were associated with ATP binding, metal ion binding, and protein binding.

Ortholog identification

We identified orthologous gene families by comparing the proteomes of *Q. ilex*, which were predicted in our project, with those of seven other flowering plant species, including *Q. suber*, *Q. lobata*, *Q. robur*, *J. regia*, *P. trichocarpa*, and *A. thaliana* (Supplementary Table S8). OrthoFinder assigned 35,861 orthogroups and 273,514 genes in orthogroups (out of 305,119 genes, 89.6%), with 31,605 genes remaining unassigned (10.4%). A total of 11,149 orthogroups were found to be common to all the species, and 11,807 were found to be species specific. Out of the 273,514 genes, 1,206 were found to be unique to the *Q. ilex* genome (Supplementary Table S9). The GO representation of the *Q. ilex* specific gene set according to their biological processes, cellular components, and molecular functions is shown in Supplementary Table S9. Within the biological

processes, the top three associated categories were DNA integration, the oxidation–reduction process, and signal transduction (Figure 1A). In the cellular component category, the majority of genes were classified in terms of integral components of the membrane, plasma membrane, and nucleus (Figure 1B). In the case of the molecular functions, many genes were associated with ATP binding, metal ion binding, and protein binding (Figure 1C).

Genome synteny and whole-genome duplications

In the synteny analysis between *Q. ilex* and *Q. lobata* and between *Q. ilex* and *Q. robur*, a conserved collinearity among genomic blocks and a clear one-to-one relationship of the 12 chromosomes were observed (Figures 2A, B). This analysis allowed us to infer that all misassembled scaffolds in our assembly (placed in chr0) likely belong to chromosome 8 (Figure 2B) (Sork et al., 2022). Liu et al. (2022) and Han et al. (2022) reported a largely conserved gene synteny between *Q. variabilis* and *Q. acutissima* and between *Q. variabilis* and *Q. lobata* and *Q. robur*, respectively. Therefore, this suggests little rearrangement of chromosomes of the genus *Quercus*, indicating conserved evolution in Fagaceae.

The potential WGD events in the evolutionary history of *Q. ilex* were studied by the distribution of the Ks between homologous gene pairs derived from *Q. ilex*, *Q. lobata*, *Q. suber*, and *J. regia*. One main peak was observed in the *Q. ilex* genome based on the abundance of Ks value (a Ks value of 1.31), indicating a single major WGD event (Figure 2C) (Han et al., 2022; Zhou et al., 2022). The low Ks values between *Q. ilex* and *Q. lobata* (0.04–0.06 Ks units) and between *Q. ilex* and *Q. suber* (0.02–0.04 Ks units) indicated little divergence between these three species, supported by the higher divergence with *J. regia* (0.41–0.42 Ks units) (Figure 2C).

Expanded and contracted gene families

We analyzed gene family expansions and contractions in the seven species used in the OrthoFinder analysis. We found that the ratio between contracted versus expanded gene families in *Q. ilex* (839/993) was much lower than in the other *Quercus* spp. included in our analysis (*Q. suber*, 1132/1958; *Q. robur*, 1041/1464; and *Q. lobata*, 876/1632) (Figure 3). In addition, this ratio was also lower than that in *Q. acutissima* (2390/3897) (Liu et al., 2022). *Q. ilex* exhibited a notable enrichment of expanded genes involved in “glutathione metabolism,” “plant hormone signal transduction,” “oxidative phosphorylation,” “peptidases and inhibitors,” “phenylpropanoid biosynthesis,” and “sesquiterpenoid and triterpenoid biosynthesis,” while contracted genes were enriched in “glycosyltransferases” and “ribosome” (Supplementary Table S10).

Chloroplast genome assembly and annotation

One contig (ctg011470, ~140 Kbp) showed high homology with *Medicago sativa* chloroplast sequences, and five contigs (ctg008800, ctg009160, ctg009300, ctg009310, and ctg011650) showed homology

with mitochondrial sequences of different plant species (*Q. robur*, *Q. acutissima*, and *Lithocarpus litseifolius*). The final chloroplast assembly was 142.3 Kbp in size with a GC content of 33.5%. We annotated 149 genes of which 92 were protein-coding genes, 51 were rRNAs, and six were tRNAs (Figure 4). A comparison of features across 28 chloroplast genomes of the genus *Quercus* is included in Supplementary Table S11. All the chloroplast genomes exhibited a consistent size, except for *Q. ilex*, which displayed a smaller genome size. The number of genes ranged from 113 (e.g., *Quercus wutaishanica*, *Quercus rubra*, and *Quercus macrocarpa*, among others) to 140 (*Q. ilex*). In terms of GC content, *Q. ilex* exhibited the lowest value, while *Q. lobata* displayed the highest.

Conclusion

We generated the first draft of the *Q. ilex* genome with a high level of completeness and continuity. The final assembly is 842.2 Mb which is similar to other *Quercus* species. We found no strong evidence of chromosome fusion or species-specific WGD events, and there was high synteny between *Q. ilex* chromosomes and those of *Q. lobata* and *Q. robur*. Gene evolution analysis revealed that *Q. ilex* experienced a much lower frequency of gene family expansions if compared to other *Quercus* species, although the reason behind this pattern remains unclear. This assembly could be improved in the future by the inclusion of sequencing data from other platforms such as Illumina and Oxford Nanopore. However, the available first draft of the assembled genome, together with previously published transcriptomics, proteomics, and metabolomics datasets (Guerrero-Sánchez et al., 2021; López-Hidalgo et al., 2021; San-Eufrasio et al., 2021; Maldonado-Alconada et al., 2022; Tienda-Parrilla et al., 2022), provides valuable information for future studies on comparative genomics, adaptation, hybridization, and epigenetics that provide the response to the changed environment as well as nutritional quality in acorns of *Q. ilex*.

Data availability statement

The raw sequencing data were deposited in the NCBI SRA database under the BioProject accession number PRJNA687489 for raw data and accession number SAMN17145625 JAVFWQ000000000 for the assembled genomes (nuclear + chloroplast).

Author contributions

JJ-N: sample collection. M-DR: genomic DNA extraction. ML-O: microsatellite analysis. VR: whole-genome sequencing analysis. M-DR, ML-O, VG-S, RC, MC, and JJ-N: data interpretation. M-DR, RC, and JJ-N: manuscript writing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by an ENCINOMICS-2 PID 2019-109038RB-100 grant from the Spanish Ministry of Economy and

Competitiveness, UCO_FEDER-18-12575795R and PROYEXCEL_00881 grants from the Junta de Andalucía.

Acknowledgments

MC and M-DR are grateful for the award of a Ramón y Cajal (RYC-2017-23706) and a Juan de la Cierva-Incorporación (IJC2018-035272-I) contract, respectively, by the Spanish Ministry of Science, Innovation, and Universities. They thank Bruce Kingham and Olga Shevchenko at the Sequencing and Genotyping Center at the University of Delaware DNA for assistance and sequencing of the raw data of the *Q. ilex* genome. They also thank Richard J. Buggs and Antonio Rodríguez-Franco for assistance in data interpretation and Ignacio Rivera for harvesting leaf samples.

Conflict of interest

Author VR was employed by Biomeets Consulting ITNIG—Carrer d' Alaba 61 08005 Catalonia.

References

- Ai, W., Liu, Y., Mei, M., Zhang, X., Tan, E., Liu, H., et al. (2022). A chromosome-scale genome assembly of the Mongolian oak (*Quercus mongolica*). *Mol. Ecol. Resour.* 22, 2396–2410. doi:10.1111/1755-0998.13616
- Alexander, L. W., and Woeste, K. E. (2014). Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. *Tree Genet. Genomes* 10, 803–812. doi:10.1007/s11295-013-0681-1
- Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., et al. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol.* 20, 224–317. doi:10.1186/s13059-019-1829-6
- Amaral, J. (1983). "Quercus," in *Flora iberica II: 15-20*. Editors S. Castroviejo, M. Laínz, G. López, P. Montserrat, F. Muñoz, J. Paiva, et al. (Madrid, Spain: Real Jardín Botánico. Consejo Superior de Investigaciones Científicas).
- Backs, J. R., and Ashley, M. V. (2021). *Quercus* genetics: insights into the past, present, and future of oaks. *Forests* 12, 1628. doi:10.3390/f12121628
- Barbata, A., and Peñuelas, J. (2016). Sequence of plant responses to droughts of different timescales: lessons from holm oak (*Quercus ilex*) forests. *Plant Ecol. Divers.* 9, 321–338. doi:10.1080/17550874.2016.1212288
- Bedell, J. A., Korf, I., and Gish, W. (2000). MaskerAid: A performance enhancement to RepeatMasker. *Bioinformatics* 16, 1040–1041. doi:10.1093/bioinformatics/16.11.1040
- Brasier, C. M. (1992). Oak tree mortality in Iberia. *Nature* 360, 539. doi:10.1038/360539a0
- Brúna, T., Hoff, K. J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genom. bioinform.* 3, lqaa108. doi:10.1093/nargab/lqaa108
- Burgarella, C., Lorenzo, Z., Jabbour-Zahab, R., Lumaret, R., Guichoux, E., Petit, R. J., et al. (2009). Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity* 102, 442–452. doi:10.1038/hdy.2009.8
- Camilo-Alves, C. S. P., Clara, M. I. E., and Ribeiro, N. M. C. A. (2013). Decline of mediterranean oak trees and its association with *Phytophthora cinnamomi*: A review. *Eur. J. For. Res.* 132, 411–432. doi:10.1007/s10342-013-0688-z
- Campbell, M. S., Law, M., Holt, C., Stein, J. C., Moghe, G. D., Hufnagel, D. E., et al. (2014). MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* 164, 513–524. doi:10.1104/pp.113.230144
- Cortés, A. J., Restrepo-Montoya, M., and Bedoya-Canas, L. E. (2020). Modern strategies to assess and breed forest tree adaptation to changing climate. *Front. Plant Sci.* 11, 583323. doi:10.3389/fpls.2020.583323
- De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). Cafe: A computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. doi:10.1093/bioinformatics/btl097
- De Rigo, D., and Caudullo, G. (2016). "Quercus ilex in europe: distribution, habitat, usage and threats," in *Book: European atlas of forest tree species publisher* (Luxembourg: Publications Office of the European Union), 152–153.
- Díaz-Caro, C., García-Torres, S., Elghannam, A., Tejerina, D., Mesias, F. J., and Ortiz, A. (2019). Is production system a relevant attribute in consumers' food preferences? The case of iberian dry-cured ham in Spain. *Meat Sci.* 158, 107908. doi:10.1016/j.meatsci.2019.107908
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238–314. doi:10.1186/s13059-019-1832-y
- Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157. doi:10.1186/s13059-015-0721-2
- Feng, L., Yang, X., Jiao, Q., Wang, C., and Yin, Y. (2020). The complete chloroplast genome of *Quercus robur* 'Fastigiata'. *Mitochondrial DNA B Resour.* 5, 129–130. doi:10.1080/23802359.2019.1692724
- Fernández i Martí, A., Romero-Rodríguez, C., Navarro-Cerrillo, R. M., Abril, N., Jorriñ-Novo, J. V., and Dodd, R. S. (2018). Population genetic diversity of *Quercus ilex* subsp. *ballota* (Desf.) samp. reveals divergence in recent and evolutionary migration rates in the Spanish dehesas. *Forests* 9, 337. doi:10.3390/f9060337
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi:10.1093/nar/gkz238
- Guerrero-Sánchez, V. M., Castillejo, M. Á., López-Hidalgo, C., Maldonado-Alconada, A. M., Jorriñ-Novo, J. V., and Rey, M. D. (2021). Changes in the transcript and protein profiles of *Quercus ilex* seedlings in response to drought stress. *J. Proteom.* 243, 104263. doi:10.1016/j.jprot.2021.104263
- Guerrero-Sanchez, V. M., Maldonado-Alconada, A. M., Amil-Ruiz, F., and Jorriñ-Novo, J. V. (2017). Holm oak (*Quercus ilex*) transcriptome. *De novo* sequencing and assembly analysis. *Front. Mol. Biosci.* 4, 70. doi:10.3389/fmolb.2017.00070
- Guerrero-Sanchez, V. M., Maldonado-Alconada, A. M., Amil-Ruiz, F., Verardi, A., Jorriñ-Novo, J. V., and Rey, M. D. (2019). Ion Torrent and Illumina, two complementary RNA-seq platforms for constructing the holm oak (*Quercus ilex*) transcriptome. *PLoS One* 14, e0210356. doi:10.1371/journal.pone.0210356
- Han, B., Wang, L., Xian, Y., Xie, X. M., Li, W. Q., Zhao, Y., et al. (2022). A chromosome-level genome assembly of the Chinese cork oak (*Quercus variabilis*). *Front. Plant Sci.* 13, 1001583. doi:10.3389/fpls.2022.1001583
- Hu, H. L., Zhang, J. Y., Li, Y. P., Xie, L., Chen, D. B., Li, Q., et al. (2019). The complete chloroplast genome of the daimyo oak, *Quercus dentata* Thunb. *Conserv. Genet. Resour.* 11, 409–411. doi:10.1007/s12686-018-1034-z

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1242943/full#supplementary-material>

- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi:10.1093/bioinformatics/btn129
- Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 24, 1384–1395. doi:10.1101/gr.170720.113
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi:10.1101/gr.229202
- Kumar, S., Suleski, M., Craig, J. M., Kaspróvicz, A. E., Sanderford, M., Li, M., et al. (2022). TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* 39, msac174. doi:10.1093/molbev/msac174
- Laureano, R. G., García-Nogales, A., Seco, J. I., Linares, J. C., Martínez, F., and Merino, J. (2016). Plant maintenance and environmental stress. Summarising the effects of contrasting elevation, soil, and latitude on *Quercus ilex* respiration rates. *Plant Soil* 409, 389–403. doi:10.1007/s11104-016-2970-6
- Leroy, T., Plomión, C., and Kremer, C. (2020). Oak symbolism in the light of genomics. *New Phytol.* 226, 1012–1017. doi:10.1111/nph.15987
- Li, X., Li, Y., Zang, M., Li, M., and Fang, Y. (2018). Complete chloroplast genome sequence and phylogenetic analysis of *Quercus acutissima*. *Int. J. Mol. Sci.* 19, 2443. doi:10.3390/ijms19082443
- Liu, D., Xie, X., Tong, B., Zhou, C., Qu, K., Guo, H., et al. (2022). A high-quality genome assembly and annotation of *Quercus acutissima* Carruth. *Front. Plant Sci.* 13, 1068802. doi:10.3389/fpls.2022.1068802
- Liu, X., Chang, E. M., Liu, J. F., Huang, Y. N., Wang, Y., Yao, N., et al. (2019). Complete chloroplast genome sequence and phylogenetic analysis of *Quercus bawanglingensis* Huang, Li et Xing, a vulnerable oak tree in China. *Forests* 10, 587. doi:10.3390/f10070587
- López De Heredia, U., Sanchez, H., and Soto, A. (2018). Molecular evidence of bidirectional introgression between *Quercus suber* and *Quercus ilex*. *IFOREST* 11, 338–343. doi:10.3832/ifor2570-011
- López-Hidalgo, C., Menéndez, M., Jorrin-Novo, J. V., and Jorrin-Novo, J. V. (2021). Phytochemical composition and variability in *Quercus ilex* acorn morphotypes as determined by NIRS and MS-based approaches. *Food Chem.* 338, 127803. doi:10.1016/j.foodchem.2020.127803
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi:10.1093/nar/25.5.955
- Lu, S., Hou, M., Du, F. K., Li, J., and Yin, K. (2016). Complete chloroplast genome of the oriental white oak: *quercus aliena* blume. *Mitochondrial DNA A Resour.* 27, 2802–2804. doi:10.3109/19401736.2015.1053074
- Madritsch, S., Wischnitzki, E., Kotrade, P., Ashoub, A., Burg, A., Fluch, S., et al. (2019). Elucidating drought stress tolerance in European oaks through cross-species transcriptomics. *G3 Genes. Genomes Genet.* 9, 3181–3199. doi:10.1534/g3.119.400456
- Maldonado-Alconada, A. M., Castillejo, M. Á., Rey, M. D., Labella-Ortega, M., Tienda-Parrilla, M., Hernández-Lao, T., et al. (2022). Multiomics molecular research into the recalcitrant and orphan *Quercus ilex* tree species: why, what for and how. *Int. J. Mol. Sci.* 23, 9980. doi:10.3390/ijms23179980
- Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4325. doi:10.1093/nar/8.19.4321
- Muthuramalingam, P., Jeyasri, R., Rakkammal, K., Satish, L., Shamili, S., Karthikeyan, A., et al. (2022). Multi-omics and integrative approach towards understanding salinity tolerance in rice: A review. *Biology* 11, 1022. doi:10.3390/biology11071022
- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R., Hellinga, A. J., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20, 275–318. doi:10.1186/s13059-019-1905-y
- Paajanen, P., Kettleborough, G., López-Girona, E., Giolai, M., Heavens, D., Baker, D., et al. (2019). A critical comparison of technologies for a plant genome sequencing project. *Gigascience* 8, giy163. doi:10.1093/gigascience/giy163
- Pang, X., Liu, H., Wu, S., Yuan, Y., Li, H., Dong, J., et al. (2019). Species identification of oaks (*Quercus* L., Fagaceae) from gene to genome. *Int. J. Mol. Sci.* 20, 5940. doi:10.3390/ijms20235940
- Plieninger, T., Flinzberger, L., Hetman, M., Horstmannshoff, I., Reinhard-Kolempas, M., Topp, E., et al. (2021). Dehesas as high nature value farming systems: A social-ecological synthesis of drivers, pressures, state, impacts, and responses. *Ecol. Soc.* 26, art23. doi:10.5751/ES-12647-260323
- Plomión, C., Aury, J. M., Amselem, J., Alaeitabar, T., Barbe, V., Belsler, C., et al. (2016). Decoding the oak genome: public release of sequence data, assembly, annotation and publication strategies. *Mol. Ecol. Resour.* 16, 254–265. doi:10.1111/1755-0998.12425
- Plomión, C., Aury, J. M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., et al. (2018). Oak genome reveals facets of long lifespan. *Nat. Plants* 4, 440–452. doi:10.1038/s41477-018-0172-3
- Ramos, A. M., Usié, A., Barbosa, P., Barros, P. M., Capote, T., Chaves, I., et al. (2018). The draft genome sequence of cork oak. *Sci. Data* 5, 180069. doi:10.1038/sdata.2018.69
- Rey, M. D., Castillejo, M. Á., Sánchez-Lucas, R., Guerrero-Sanchez, V. M., López-Hidalgo, C., Romero-Rodríguez, C., et al. (2019). Proteomics, holm oak (*Quercus ilex* L.) and other recalcitrant and orphan forest tree species: how do they see each other? *Int. J. Mol. Sci.* 20, 692. doi:10.3390/ijms20030692
- Roach, M. J., Schmidt, S. A., and Borneman, A. R. (2018). Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinform* 19, 460–510. doi:10.1186/s12859-018-2485-7
- Roberts, R. J., Carneiro, M. O., and Schatz, M. C. (2013). The advantages of SMRT sequencing. *Genome Biol.* 14, 405. doi:10.1186/gb-2013-14-6-405
- Ruiz Gómez, F. J., Pérez-de-Luque, A., Sánchez-Cuesta, R., Quero, J. L., and Navarro-Cerrillo, R. M. (2018). Differences in the response to acute drought and *Phytophthora cinnamomi* Rands infection in *Quercus ilex* L. seedlings. *Forests* 9, 634. doi:10.3390/f9100634
- San-Eufrasio, B., Castillejo, M. Á., Labella-Ortega, M., Ruiz-Gómez, F. J., Navarro-Cerrillo, R. M., Tienda-Parrilla, M., et al. (2021). Effect and response of *Quercus ilex* subsp. *ballota* [Desf.] Samp. seedlings from three contrasting Andalusian populations to individual and combined *Phytophthora cinnamomi* and drought stresses. *Front. Plant Sci.* 12, 722802. doi:10.3389/fpls.2021.722802
- Sensalari, C., Maere, S., and Lohaus, R. (2022). ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. *Bioinformatics* 38, 530–532. doi:10.1093/bioinformatics/btab602
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi:10.1093/bioinformatics/btv351
- Sork, V. L., Cokus, S. J., Fitz-Gibbon, S. T., Zimin, A. V., Puiui, D., Garcia, J. A., et al. (2022). High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. *Nat. Commun.* 13, 2047–2115. doi:10.1038/s41467-022-29584-y
- Sork, V. L., Fitz-Gibbon, S. T., Puiui, D., Crepeau, M., Gugger, P. F., Sherman, R., et al. (2016). First draft assembly and annotation of the genome of a California Endemic oak *Quercus lobata* Née (Fagaceae). *G3 Genes. Genomes Genet.* 6, 3485–3495. doi:10.1534/g3.116.030411
- Stocks, J. J., Metheringham, C. L., Plumb, W. J., Lee, S. J., Kelly, L. J., Nichols, R. A., et al. (2019). Genomic basis of European ash tree resistance to ash dieback fungus. *Nat. Ecol. Evol.* 3, 1686–1696. doi:10.1038/s41559-019-1036-6
- Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi:10.1126/science.1153917
- Tienda-Parrilla, M., López-Hidalgo, C., Guerrero-Sanchez, V. M., Infantes-González, Á., Valderrama-Fernández, R., Castillejo, M. Á., et al. (2022). Untargeted MS-based metabolomics analysis of the responses to drought stress in *Quercus ilex* L. leaf seedlings and the identification of putative compounds related to tolerance. *Forests* 13, 551. doi:10.3390/f13040551
- Tillich, M., Lehwerk, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi:10.1093/nar/gkx391
- Vinha, A. F., Costa, A. S. G., Barreira, J. C., Pacheco, R., and Oliveira, M. B. P. (2016). Chemical and antioxidant profiles of acorn tissues from *quercus* spp.: potential as new industrial raw materials. *Ind. Crops Prod.* 94, 143–151. doi:10.1016/j.indcrop.2016.08.027
- Wang, T. R., Wang, Z. W., Song, Y. G., and Kozłowski, G. (2021). The complete chloroplast genome sequence of *Quercus ninggangensis* and its phylogenetic implication. *Plant Fungal Syst.* 66, 155–165. doi:10.35535/pfsyst-2021-0014
- Wang, W. B., Xiang-Feng, H., Yan, X. M., Ma, B., Lu, C. F., Wu, J., et al. (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.* 238, 2016–2032. doi:10.1111/nph.18814
- Xia, E. H., Zhang, H. B., Sheng, J., Li, K., Zhang, Q. J., Kim, C., et al. (2017). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol. plant* 10, 866–877. doi:10.1016/j.molp.2017.04.002
- Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., and Zhao, G. (2016). Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front. Plant Sci.* 7, 959. doi:10.3389/fpls.2016.00959
- Yang, Y., Zhou, T., Zhu, J., Zhao, J., and Zhao, G. (2018a). Characterization of the complete plastid genome of *Quercus tarokoensis*. *Conserv. Genet. Resour.* 10, 191–193. doi:10.1007/s12686-017-0796-z
- Yang, Y., Zhu, J., Feng, L., Zhou, T., Bai, G., Yang, J., et al. (2018b). Plastid genome comparative and phylogenetic analyses of the key genera in Fagaceae: highlighting the effect of codon composition bias in phylogenetic inference. *Front. Plant Sci.* 9, 82. doi:10.3389/fpls.2018.00082
- Zhao, D., Hamilton, J. P., Pham, G. M., Crisovan, E., Wiegert-Rininger, K., Vaillancourt, B., et al. (2017). De novo genome assembly of *Camptotheca acuminata*, a natural source of the anti-cancer compound camptothecin. *Gigascience* 6, 1–7. doi:10.1093/gigascience/gix065
- Zhou, X., Liu, N., Jiang, X., Qin, Z., Farooq, T. H., Cao, F., et al. (2022). A chromosome-scale genome assembly of *Quercus gilva*: insights into the evolution of quercus section cyclobalanopsis (Fagaceae). *Front. Plant Sci.* 13, 1012277. doi:10.3389/fpls.2022.1012277