



## OPEN ACCESS

EDITED BY  
Gabriele Casazza,  
University of Genoa, Italy

REVIEWED BY  
Juan Pablo Ortiz Brunel,  
University of Guadalajara, Mexico  
Carlos I. Arbizu,  
Universidad Nacional Toribio Rodríguez de  
Mendoza de Amazonas, Peru

\*CORRESPONDENCE  
Xuemei Zhang  
✉ zhangmei103127@sina.com

RECEIVED 09 May 2024  
ACCEPTED 31 July 2024  
PUBLISHED 19 August 2024

CITATION  
Huang K, Li B, Chen X, Qin C and Zhang X  
(2024) Comparative and phylogenetic analysis  
of chloroplast genomes from ten species in  
*Quercus* section *Cyclobalanopsis*.  
*Front. Plant Sci.* 15:1430191.  
doi: 10.3389/fpls.2024.1430191

COPYRIGHT  
© 2024 Huang, Li, Chen, Qin and Zhang. 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.

# Comparative and phylogenetic analysis of chloroplast genomes from ten species in *Quercus* section *Cyclobalanopsis*

Ke Huang, Buyu Li, Xiaoli Chen, Chun Qin and Xuemei Zhang\*

College of Life Sciences, China West Normal University, Nanchong, China

The genus *Quercus* L. is widely acknowledged as a significant assemblage within East Asia tropical and subtropical broadleaf evergreen forests, possessing considerable economic importance. Nevertheless, the differentiation of *Quercus* species is deemed arduous, and the interrelations among these species remain enigmatic. Leveraging Illumina sequencing, we undertook the sequencing and assembly of the chloroplast (cp) genomes of seven species belonging to *Quercus* section *Cyclobalanopsis* (*Quercus argyrotricha*, *Q. augustinii*, *Q. bambusifolia*, *Q. bella*, *Q. edithiae*, *Q. jenseniana*, and *Q. poilanei*). Furthermore, we collated three previously published cp genome sequences of *Cyclobalanopsis* species (*Q. litseoides*, *Q. obovatifolia*, and *Q. saravanensis*). Our primary objective was to conduct comparative genomics and phylogenetic analyses of the complete cp genomes of ten species from *Quercus* section *Cyclobalanopsis*. This investigation unveiled that *Quercus* species feature a characteristic circular tetrad structure, with genome sizes ranging from 160,707 to 160,999 base pairs. The genomic configuration, GC content, and boundaries of inverted repeats/single copy regions exhibited marked conservation. Notably, four highly variable hotspots were identified in the comparative analysis, namely *trnK-rps16*, *psbC-trnS*, *rbcl-accD*, and *ycf1*. Furthermore, three genes (*atpF*, *rpoC1*, and *ycf2*) displayed signals of positive selection pressure. Phylogenetic scrutiny revealed that the four sections of *Cyclobalanopsis* clustered together as sister taxa. The branch support values ranged from moderate to high, with most nodes garnering 100% support, underscoring the utility of cp genomic data in elucidating the relationships within the genus. Divergence time analysis revealed that Section *Cyclobalanopsis* represents the earliest type of *Quercus* genus. The outcomes of this investigation establish a foundation for forthcoming research endeavors in taxonomy and phylogenetics.

## KEYWORDS

chloroplast (cp) genome, phylogenetic analysis, *Quercus* section *Cyclobalanopsis*, DNA sequencing, sequence annotation

## Introduction

*Quercus* L. is the most diverse genus in *Fagaceae* (Manos et al., 1999), with 430 species worldwide, making it one of the most widely distributed woody genera in the Northern Hemisphere. It has long been closely associated with human ecology (Welter et al., 2012; Gil-Pelegrín et al., 2017). From an ecological conservation and balance perspective, these plants are crucial in maintaining soil structure and fertility, regulating local climatic conditions, and contributing to biodiversity. In terms of economic value and practical applications, the physical properties of these plants, such as high hardness, strength, and good corrosion resistance, provide a solid basis for their use in various industries. Whether the systematic classification of cycads belongs to the genus *Quercus* or is a separate genus within the family *Fagaceae* has long been debated. The classification of the genus *Quercus* is challenging due to its complex evolutionary history, and it remains an ongoing endeavor (Manos et al., 1999; Kremer et al., 2010; Deng et al., 2018). A novel global phylogenetic framework for the genus *Quercus* has been developed to address this challenge, utilizing nuclear molecular markers and pollen characteristics. This framework divides the genus into two subgenera, *Quercus* and *Cerris*, each further divided into eight sections (Denk et al., 2017). China includes all four previously recognized sections: *Quercus*, *Cerris*, *Cyclobalanopsis*, and *Ilex* (Hubert et al., 2014; Vitelli et al., 2017). Yet, the inclusion of Chinese species in this structure is restricted and significantly less than that of species from North America and Europe.

Historically, the classification of the genus *Quercus* has traditionally been based on morphological features, leaf epidermis characteristics, pollen analysis, and the study of evolutionary processes in a limited group of species (Oh and Manos, 2008; Denk and Grimm, 2009; Deng et al., 2014; Muthreich et al., 2020). Contradictions arose when molecular sequence data were introduced to delineate (sub) sections and series, leading to discrepancies with the classifications previously established through morphological analysis (Denk et al., 2017). For example, investigations using ITS sequences have revealed a close genetic relationship between species grouped under the compound trichome base (CTB) within the section *Cyclobalanopsis* and those in the section *Cerris*, which contradicts conventional morphological taxonomy (Deng et al., 2014). This divergence highlights the lack of a unified approach to classifying the genus *Quercus*. Despite extensive morphological research on *Cyclobalanopsis*, additional molecular data is essential to fully understand interspecific relationships and the broader phylogenetic framework within the genus, especially given the observed gene introgression and similarities in leaf characteristics among different sections (Li et al., 2024).

Chloroplasts, renowned for their semi-autonomous nature within the cell, play crucial roles in photosynthesis and synthesizing essential compounds. Studies have confirmed the presence of a distinct genetic system within these organelles (Bendich, 1987; Neuhaus and Emes, 2000). In flowering plants, the cp DNA's structure, composition, and layout remain consistent across various plant families. Yet, within species, variations are

observed that result in the alteration of genes and introns, including their acquisition or loss, and structural modifications such as expansions, contractions, and reversals at the inverted repeat regions. Analyzing cp DNA contrasts offers invaluable insights into plant phylogeny and evolutionary trajectories. Furthermore, the cp genome is now recognized as a superior genomic resource for elucidating evolutionary links and species-specific diversities over traditional taxonomic methodologies (Liu X. et al., 2019).

Since the release of the *Q. rubra* cp genome in 2014, there has been a notable rise in the sequencing of cp genomes across different oak species. This pattern showcases rapid advancements in next-generation sequencing technologies as well as a growing enthusiasm for using cp genomes to establish evolutionary relationships (Alexander and Woeste, 2014; Jiang et al., 2021; Wang et al., 2021; Yang et al., 2021; Li et al., 2024). The cp genomes of only 16 species within the *Cyclobalanopsis* section, a subgenus within *Quercus*, have been fully sequenced and analyzed. The limited availability of information hinders the widespread use of phylogenetic studies and molecular identification within the *Quercus* genus. Therefore, expanding the research on cp genome sequences is imperative to clarify taxonomic ambiguities within *Quercus*.

In this study, we newly published the cp genome sequences of seven *Quercus* section *Cyclobalanopsis* species, including *Q. argyrotricha*, *Q. augustinii*, *Q. bambusifolia*, *Q. bella*, *Q. edithiae*, *Q. jenseniana*, and *Q. poilanei*. Then we downloaded three already published data from NCBI. They are: *Q. litseoides*, *Q. obovatifolia*, and *Q. saravanensis*. Using these ten cp genomes, we performed (1) the structure and gene annotation, (2) comparative genomics analyses, (3) selection pressures, (4) phylogenetic analyses and divergence time estimation. Our study will provide valuable material for phylogenetic analyses of *Quercus* section *Cyclobalanopsis* (Chen X. et al., 2023).

## Materials and methods

### Plant material and DNA sequencing

Tender, unwounded leaf of seven *Quercus* section *Cyclobalanopsis* species (*Quercus argyrotricha*, *Q. augustinii*, *Q. bambusifolia*, *Q. bella*, *Q. edithiae*, *Q. jenseniana* and *Q. poilanei*) were harvested from four provinces in China: Yunnan, Hainan, Guangxi and Guizhou. Silica gel was used to dry the materials collected. Voucher specimens were saved at China West Normal University (CWNU), and sample information is listed in [Supplementary Table S1](#). Associate Professor Xuemei Zhang of CWNU identified these specimens. Shanghai Tianhao Genomics Company carried out the extraction and sequencing of cp genomic DNA. The extracted DNA underwent double-ended sequencing through Illumina NovaSeq platforms. Quality control on the raw data used FastQC (Lo and Chain, 2014). For junction-dependent sequencing technologies (e.g., Illumina), special attention must be paid to removing contaminating junctions. Tainted junctions are located at the 3' end of the sequencing reads. So we use Adapter

Removal to remove junction contamination at the 3' end; then we use the sliding window method for quality filtering.

## Cp genome assembly and genes annotation

The filtered readings were assembled using GetOrganelle v1.7.6.1 (Jin et al., 2020) to obtain the cp genome, with *Q. kerrii* (sequence number OP679796.1) as the reference sequence. The assembled cp genome was then annotated using GAVAS2 (<http://47.96.249.172:16019/analyzer/annotate>) (Shi et al., 2019). To ensure accuracy, the annotation process included manual corrections using Geneious R9.0.2 (Kearse et al., 2012) software, which incorporated start and stop codon positions and intron and exon boundaries. To enhance the precision of the annotation outcomes, a comparative study was conducted with closely related species. The annotated sequences of the cp genome were then uploaded to the U.S. National Center for Biotechnology Information (NCBI) database for analysis. The cp genome sequences map was drawn on OGDRAW (<http://47.96.249.172:16085/CPGView/home>) (Lohse et al., 2007).

## SSR and sporadic repeat sequences

The study examined scattered repetitive sequences in the cp genomes of ten different sections of *Cyclobalanopsis* species using the online software REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) (Kurtz et al., 2001). The analysis investigated forward (F), reverse (R), palindrome (P), and complementary (C) repeats. Various parameters were employed, including a minimum repeat length of 30 and a Hamming distance of 3, requiring repeat sequences to have a minimum similarity of 90%. The default settings were used for the other parameters, with 1000 parameters being configured. For the analyses Simple Sequence Repeats (SSRs), MISA (<https://webblast.ipk-gatersleben.de/misa/>) (Beier et al., 2017) software was utilized. Different cut-off points were used for diverse nucleotide repeats, encompassing parameters like 1-10 (single nucleotide repeats happening at least ten times), 2-5, 3-4, 4-3, 5-3, and 6-3 for SSR analysis. All the other parameters stayed in their default states. Manual verification was carried out on all the analyzed repetitions, and any redundant outcomes were removed (Li et al., 2024).

## Codon bias analysis

This research initially sifted through 52 distinctive non-repetitive sequences, each surpassing 300 base pairs and containing the ATG start codon, to get ready for subsequent analysis. We used the CodonW 1.4.2 program (Peden, 1997) to calculate various codon usage indices and base composition statistics for each coding sequence. The study involved analyzing various measurements related to the utilization of codons, such as relative synonymous codon usage (RSCU), codon adaptation index

(CAI), optimal number of codons (ENC), index of codon bias (CBI), incidence of preferred codons (FOP), ENC scores, RSCU scores, and the likelihood of each nucleotide appearing in the third position of a codon (Li et al., 2024). The coding sequences' GC1, GC2, and GC3 contents were determined using EMBOSS software (Rice et al., 2000).

## Sequence variation of cp genome

To assess gene rearrangements and boundaries within the large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions of ten section *Cyclobalanopsis* species, the researchers utilized the IRscope (<https://github.com/Limpfrog/irscope>) (Amiryousefi et al., 2018) online tool to generate horizontal visualizations. They carried out a comparison analysis of sequence variations among ten genomes by employing the MVISTA software in shuffle-LAGAN mode. The MVISTA program can be accessed at the following URL: (<http://genome.lbl.gov/vista/mvista/submit.shtml>). They assessed nucleotide variation within cp genomes by screening for sites with high variability using DNAsp6 software (Rozas et al., 2017) based on the nucleotide diversity index ( $\pi$ ).

## Selection pressure and phylogenetic analyses

KaKs\_Calculator (<https://sourceforge.net/projects/kakscalculator/>) (Wang et al., 2010) was adopted to calculate the rate of nonsynonymous mutation (Ka) and synonymous mutation (Ks) in protein-coding genes. The results of Ka/Ks could be used to assess the role of selection for each gene in CPGs of 10 *Quercus* species.

A phylogenetic tree of the genus *Quercus* was established using Bayesian (BI) analysis based on cp genomic data to understand the phylogenetic relationships of the genus *Quercus*. This tree included four sections of the genus *Quercus* endemic to China, encompassing 27 species (All 27 species in Supplementary Table S2), *Trigonobalanus doichangensis* as an outgroup. Apply all selected cp genomic sequences to align MAFFT v7.427 (Katoh and Standley, 2013). Later, MrBayes v3.2.7 (Ronquist et al., 2012) was utilized to carry out the BI tree analysis based on the following processes: infer the best-fit nucleotide substitution model (GTR+F+I+G4) by Modeltest (Posada and Crandall, 1998) and PAUP (Matthews and Rosenberger, 2008). A Markov chain Monte Carlo (MCMC) analysis was run for 6,000,000 generations, sampling the tree every 1,000 generations and ignoring the initial 0.25 as the burnin score.

## Divergence time estimate

The estimation of differentiation times for plants in Section *Cyclobalanopsis* was conducted using the BEAST v1.10.4 (Suchard et al., 2018) software package. Both Section *Cyclobalanopsis* and its closely related taxa have a rich fossil record for temporal calibration. Two crown node calibrations from Hipp et al.'s (Hipp et al., 2020)

study on oak fossil calibration were used to calibrate differentiation times. Cp genome sequences were inputted using BEAUti v1.10.4, with species taxa set based on the node age calibration. Taxon differentiation time points were calibrated with a Normal prior distribution and a Yule tree prior with an uncorrelated relaxed lognormal molecular clock. The Markov-Chain Monte Carlo (MCMC) chain length was set to 200 million, run twice independently, with parameter sampling every 1000 generations. The xml files were then imported into BEAST v1.10.4. The tree and log files were combined using LOGCOMBINER, and convergence was assessed using Tracer v1.7.2 to detect effective sample sizes (ESSs) for all parameters. A 20% burn-in was set in TreeAnnotator v1.10.4 to generate the Maximum Clade Credibility tree (MCC). The tree file was displayed and edited in FigTree v1.4, and the 95% confidence intervals (95% HPD) were examined.

## Result

### Basic properties of the cp genome

We managed to sequence the entire cp genomes of seven *Cyclobalanopsis* species (*Q. argyrotricha*, *Q. augustinii*, *Q.*

*bambusifolia*, *Q. bella*, *Q. edithiae*, *Q. jenseniana*, and *Q. poilanei*). We compared them with the three published cp genomes sequences (*Q. litseoides*, *Q. obovatifolia*, and *Q. saravanensis*), whose lengths ranged from 160,707 to 160,999 base pairs (Figure 1). The genes were conserved in the composition and sequence within these cp genomes. We also found that these genomes have a very typical tetragonal structure, with an LSC region (90,147 bp to 90,326 bp) flanked by two IR regions (25,809 bp to 25,860 bp), separating it from the SSC region (18,788 bp to 18,993 bp). In all, 132 genes were identified, including 8 rRNA genes, 37 tRNA genes, and 87 protein-coding genes (Table 1).

### Codon usage

The codon usage patterns were presented in Table 2. The protein-coding genes, spanning from 62,493 to 62,991 bp, were obtained from ten species in the section *Cyclobalanopsis*. The quantities of synonymous codons fluctuated from 20,831 to 20,992. The ENC (Effective Number of Codons) value lay between 49.89 and 49.96. The FOP (Frequency of Optimal Codons) value was 0.355 in *Q. augustinii*, *Q. bella* and 0.354 in the other eight samples. The GC content was between 37.94 and 38.00%. The codon preference indexes of the ten species varied slightly, indicating similar codon

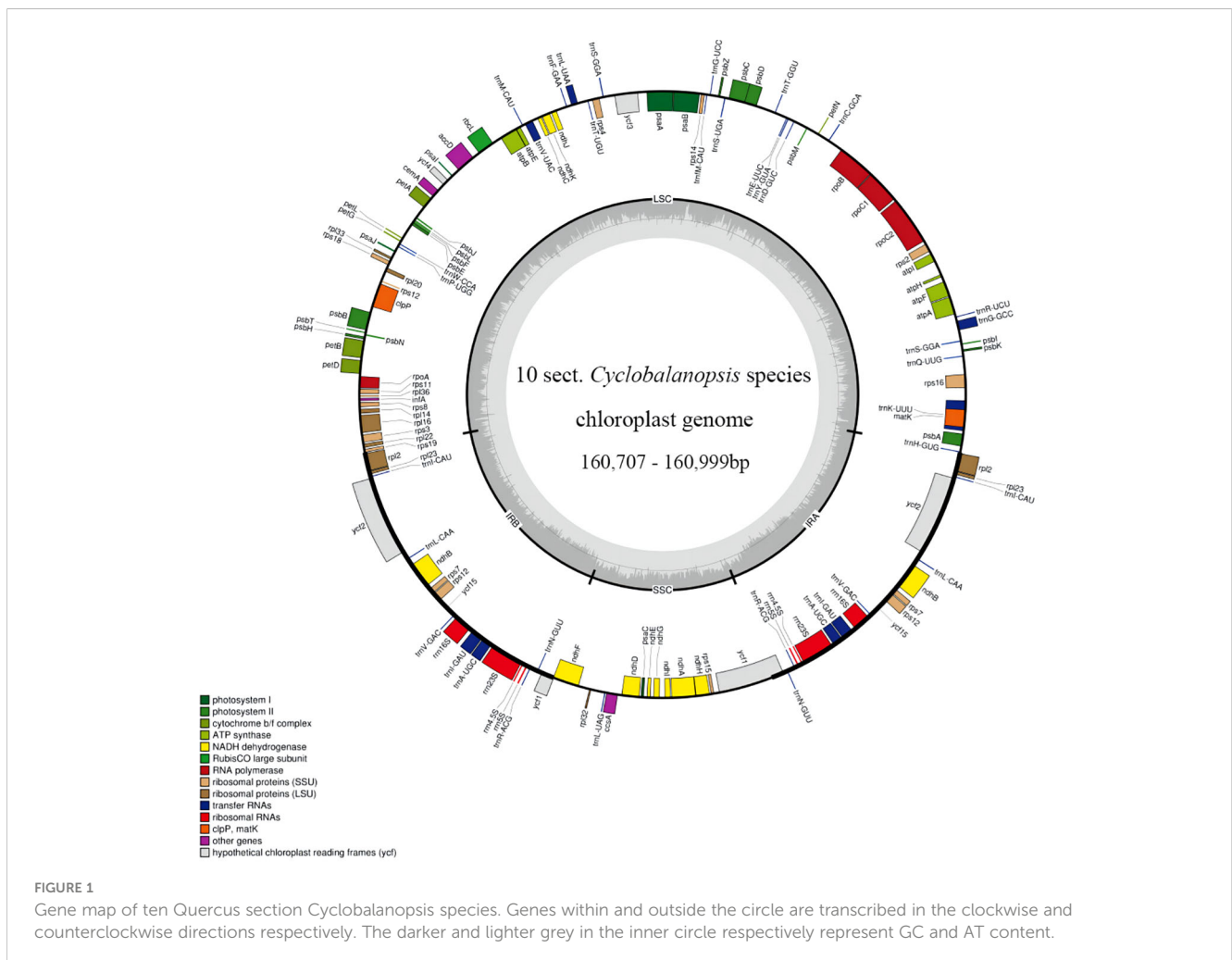


TABLE 1 A summary of the statistics regarding the cp genomes of 10 *Quercus* sect. *Cyclobalanopsis* species.

Species	<i>Q. argyrotricha</i>	<i>Q. augustinii</i>	<i>Q. bambusifolia</i>	<i>Q. bella</i>	<i>Q. edithiae</i>	<i>Q. jenseniana</i>	<i>Q. litseoides</i>	<i>Q. obovatifolia</i>	<i>Q. poilanei</i>	<i>Q. saravanensis</i>
Genome size (bp)	160,999	160,846	160,803	160,798	160,707	160,802	160,782	160,817	160,784	160,767
Length of LSC (bp)	90,326	90,228	90,274	90,255	90,147	90,236	90,235	90,210	90,216	90,220
Length of IRs (bp)	25,840	25,860	25,811	25,833	25,842	25,842	25,840	25,809	25,842	25,825
Length of SSC (bp)	18,993	18,898	18,907	18,788	18,876	18,882	18,867	18,989	18,884	18,897
Number of genes	132	132	132	132	132	132	132	132	132	132
protein-coding genes	87	87	87	87	87	87	87	87	87	87
tRNA genes	37	37	37	37	37	37	37	37	37	37
rRNA genes	8	8	8	8	8	8	8	8	8	8
GC content (%)	36.90	36.89	36.88	36.89	36.90	36.89	36.90	36.90	36.89	36.89
GC content of LSC (%)	34.78	34.75	34.73	34.74	34.75	34.74	34.74	34.75	34.74	34.74
GC content of IRs (%)	42.77	42.75	42.79	42.77	42.77	42.77	42.77	42.78	42.77	42.76
GC content of SSC (%)	30.97	31.09	31.06	31.11	31.12	31.10	31.13	31.10	31.11	31.08
Species origin	new sequencing	new sequencing	new sequencing	new sequencing	new sequencing	new sequencing	NC_065782	NC_039972	new sequencing	NC_060377

usage. The GC3 of ten species ranged between 29.85 and 29.94%. The analysis of codon usage preference in the *Cyclobalanopsis* cp genomes revealed 30 high-frequency codons with Relative Synonymous Codon Usage (RSCU) values greater than 1, of which 28 ended with A or U bases and only two with G or C bases. This demonstrates a bias towards A/U endings in the *Cyclobalanopsis* cp genomes.

Nonetheless, the tryptophan codon (UGG) and the methionine codon (AUG) did not exhibit a clear preference, as indicated by their RSCU values. Figure 2 shows a significant conservation of codon usage within the section *Cyclobalanopsis*, although species-specific variations were present (Specific RSCU values are shown in Supplementary Table S3).

TABLE 2 Codon preference indicators of 10 species of section *Cyclobalanopsis*.

Index	<i>Q. argyrotricha</i>	<i>Q. augustinii</i>	<i>Q. bambusifolia</i>	<i>Q. bella</i>	<i>Q. edithiae</i>	<i>Q. jenseniana</i>	<i>Q. litseoides</i>	<i>Q. obovatifolia</i>	<i>Q. poilanei</i>	<i>Q. saravanensis</i>
Length (bp)	62,991	62,931	62,829	62,976	62,838	62,940	62,838	62,493	62,940	62,823
Codon number	20,980	20,977	20,943	20,992	20,946	20,980	20,946	20,831	20,980	20,941
Effective number of codons	49.89	49.94	49.94	49.94	49.96	49.92	49.95	49.95	49.91	49.94
Codon adaptation index	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167	0.167
Codon bias index	-0.099	-0.099	-0.099	-0.099	-0.100	-0.099	-0.100	-0.100	-0.100	-0.099
Frequency of optimal codons	0.354	0.355	0.354	0.355	0.354	0.354	0.354	0.354	0.354	0.354
GC content (%)	37.94	37.96	37.97	37.96	37.99	37.95	37.98	38.00	37.95	37.97
GC1 content (%)	46.12	46.14	46.16	46.11	46.18	46.13	46.19	46.24	46.12	46.17
GC2 content (%)	37.86	37.85	37.83	37.85	37.84	37.85	37.83	37.85	37.85	37.83
GC3 content (%)	29.85	29.89	29.92	29.91	29.94	29.88	29.92	29.92	29.89	29.91

## Analysis of the cp genome structure

The SSR analysis detected five types of SSRs: single, dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide repeats. The cp genomes of the ten plant species mainly had mononucleotide SSRs, and pentanucleotide SSRs were the least common. *Q. augustinii* and *Q. obovatifolia* had the greatest number of SSRs (120), while *Q. edithiae* and *Q. jenseniana* had the lowest (115). The difference in the number of SSRs among the species was not statistically significant. Among the ten *Cyclobalanopsis* species, A/T simple repeats were the most frequent, with AAG/CTT, AATG/ATTC, and AAAAT/ATTTT repeats each occurring only once. Uniquely, the SSR AAATT/AATTT was only identified in *Q. augustinii* and *Q. bambusifolia* (Figure 3 and Supplementary Table S4). The search for dispersed repetitive sequences uncovered four kinds: palindromic (P), complement (C), reverse (R), and forward (F). Although there were minor variations in their quantities, P were the most abundant, while C were the least known (Figure 4 and Supplementary Table S5). Notably, *Q. argyrotricha* had more repetitive sequences.

A common feature found in the cp genomes of the ten species belonging to section *Cyclobalanopsis* was the presence of a circular tetrad structure defined by four distinct boundaries: IRa-LSC, SSC-IRa, IRb-SSC, and LSC-IRb (Figure 5). While the genome sizes of these species were similar, with minor variations, there were differences observed in the genes flanking these regions, specifically *rps19*, *ndhF*, and *ycf1*. It was noted that *rps19* consistently resided within the LSC region, albeit at varying distances from the LSC-IRb boundary, specifically at 3 bp or 10 bp. The genes *ndhF* and *ycf1* were positioned in close proximity to the SSC-IR boundaries. Furthermore, a segment of *ycf1* of the same length was detected in the IRb region of all ten *Cyclobalanopsis* species, however, it was categorized as a pseudogene (*ψycf1*).

## Sequence divergence, hotspots and selection pressure estimation

The study conducted an analysis of sequence variation in the cp genomes of ten species within section *Cyclobalanopsis* using the Shuffle-LAGAN model in the mVISTA online software, with the cp genome sequence of *Q. kerrii* as a reference. Results depicted in

Figure 6 revealed high conservation in the coding regions of the cp genomes of the ten *Cyclobalanopsis* species, with minimal variation in the rRNA genes. Despite the typical conservation of the cp genome in section *Cyclobalanopsis*, some differences were observed across various groups. Nucleotide polymorphisms were identified in the cp genes of the ten *Cyclobalanopsis* species. Figure 7 displayed a relatively higher sequence conservation in the intergenic spacer of the reverse repeat (IR) region compared to the single-copy region. Pi values can be found in Supplementary Table S6. The study then identified the *trnK-rps16*, *psbC-trnS*, *rbcL-accD*, and *ycf1* genes ( $\pi > 0.006$ ) from the coding regions as potential DNA barcodes for investigating the genetic and phylogenetic relationships among *Quercus* species.

To estimate the role of selection of the *Quercus* section *Cyclobalanopsis* species, Ka and Ks values in 10 cp genomes using *Quercus chenii* as a reference. The Ka/Ks values were calculated and recorded in Supplementary Table S7, ranging from 0 to 1.52384. Among these, 36 protein-coding genes showed significance (Figure 8) in 10 species. Based on the calculation results, we speculated that the purification selection may have affected most protein-coding genes, as their Ka/Ks values were less than 1. At the same time, when  $Ka/Ks > 1$  demonstrated that the positive selection was working on the genes. Therefore we identified three genes that were under the positive selection, namely the *atpF* gene in *Q. argyrotricha*, *Q. augustinii*, *Q. bella*, *Q. jenseniana*, and *Q. poilanei*, *rpoC1* gene in *Q. bambusifolia* and *ycf2* gene in all 10 species.

## Phylogenetic relationships

To explore the phylogenetic relationships among species of Section *Cyclobalanopsis*, 7 species assembled and annotated in this study were combined with 20 species downloaded from NCBI to construct a phylogenetic tree. This study used Bayesian inference (BI) to build a phylogenetic tree (Figure 9). Based on the phylogenetic tree, *Trigonobalanus doichangensis* was located at the base of the phylogenetic tree and represents an early divergent group within the Fagaceae family. The phylogenetic tree consists of two main branches: Subgenus *Cerris* and Subgenus *Quercus*. The former category includes 14 different species belonging to Section *Cyclobalanopsis*, 4 from Section *Ilex*, and 3 from Section *Cerris*. The latter shall consist of 5 species from Section



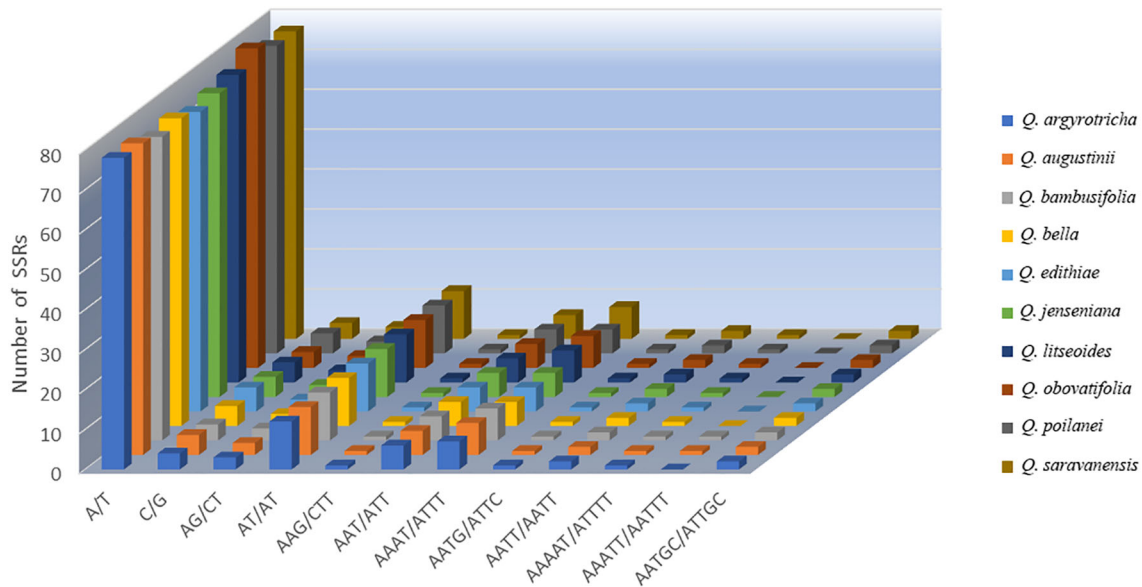


FIGURE 3 The number of SSR units detected in 10 section *Cyclobalanopsis* cp genome.

*Quercus*. All branches had support rates ranging from moderate (greater than 70%) to advanced (greater than 90%), and most nodes had support rates of 100% (All 27 species in Supplementary Table S2).

back to Neogene (18.03 Ma), which was earlier than the divergence time of other taxa. This study demonstrated that Section *Cyclobalanopsis* is the earliest occurring type in the genus *Quercus*.

### Divergence time estimate

This study used phylogenetic species to estimate the differentiation time of Section *Cyclobalanopsis* species based on chloroplast genome sequences (Figure 10). The results showed that the most recent common ancestor of 14 Section *Cyclobalanopsis* plants can be traced

### Discussion

#### Cp genome architectures in ten *Quercus* section *Cyclobalanopsis* species

We successfully constructed the cp genomes of seven *Quercus* section *Cyclobalanopsis* species and downloaded three species from

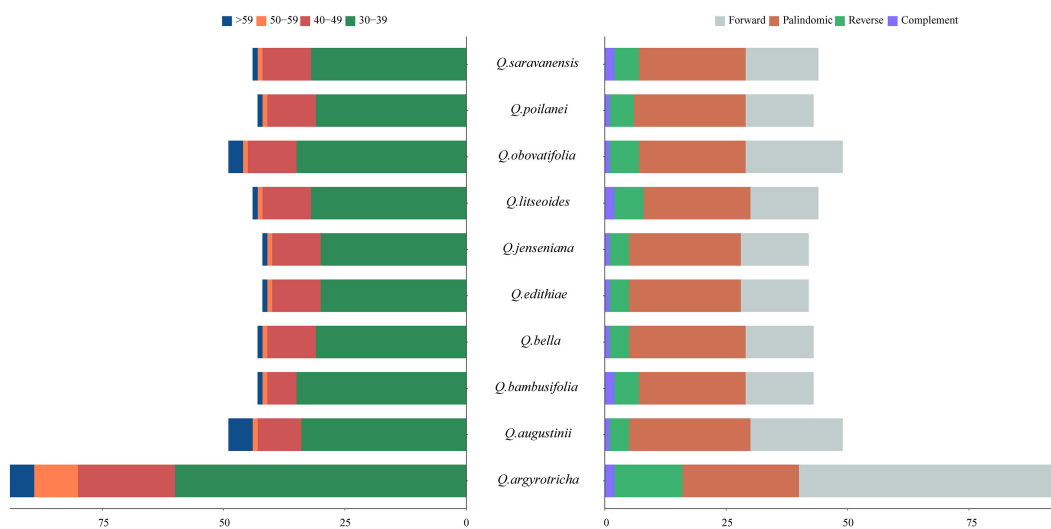


FIGURE 4 Type and quantity of scattered repeats in the cp genome of 10 section *Cyclobalanopsis* cp genomes.

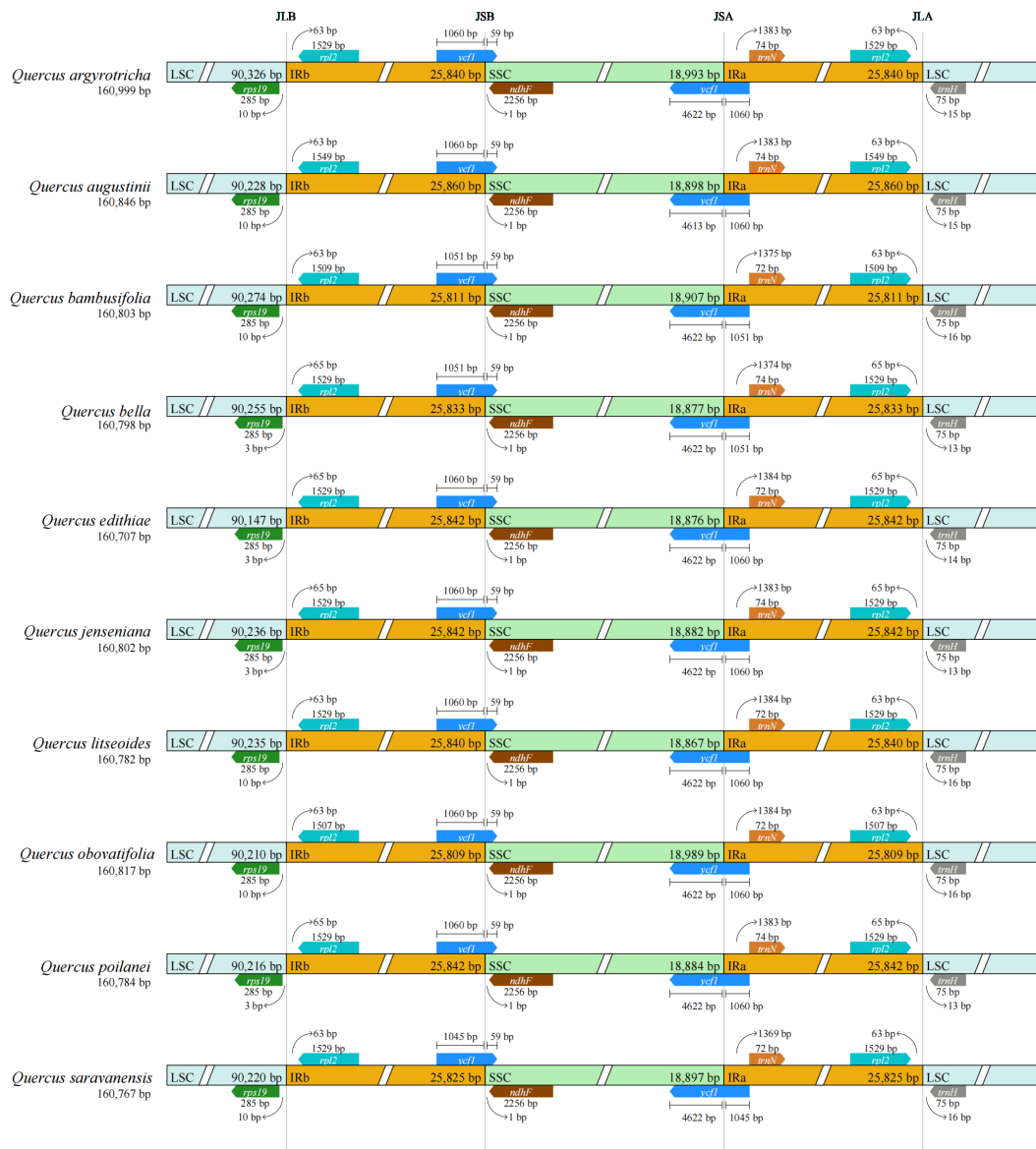


FIGURE 5 LSC, SSC and IR boundaries of the cp genomes in 10 section Cyclobalanopsis cp genome.

NCBI for joint analysis. The size of the 10 cp genomes (ca. 160 kb) is consistent with the plastid chromosomes of photosynthetic land plants, which range in size from 120 to 160kb (Wicke et al., 2011). The same quadripartite circular structure was found in the ten *Quercus* section *Cyclobalanopsis* species (Li et al., 2018; Wang et al., 2021; Chen et al., 2022). Moreover, the GC content is an essential indicator of phylogenetic relationships between species (Mitreva et al., 2006). Therefore, the overall GC content in the IR and SC regions of ten *Quercus* species was analyzed. The results suggest that the overall GC content and the GC content within the IR and SC regions are largely consistent, with the GC content in the intergenic spacer (IGS) region showing a statistically significant rise compared to the GC content in the large single-copy (LSC) region and the small single-copy (SSC) region (Li et al., 2024).

The position of boundaries between cp sequences is a critical component in the study of cp genome evolution, and the contraction and expansion of IR boundaries can help elucidate phylogenetic relationships between different taxa (Chen J. et al., 2023; Huang et al., 2022). Based on common evolutionary phenomena in plant genomes, we must be aware that the position of the IR/SC boundary may change due to contraction or expansion of the IR region (Wang and Messing, 2011; Huang et al., 2014). The IRs/SCs boundaries of all species compared in this study are located in similar positions, with slight differences in displacement between species. The conservatism of the *Quercus* section *Cyclobalanopsis* was demonstrated by the relatively constant length of cp genomes and the minor variations in their region borders, which are the same conditions as other *Quercus* species (Xu et al., 2015; Wang et al., 2021).



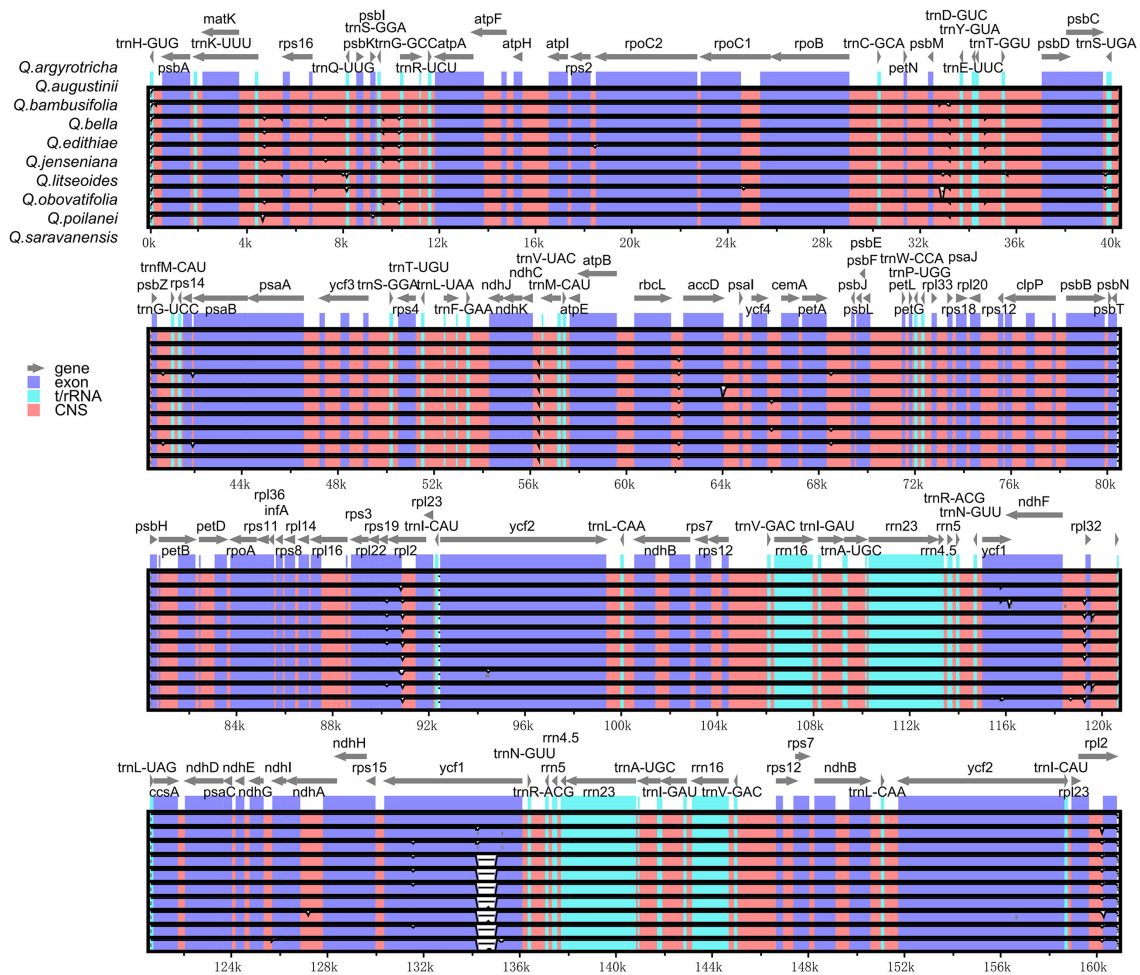


FIGURE 6 Comparison of cp genomes among 10 section *Cyclobalanopsis* species.

The preference for codon usage is a key aspect of biological evolution, shaped by a variety of factors that influence genetic code functionality. These influencing factors include genome size, base mutations, genetic drift, natural selection, gene expression levels, and protein structure (Angellotti et al., 2007). It is important to recognize that synonymous codons arise from mutations, and their relative usage can be assessed using the measure of relative synonymous codon usage (RSCU). The RSCU provides insights into differences in codon preference across various genes (Parvathy et al., 2022). By analyzing RSCU, we observed 64 codons with a distinct preference for A/T-ending sequences in these oak species. Interestingly, this pattern was also evident in an analysis of GC3 content, indicating a predilection for A/T-ending codons in the *Quercus* genus studied. This phenomenon of codon usage bias is frequently observed in angiosperms (Li et al., 2024).

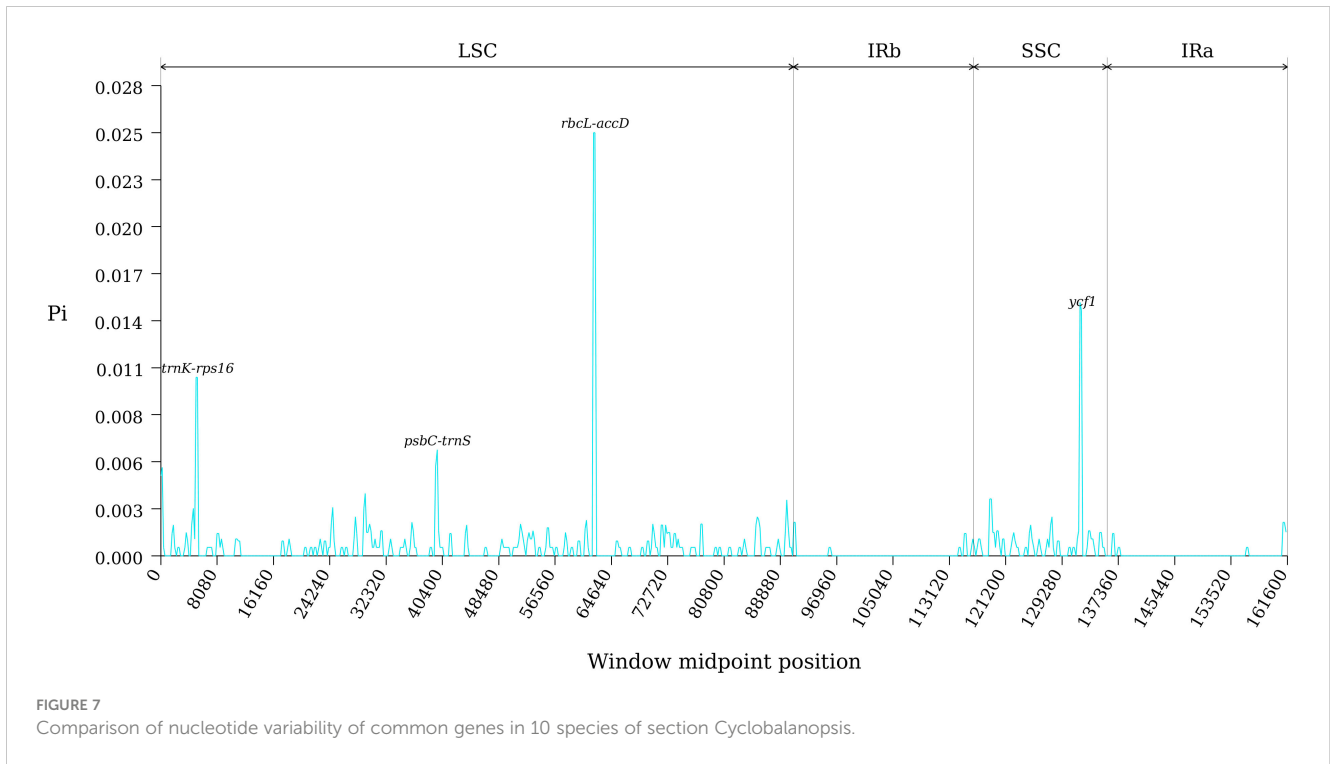
inheritance and evolution of plant species (Moore et al., 2010). Our study examined ten species within the *Cyclobalanopsis* section and found that the number of simple sequence repeats (SSRs) ranged from 115 to 120. Single nucleotide repeats were the most common type of SSRs, followed by dinucleotide and tetranucleotide repeats. The SSRs in *Cyclobalanopsis* cp genomes had a high A/T base content, indicating a preference for A/T bases. Interestingly, no hexanucleotide repeats were found, consistent with previous research on *Cyclobalanopsis* section (Wang et al., 2021; Li et al., 2022). Dispersed repeats were also observed in the ten distinct species of *Cyclobalanopsis* section, consisting mainly of forward and palindromic sequences. However, differences in the number of tandem repeats among the species suggested varying rates of mutational events (Li et al., 2024).

### Diversity of repetitive sequences and SSRs

Repetitive sequences play a crucial role in storing genetic information, influencing gene expression, and impacting the

### Conservatisms, highly variable regions and selection pressure estimation

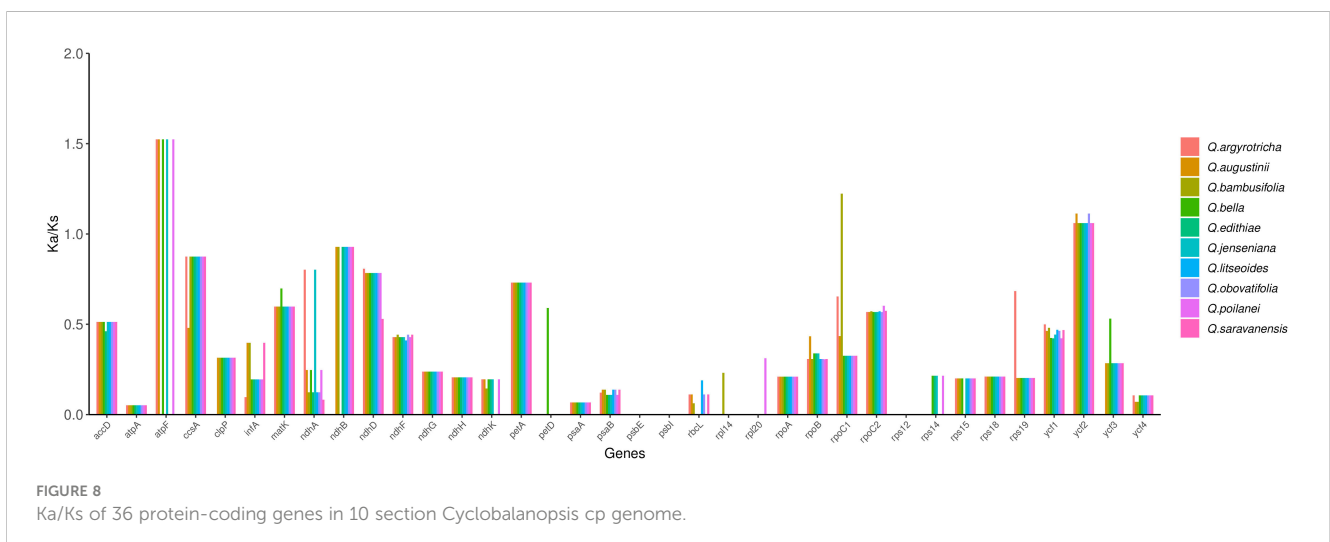
Our study compared 10 species by mVISTA sequence identity analysis using the *Q. kerrii* genome as a reference. The results

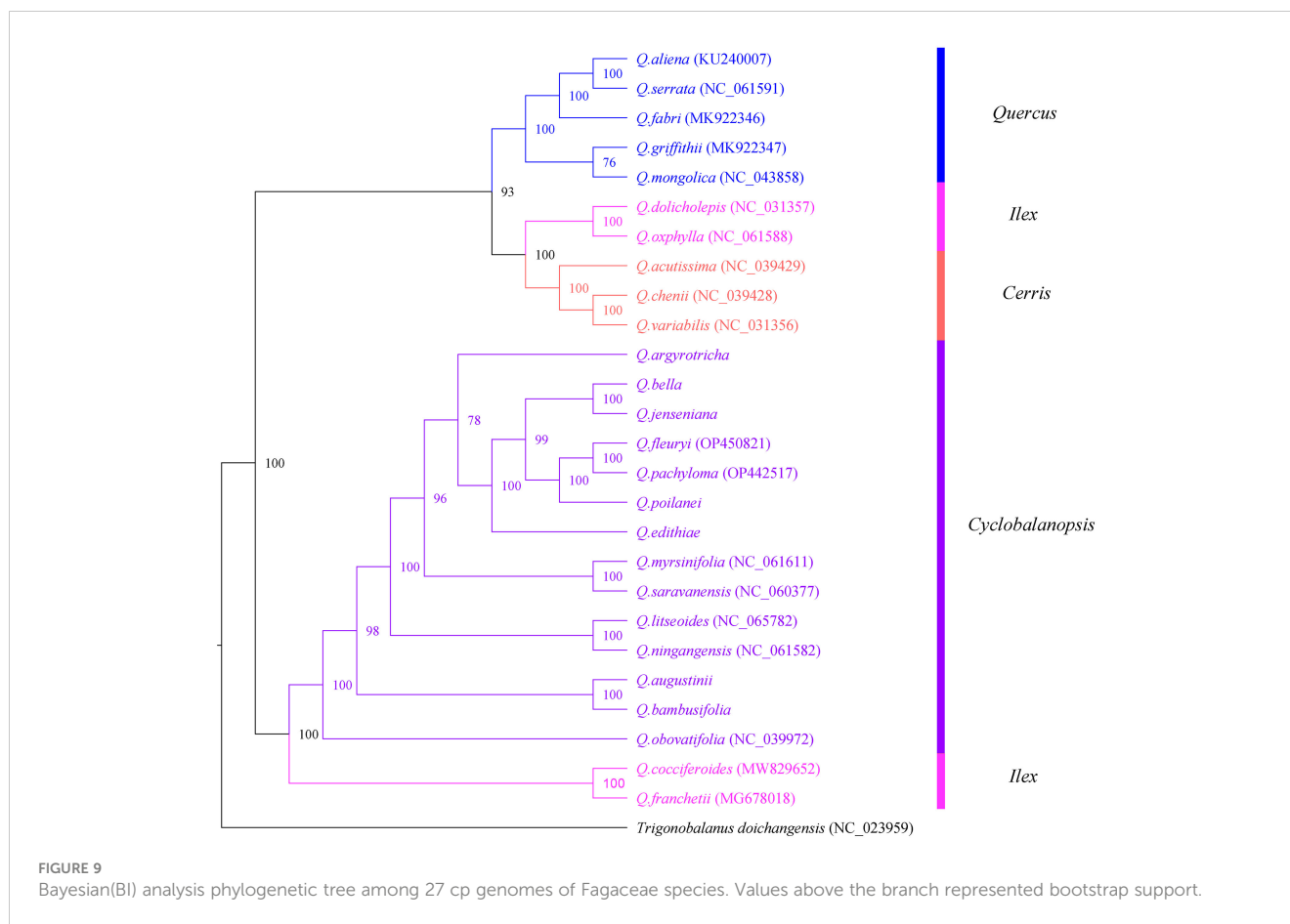


showed that these cp genomes exhibit high conservation in coding regions and significant differences in non-coding regions, especially in the LSC and SSC regions. This is similar to the findings of most angiosperm cp genomes showing high regional diversity (Xu et al., 2020; Fan et al., 2021, 2022). The non-coding regions showed a high degree of variation, especially in the LSC and SSC regions. *Q. bella*, *Q. edithiae*, *Q. jenseniana*, *Q. litseoides*, *Q. obovatifolia*, *Q. poilanei* and *Q. saravanensis* showed significant sequence variation in the *ycf1* gene region. Since the role of the *ycf1* gene in cp function is not fully understood, these variations may reflect the evolution of specific adaptations in these species. Sliding window analysis by dnaSP v6 software revealed the nucleotide diversity values (Pi) range in the cp genomes of 10 plant species. Four highly mutated regions were screened, namely *trnK-rps16*, *psbC-trnS*, *rbcL-accD*

and *ycf1* genes. These highly variable regions provide a valuable resource for developing molecular markers. Because of their genetic diversity, they are ideal candidates for developing species-specific DNA barcodes (Kress, 2017; Antil et al., 2023).

In our study, most of the Ka/Ks values were less than 1 or unavailable, suggesting that the emergence frequency of synonymous nucleotide substitution was more than that of non-synonymous nucleotide substitution due to the purify selection process (Matsuoka et al., 2002; Castle, 2011). Therefore we identified three genes that were under the positive selection, namely the *atpF* gene in *Q. argyrotricha*, *Q. augustinii*, *Q. bella*, *Q. jenseniana*, and *Q. poilanei*, *rpoC1* gene in *Q. bambusifolia* and *ycf2* gene in all 10 species. Whether these divergence hotspots found in the above analysis could be utilized for DNA barcodes or





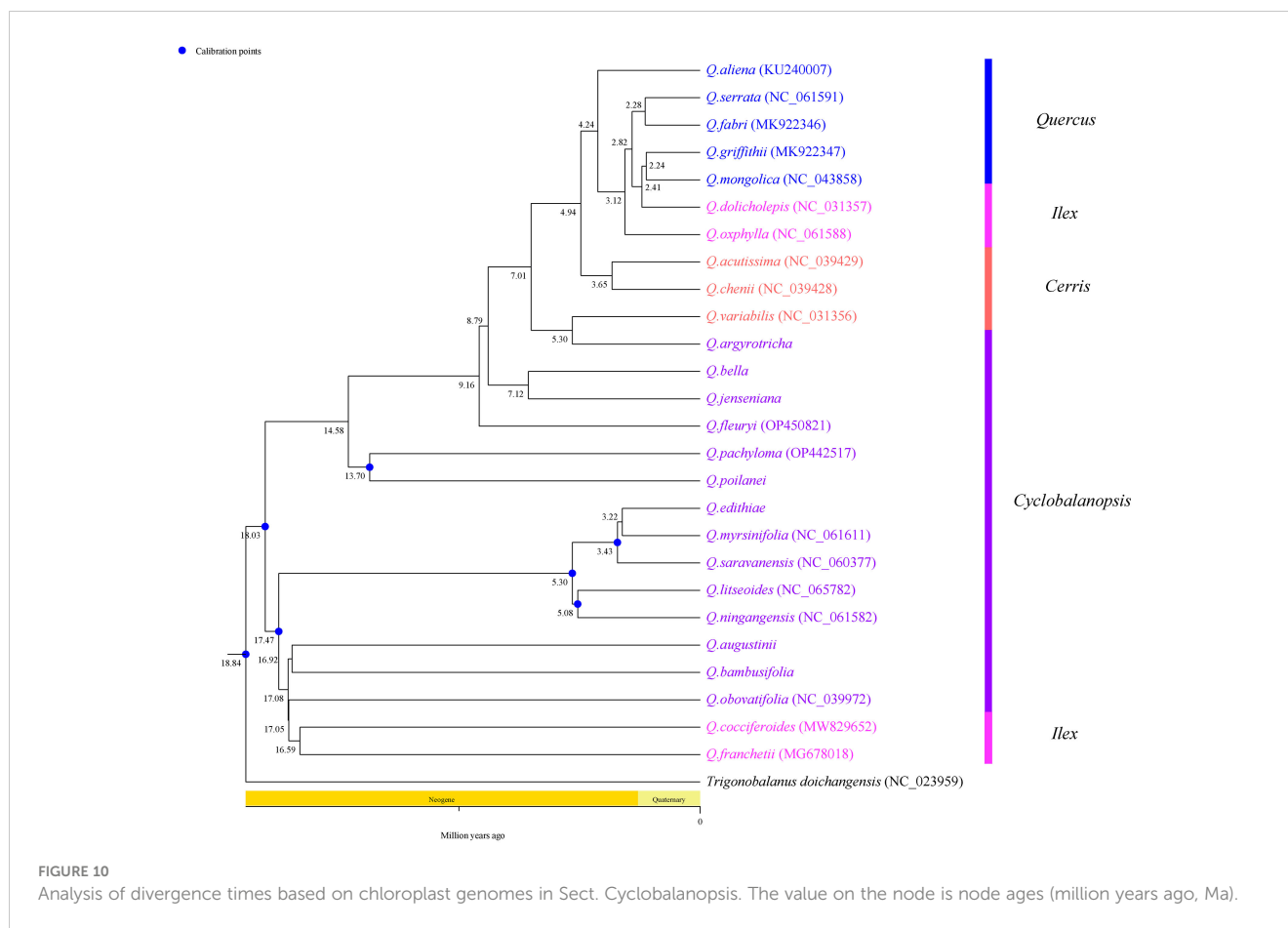
estimating taxonomic evolution in genus *Quercus* needs further research.

## Inference of phylogenetic relationship

China is one of the centers of diversity of the subgenus *Cyclobalanopsis* and presents significant challenges in understanding the evolution of oak species (Carrero et al., 2020). Studies of the taxonomy based on the morphology of oak trees are limited due to convergent evolution and the common occurrence of hybridization among different species. In spite of these challenges, Deng Min successfully developed a classification system for the *Cyclobalanopsis* section. However, the only molecular phylogenetic analysis of the genus *Quercus* in China has depended on RAD-seq sequencing (Deng et al., 2018). Most studies utilizing cp genomes have successfully yielded high-resolution and well-supported phylogenetic trees, even in phylogenetically challenging plant taxa (Feng et al., 2022; Lee et al., 2022; Feng et al., 2023; Li et al., 2023). The phylogenetic tree comprises two primary branches: Subgenus *Cerris* and Subgenus *Quercus*. The first group encompasses fourteen distinct species within Section *Cyclobalanopsis*, four species from Section *Ilex*, and another three species from Section *Cerris*. The latter shall consist of five species from Section *Quercus*—the species of Sect. *Quercus* forms a separate branch, whereas the species of Sect. *Cyclobalanopsis*, Sect. *Ilex* and Sect. *Cerris* forms paraphyletic

groups. It has been found in existing studies of the genus *Quercus* that the reason Sect. *Ilex* is not a monophyletic group, possibly due to incomplete genealogical screening or gradual infiltration between ancestral taxa (Hipp et al., 2020).

Further investigation of species within the *Cyclobalanopsis* section revealed that *Q. bella* and *Q. jenseniana* formed sister branches, while *Q. augustinii* and *Q. bambusifolia* also clustered together. However, a previous study by Deng Min et al. (2014) they were proposed that *Q. jenseniana* might be more closely related to *Q. augustinii*. This discrepancy could be due to two main reasons: first, Deng Min et al. primarily classified the species based on morphological characteristics, such as leaf epidermis, where *Q. jenseniana* and *Q. augustinii* share similarities. Second, there might have been genetic introgression, as *Q. jenseniana* was collected from the same location as *Q. bella*. Collecting samples of *Q. jenseniana* from various populations and geographic ranges is recommended for further analysis of their morphological traits. Interestingly, *Q. argyrotricha* does not group with the other species in the phylogenetic tree, contradicting previous studies based on morphology and molecular phylogenetics. We propose that this discrepancy in the relationship and systematic position of *Q. argyrotricha* could be attributed to unique repetitive sequences in its cp genome compared to other species, as well as its higher altitude of collection. *Q. argyrotricha* displays the highest number of repetitive sequences among the analyzed species, suggesting potential recombination events and greater genetic diversity.



Trichomes are typically located on the leaves, particularly during the early stages of leaf development. Some trichomes shed and vanish as the leaf matures, leaving only their bases on the leaf surface. This characteristic holds significant taxonomic importance in the *Quercus* genus. Plants in the *Cyclobalanopsis* section generally possess unicellular and composite trichomes, with the former being more prevalent. In this study, all species except *Q. argyrotricha* exhibited unicellular trichomes, aligning with previous findings (Deng et al., 2014). Although genomic data could potentially aid in resolving species classification challenges within *Quercus*, it is noted that the chloroplast genome represents only a portion of the plant's total genetic material, underscoring the issue's complexity. Therefore, developing advanced genetic techniques for phylogenetic and population analysis is expected to deepen our insights into the evolutionary history of the *Quercus* genus.

## Divergence time estimate

Thanks to the matrilineal nature of inheritance, cp genomic data are relatively stable, and their distribution is not prone to change in any species, allowing them to be effectively utilized to study the evolutionary history of species (Daniell et al., 2016). In this study, we estimated species divergence time based on the chloroplast genome and showed that the Section *Cyclobalanopsis* branch can be traced back to Neogene (18.03 Ma). Through the

study of the world's first fossil of Section *Cyclobalanopsis* from the Maoming Basin, Guangdong Province, China, the previous authors concluded that Section *Cyclobalanopsis* plants reached the tropical and subtropical regions of South China in the Early Oligocene, which also suggests that the modern distribution pattern of Section *Cyclobalanopsis* may have originated in the Oligocene (Liu X. Y. et al., 2019). Fossil evidence indicates that the genus *Palaeocrustacea* is the earliest-appearing taxon of the genus *Quercus*, and its leaf structure and cuticle characteristics are closest to those of Section *Cyclobalanopsis*, so it can be concluded that Section *Cyclobalanopsis* is the earliest-appearing type of the genus *Quercus*. In the present study, this argument can also be proved. In this study, Section *Cyclobalanopsis* mainly diverged in Neogene (ca. 3.22–18.03 Ma), a period of rapid uplift of the Tibetan Plateau (17–25 Ma) (Wen et al., 2014; Favre et al., 2015), which may have allowed species from Eurasia to diverge in response to geographic and environmental isolation (Li et al., 2019). In addition, the formation of the East Asian summer monsoon system was an essential climatic event during this period, which brought abundant precipitation and facilitated the lineage differentiation of plant groups (Sun and Wang, 2005; Kong et al., 2017). In this study, Section *Cyclobalanopsis* plants successively diverged in 17–25 Ma, reflecting the diversification of Section *Cyclobalanopsis* species, which is close to previous estimates based on the RAD-seq dataset (ca 10 Ma) (Hipp et al., 2020). The rapid species differentiation during this time was mainly driven by Himalayan

movements and climate change, such as the rapid uplift of the Tibetan Plateau around 13-15 Ma, 7-8 Ma, and 1.6-3.5 Ma, and the significant increase in the strength of the East Asian monsoon during the same period, which drove the rapid differentiation of plant taxa (Sun and Wang, 2005; Wen et al., 2014).

## Conclusion

In this study, we sequenced seven species in *Quercus* section *Cyclobalanopsis* and downloaded three species from NCBI for analysis together. Cp genome assembly and annotation revealed that the cp genome of *Quercus* plants has a typical circular tetrameric structure, and the cp genome size ranges from 160,707 to 160,999 bp. By comparative genome analysis, the species we studied prefer codons ending in A/U. The most numerous simple repeat sequences in *Quercus* plants are single nucleotides, mainly A/T-based. There was no obvious contraction and expansion of the IR/SC boundary in these plants. Only *ycf1* gene showed  $\psi ycf1$  gene in the IRB/SC boundary. Four highly variable hotspots were detected in comparison analysis, they are *trnK-rps16*, *psbC-trnS*, *rbcL-accD*, and *ycf1*. Besides, three genes (*atpF*, *rpoC1*, and *ycf2*) were detected under positive selection pressure. Phylogenetic analysis shows that *Q. bella* and *Q. jenseniana*, *Q. augustinii*, and *Q. bambusifolia* have a recent relationship. It is noteworthy that *Q. argyrotricha* is individually differentiated, possibly related to its repetitive sequence differences. Divergence time analysis revealed that Section *Cyclobalanopsis* represents the earliest type of *Quercus* genus. The findings obtained will facilitate further investigations into the taxonomy, phylogenetic evolution, and preservation of *Quercus* genus.

## Data availability statement

The datasets generated and/or analyzed during the current study are available in the National Center for Biotechnology Information repository, Accession Number: OQ534364, PP498793, PP498794, PP471977, PP471975, PP498796, PP498797.

## Ethics statement

The plant materials directly used in this study do not contain any rare or endangered plants. All the plant materials in this study were obtained from the wild, and the collection of plant materials has been permitted by the National Park of Hainan Tropical Rainforest, Daweishan Nature Reserve of Yunan, Guangxi Damingshan National Nature Reserve and Guizhou Wangmo

*Cycas* Nature Reserve. This study protocol also complies with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

## Author contributions

KH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing – original draft, Writing – review & editing. BL: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft. XC: Conceptualization, Investigation, Software, Writing – review & editing. CQ: Conceptualization, Investigation, Methodology, Software, Writing – review & editing. XZ: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from: National Specimen Platform Teaching Standard Subplatform (<http://mnh.scu.edu.cn/>) (2005DKA21403-JK); Research and Innovation Team of China West Normal University (KCXTD2022-4).

## Conflict of interest

The 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/fpls.2024.1430191/full#supplementary-material>

## References

- 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

- Amiryousefi, A., Hyvönen, J., and Pocza, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34, 3030–3031. doi: 10.1093/bioinformatics/bty220
- Angellotti, M. C., Bhuiyan, S. B., Chen, G., and Wan, X. F. (2007). CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Res.* 35, W132–W136. doi: 10.1093/nar/gkm392
- Antil, S., Abraham, J. S., Sripoorna, S., Maurya, S., Dagar, J., Makhija, S., et al. (2023). DNA barcoding, an effective tool for species identification: a review. *Mol. Biol. Rep.* 50, 761–775. doi: 10.1007/s11033-022-08015-7
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198
- Bendich, A. J. (1987). Why do chloroplasts and mitochondria contain so many copies of their genome. *Bioessays* 6, 279–282. doi: 10.1002/bies.950060608
- Carrero, C., Jerome, D., Beckman, E., Byrne, A., Coombes, A. J., Deng, M., et al. (2020). *The red list of oaks 2020* (Lisle: The Morton Arboretum).
- Castle, J. C. (2011). SNPs occur in regions with less genomic sequence conservation. *PLoS One* 6, e20660. doi: 10.1371/journal.pone.0020660
- Chen, X., Li, B., and Zhang, X. (2023). Comparison of chloroplast genomes and phylogenetic analysis of four species in Quercus section Cyclobalanopsis. *Sci. Rep.* 13, 18731. doi: 10.1038/s41598-023-45421-8
- Chen, J., Wang, F., Zhao, Z., Li, M., Liu, Z., and Peng, D. (2023). Complete chloroplast genomes and comparative analyses of three paraphalaenopsis (Aceridinae, Orchidaceae) species. *Int. J. Mol. Sci.* 24, 11167. doi: 10.3390/ijms241311167
- Chen, S., Zhang, W., Li, Y., Ge, X., Zhou, X., Hu, Y., et al. (2022). The complete chloroplast genome sequence of *Quercus sessilifolia* Blume (Fagaceae). *Mitochondr. DNA B* 7, 182–184. doi: 10.1080/23802359.2021.2017366
- Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 1–29. doi: 10.1186/s13059-016-1004-2
- Deng, M., Hipp, A., Song, Y. G., Li, Q. S., Coombes, A., and Cotton, A. (2014). Leaf epidermal features of *Quercus* subgenus Cyclobalanopsis (Fagaceae) and their systematic significance. *Bot. J. Linn. Soc.* 176, 224–259. doi: 10.1111/boj.12207
- Deng, M., Jiang, X. L., Hipp, A. L., Manos, P. S., and Hahn, M. (2018). Phylogeny and biogeography of East Asian evergreen oaks (*Quercus* section Cyclobalanopsis; Fagaceae): insights into the Cenozoic history of evergreen broad-leaved forests in subtropical Asia. *Mol. Phylogenet. Evol.* 119, 170–181. doi: 10.1016/j.ympev.2017.11.003
- Denk, T., and Grimm, G. W. (2009). Significance of pollen characteristics for infrageneric classification and phylogeny in *Quercus* (Fagaceae). *Int. J. Plant Sci.* 170, 926–940. doi: 10.1086/600134
- Denk, T., Grimm, G. W., Manos, P. S., Deng, M., and Hipp, A. L. (2017). “An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns,” in *Oaks Physiological Ecology. Exploring the Functional Diversity of Genus Quercus L.* Eds. E. Gil-Pelegrin, J. Peguero-Pina and D. Sancho-Knapik (Springer, Cham), 13–38. doi: 10.1007/978-3-319-69099-5\_1
- Fan, Y., Jin, Y. N., Ding, M., Tang, Y., Cheng, J., Zhang, K., et al. (2021). The complete chloroplast genome sequences of eight *Fagopyrum* species: Insights into genome evolution and phylogenetic relationships. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.799904
- Fan, X., Wang, W., Wagutu, G. K., Li, W., Li, X., and Chen, Y. (2022). Fifteen complete chloroplast genomes of *Trapa* species (Trapaceae): insight into genome structure, comparative analysis and phylogenetic relationships. *BMC Plant Biol.* 22, 230. doi: 10.1186/s12870-022-03608-7
- Favre, A., Päckert, M., Pauls, S. U., Jähniß, S. C., Uhl, D., Michalak, I., et al. (2015). The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol. Rev.* 90, 236–253. doi: 10.1111/brv.12107
- Feng, J. L., Wu, L. W., Wang, Q., Pan, Y. J., Li, B. L., Lin, Y. L., et al. (2022). Comparison analysis based on complete chloroplast genomes and insights into plastid phylogenomic of four *Iris* species. *BioMed. Res. Int.* 2022, 1–14. doi: 10.1155/2022/2194021
- Feng, J., Xiong, Y., Su, X., Liu, T., Xiong, Y., Zhao, J., et al. (2023). Analysis of complete chloroplast genome: Structure, phylogenetic relationships of *Galega orientalis* and evolutionary inference of Galegeae. *Genes* 14, 176. doi: 10.3390/genes14010176
- Gil-Pelegrin, E., Peguero-Pina, J. J., and Sancho-Knapik, D. (2017). “Oaks and people: a long journey together,” in *Oaks Physiological Ecology. Exploring the Functional Diversity of Genus Quercus L.* Eds. E. Gil-Pelegrin, J. Peguero-Pina and D. Sancho-Knapik (Springer, Cham), 1–11. doi: 10.1007/978-3-319-69099-5\_1
- Hipp, A. L., Manos, P. S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., et al. (2020). Genomic landscape of the global oak phylogeny. *New Phytol.* 226, 1198–1212. doi: 10.1111/nph.16162
- Huang, Y., Li, J., Yang, Z., An, W., Xie, C., Liu, S., et al. (2022). Comprehensive analysis of complete chloroplast genome and phylogenetic aspects of ten *Ficus* species. *BMC Plant Biol.* 22, 253. doi: 10.1186/s12870-022-03643-4
- Huang, H., Shi, C., Liu, Y., Mao, S. Y., and Gao, L. Z. (2014). Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol. Biol.* 14, 1–17. doi: 10.1186/1471-2148-14-151
- Hubert, F., Grimm, G. W., Jousselin, E., Berry, V., Franc, A., and Kremer, A. (2014). Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*. *Syst. Biodivers.* 12, 405–423. doi: 10.1080/14772000.2014.941037
- Jiang, X. L., Mou, H. L., Luo, C. S., and Xu, G. B. (2021). The complete chloroplast genome sequence of *Quercus chungii* (Fagaceae). *Mitochondr. DNA B* 6, 1789–1790. doi: 10.1080/23802359.2021.1931505
- Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., Depamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21, 1–31. doi: 10.1186/s13059-020-02154-5
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kong, H., Condamine, F. L., Harris, A., Chen, J., Pan, B., Möller, M., et al. (2017). Both temperature fluctuations and East Asian monsoons have driven plant diversification in the karst ecosystems from southern China. *Mol. Ecol.* 26, 6414–6429. doi: 10.1111/mec.14367
- Kremer, A., Le Corre, V., Petit, R. J., and Ducousso, A. (2010). “Historical and contemporary dynamics of adaptive differentiation in European oaks,” in *Molecular approaches in natural resource conservation*. Eds. J. A. DeWoody, J. W. Bickham and C. H. Michler (Cambridge University Press, Cambridge), 101–122. doi: 10.1017/CBO9780511777592
- Kress, W. J. (2017). Plant DNA barcodes: Applications today and in the future. *J. Syst. Evol.* 55, 291–307. doi: 10.1111/jse.12254
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633
- Lee, H. R., Kim, K. A., Kim, B. Y., Park, Y. J., Lee, Y. B., and Cheon, K. S. (2022). The complete chloroplast genome sequences of eight *Orostachys* species: Comparative analysis and assessment of phylogenetic relationships. *PLoS One* 17, e0277486. doi: 10.1371/journal.pone.0277486
- Li, Y., Zhang, X., and Fang, Y. (2019). Landscape features and climatic forces shape the genetic structure and evolutionary history of an oak species (*Quercus chenii*) in East China. *Front. Plant Sci.*
- Li, B., Huang, K., Chen, X., Qin, C., and Zhang, X. (2024). Comparative and phylogenetic analysis of chloroplast genomes from four species in *Quercus* section Cyclobalanopsis. *BMC Genomic Data* 25, 57. doi: 10.1186/s12863-024-01232-y
- 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
- Li, E., Liu, K., Deng, R., Gao, Y., Liu, X., Dong, W., et al. (2023). Insights into the phylogeny and chloroplast genome evolution of *Eriocaulon* (Eriocaulaceae). *BMC Plant Biol.* 23, 32. doi: 10.1186/s12870-023-04034-z
- Li, Y., Wang, T. R., Kozłowski, G., Liu, M. H., Yi, L. T., and Song, Y. G. (2022). Complete chloroplast genome of an endangered species *Quercus litseoides*, and its comparative, evolutionary, and phylogenetic study with other *Quercus* section Cyclobalanopsis species. *Genes* 13, 1184. doi: 10.3390/genes13071184
- 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
- Liu, X. Y., Xu, S. L., Han, M., and Jin, J. H. (2018). An early Oligocene fossil acorn, associated leaves and pollen of the ring-cupped oaks (*Quercus* subg. Cyclobalanopsis) from Maoming Basin, South China. *J. Syst. Evol.* 57, 153–168.
- Lo, C. C., and Chain, P. S. (2014). Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinform.* 15, 1–8.
- Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52, 267–274. doi: 10.1007/s00294-007-0161-y
- Manos, P. S., Doyle, J. J., and Nixon, K. C. (1999). Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Mol. Phylogenet. Evol.* 12, 333–349. doi: 10.1006/mpev.1999.0614
- Matsuoka, Y., Yamazaki, Y., Ogihara, Y., and Tsunewaki, K. (2002). Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. *Mol. Biol. Evol.* 19, 2084–2091. doi: 10.1093/oxfordjournals.molbev.a004033
- Matthews, L. J., and Rosenberger, A. L. (2008). Taxon combinations, parsimony analysis (PAUP\*), and the taxonomy of the yellow-tailed woolly monkey, *Lagothrix flavicauda*. *Am. J. Phys. Anthropol.* 137, 245–255. doi: 10.1002/ajpa.20859
- Mitrevva, M., Wendl, M. C., Martin, J., Wylie, T., Yin, Y., Larson, A., et al. (2006). Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biol.* 7, 1–19. doi: 10.1186/gb-2006-7-8-R75
- Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4623–4628. doi: 10.1073/pnas.0907801107
- Muthreich, F., Zimmermann, B., Birks, H. J. B., Viçosa, C. M., and Seddon, A. W. (2020). Chemical variations in *Quercus* pollen as a tool for taxonomic identification: Implications for long-term ecological and biogeographical research. *J. Biogeogr.* 47, 1298–1309. doi: 10.1111/jbi.13817

- Neuhaus, H., and Emes, M. (2000). Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Biol.* 51, 111–140. doi: 10.1146/annurev.arplant.51.1.111
- Oh, S. H., and Manos, P. S. (2008). Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences. *Taxon* 57, 434–451. doi: 10.2307/25066014
- Parvathy, S. T., Udayasuriyan, V., and Bhadana, V. (2022). Codon usage bias. *Mol. Biol. Rep.* 49, 539–565. doi: 10.1007/s11033-021-06749-4
- Peden, J. (1997). *CodonW* (Dublin: Trinity College).
- Posada, D., and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Rozas, J., Ferrer-Mata, A., Sánchez-Delbarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248
- Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47, W65–W73. doi: 10.1093/nar/gkz345
- Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4, vey016. doi: 10.1093/ve/vey016
- Sun, X., and Wang, P. (2005). How old is the Asian monsoon system?—Palaeobotanical records from China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 222, 181–222. doi: 10.1016/j.palaeo.2005.03.005
- Vitelli, M., Vessella, F., Cardoni, S., Pollegioni, P., Denk, T., Grimm, G. W., et al. (2017). Phylogeographic structuring of plastome diversity in Mediterranean oaks (*Quercus* Group Ilex, Fagaceae). *Tree Genet. Genomes* 13, 1–17. doi: 10.1007/s11295-016-1086-8
- Wang, W., and Messing, J. (2011). High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. *PLoS One* 6, e24670. doi: 10.1371/journal.pone.0024670
- Wang, T. R., Wang, Z. W., Song, Y. G., and Kozłowski, G. (2021). The complete chloroplast genome sequence of *Quercus ningangensis* and its phylogenetic implication. *Plant Fungal Syst.* 66, 155–165. doi: 10.35535/pfsyst-2021-0014
- Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteomics Bioinf.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3
- Welter, S., Bracho-Núñez, A., Mir, C., Zimmer, I., Kesselmeier, J., Lumaret, R., et al. (2012). The diversification of terpene emissions in Mediterranean oaks: lessons from a study of *Quercus suber*, *Quercus canariensis* and its hybrid *Quercus afares*. *Tree Physiol.* 32, 1082–1091. doi: 10.1093/treephys/tps069
- Wen, J., Zhang, J. Q., Nie, Z. L., Zhong, Y., and Sun, H. (2014). Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Front. Genet.* 5, 1–16. doi: 10.3389/fgene.2014.00004
- Wicke, S., Schneeweiss, G. M., Depamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4
- Xu, J., Deng, M., Jiang, X. L., Westwood, M., Song, Y. G., and Turkington, R. (2015). Phylogeography of *Quercus glauca* (Fagaceae), a dominant tree of East Asian subtropical evergreen forests, based on three chloroplast DNA interspace sequences. *Tree Genet. Genomes* 11, 1–17. doi: 10.1007/s11295-014-0805-2
- Xu, J., Shen, X., Liao, B., Xu, J., and Hou, D. (2020). Comparing and phylogenetic analysis chloroplast genome of three *Achyranthes* species. *Sci. Rep.* 10, 10818. doi: 10.1038/s41598-020-67679-y
- Yang, Y., Zhou, T., Qian, Z., and Zhao, G. (2021). Phylogenetic relationships in Chinese oaks (Fagaceae, *Quercus*): Evidence from plastid genome using low-coverage whole genome sequencing. *Genomics* 113, 1438–1447. doi: 10.1016/j.ygeno.2021.03.013