Postprint: Chloroplast Genome Characteristics and Phylogenetic Analysis of Four Species of Quercus Subgenus Cyclobalanopsis

Authors: Huang Ting, Tang Meng, Chen Xiaoli, Li Buyu, Zhang Xuemei

Date: 2022-08-05T00:00:00+00:00

Abstract

The phylogenetic status of subgenus Cyclobalanopsis plants has long been controversial, with interspecific relationships of some species remaining unclear. To reveal the chloroplast genome characteristics and phylogenetic relationships of Quercus ningangensis, Q. oxyodon, Q. gambleana, and Q. neglecta, this study selected mature leaves of the above four Quercus species for next-generation sequencing, analyzed their chloroplast genome structure and characteristics, and conducted phylogenetic studies in combination with related taxa. The results showed that: (1) The chloroplast genome sequence lengths of Q. ningangensis, Q. oxyodon, Q. gambleana, and Q. neglecta were 160,906 bp, 160,883 bp, 160,832 bp, and 160,784 bp, respectively, all encoding 133 genes, including 88 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. (2) The four Quercus species showed a preference for codons ending with A/T, and plastome variation regions were mainly present in non-coding regions. (3) Through IR boundary analysis, it was found that the four subgenus Cyclobalanopsis plants contained a ycf1 pseudogene and underwent expansion in the IRb/SSC region. (4) Phylogenetic analysis revealed that within Fagaceae, Fagus and Trigonobalanus diverged early, subgenus Quercus did not form a monophyletic group, and chloroplast genome-based tree results were consistent with nuclear and plastid markers, with intermingling observed between the Ilex and Cerris sections within subgenus Quercus. (5) Flora of China and some local floras treat Q. gambleana as a separate species, whereas Zhou Zhekun, Deng Min, and others, based on leaf characteristics, treat it as a subspecies of Q. oxyodon; the phylogenetic status of Q. gambleana remains controversial. Based on chloroplast genome information combined with previous morphological analysis results, this study supports the view of Q. gambleana as an independent species. This study provides fundamental data for exploring the phylogenetic status of subgenus Cyclobalanopsis, the classification of sections within subgenus Cyclobalanopsis, and resolving interspecific relationships of questionable species.

Full Text

Preamble

Comparison of Chloroplast Genomes and Phylogenetic Analysis of Four Species of Quercus subg. Cyclobalanopsis

HUANG Ting, TANG Meng, CHEN Xiaoli, LI Puyu, ZHANG Xuemei*

College of Life Sciences, China West Normal University, Nanchong 637002, Sichuan, China

Abstract

The phylogenetic status of Quercus subg. Cyclobalanopsis has long been controversial, and the interspecific relationships of some species remain unclear. To reveal the chloroplast genome characteristics and phylogenetic relationships of *Quercus ningangensis*, *Q. oxyodon*, *Q. gambleana*, and *Q. neglecta*, this study performed next-generation sequencing on mature leaves of these four species, analyzed their chloroplast genome structure and characteristics, and conducted phylogenetic analyses in combination with related taxa. The results showed: (1) The chloroplast genome sequences of *Q. ningangensis*, *Q. oxyodon*, *Q. gambleana*, and *Q. neglecta* were 160,906 bp, 160,883 bp, 160,832 bp, and 160,784 bp, respectively, all encoding 133 genes including 88 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. (2) The four species preferred codons ending in A/T, and plastome variation regions were mainly located in non-coding regions. (3) IR boundary analysis revealed that the four Quercus subg. Cyclobalanopsis species contained a *ycf1* pseudogene, with expansion occurring at the IRb/SSC boundary. (4) Phylogenetic analysis indicated that *Fagus* and *Trigonobalanus* diverged early in Fagaceae, while subg. Quercus did not form a monophyletic group. Chloroplast genome-based phylogenies were consistent with nuclear and plastid markers, showing interspersion between the Ilex and Cerris groups within subg. Quercus. (5) While *Flora of China* and some local floras treat *Q. gambleana* as a separate species, Zhou Zhekun and Deng Min et al. considered it a subspecies of *Q. oxyodon* based on leaf characteristics, leaving its phylogenetic status controversial. Based on chloroplast genome information combined with previous morphological analyses, this study supports *Q. gambleana* as an independent species. This research provides fundamental data for addressing the phylogenetic status of Quercus subg. Cyclobalanopsis, the classification of its sections, and the resolution of questionable interspecific relationships.

Keywords: Fagaceae, Quercus subg. Cyclobalanopsis, genome comparison, chloroplast genome, phylogenetic analysis

Introduction

Chloroplasts are widely present in higher plants and most algae, serving as the site of photosynthesis in green plants. Chloroplasts possess independent genetic material with a characteristic quadripartite genome structure consisting of a large single-copy region (LSC), a small single-copy region (SSC), and two inverted repeat regions (IRA and IRB). Both genome sequence and structure are relatively conserved (Zhao et al., 2018). Compared with nuclear genomes, chloroplast genomes feature simple structure, small molecular weight, low gene substitution rates, and single-copy characteristics (Korpelainen, 2004; Kwak et al., 2019). In recent years, with reduced time and cost for next-generation sequencing, increasing numbers of chloroplast genome sequencing analyses have been applied to ambiguous species identification and phylogenetic studies. Current research has utilized chloroplast genomes to resolve phylogenetic issues in Quercus subg. Quercus and *Castanea* within Fagaceae (Yang, 2018; Gao, 2020), but few studies have addressed phylogenetic problems in Quercus subg. Cyclobalanopsis using chloroplast genome information.

Quercus subg. Cyclobalanopsis (Fagaceae) comprises 122 species (Frodin & Govaerts, 1998) or 150 species (Huang et al., 1999), primarily distributed in tropical and subtropical regions of Asia. China represents an important distribution area and diversity center for this subgenus (Luo & Zhou, 2001a), with 77 species and 3 varieties recorded. The primary distribution extends south of the Qinling Mountains and Huai River, where these species serve as constructive components of evergreen broad-leaved forests (Chen & Huang, 1998; Huang et al., 1999). Cyclobalanopsis wood is valued for its hardness, corrosion resistance, and wear resistance, making it excellent material for boat oars and tools. The fruits are rich in starch (approximately 50-60% content) (Duan, 1995) and can be used for wine production, feed, adhesives, and industrial starch. The bark contains tannins suitable for tannin extraction.

The phylogenetic status of Cyclobalanopsis has long been controversial. Some scholars have treated it as a separate genus based on cupule arrangement patterns (Xu & Ren, 1976; Zheng, 1985; Chen & Huang, 1998; Huang et al., 1999). However, most researchers support its treatment as a subgenus within *Quercus* based on pollen anatomical morphology and ITS sequences (Institute of Botany, Chinese Academy of Sciences, 1972; Liu & Fang, 1986; Duan, 1992; Luo & Zhou, 2001b). Due to substantial variation in reproductive structures, relatively uniform leaf characteristics, inter-group gene introgression, limited herbarium specimens, and difficulty observing important taxonomic features such as fruits, no clear infrageneric classification exists. Interspecific relationships remain unclear for many species, and numerous controversies persist regarding species delimitation. Current phylogenetic research on Quercus subg. Cyclobalanopsis remains primarily morphological. Therefore, additional molecular systematic evidence is needed to help resolve its phylogenetic status and interspecific relationships.

This study performed next-generation sequencing on mature leaves of four Cy-

clobalanopsis species, aiming to: (1) investigate basic chloroplast genome characteristics and differences among the four species; (2) analyze codon usage preferences; and (3) clarify the phylogenetic relationship between *Q. gambleana* and *Q. oxyodon*. We hope to provide new chloroplast genome-level data for resolving phylogenetic issues, interspecific relationships, species identification, and superior timber selection in Cyclobalanopsis.

1.1 Experimental Materials

Fresh, pest-free mature leaves of *Quercus ningangensis*, *Q. oxyodon*, *Q. gambleana*, and *Q. neglecta* were collected in the field (Table 1). After wiping with moist gauze, leaves were stored in sealed bags with color-changing silica gel. Once dried, samples were stored at -80°C. Voucher specimens are deposited in the Herbarium of the College of Life Sciences, China West Normal University.

1.2 Genome DNA Extraction and Sequencing

Total DNA was extracted from leaves of the four Cyclobalanopsis species using a modified CTAB method. Paired-end sequencing of extracted DNA was performed using Illumina HiSeq. Chloroplast genome DNA extraction and sequencing were completed by Nanjing Auwigene Technology Co., Ltd.

1.3 Assembly and Annotation

Data quality control was performed using FASTQ software. Clean data were assembled using GetOrganelle software (Jin et al., 2020). Raw reads were mapped to assembled sequences to check coverage, and gaps were filled using Gapcloser software (Zuo et al., 2017). The chloroplast genome of *Quercus stewardiana* (MN199023) was downloaded from NCBI as a reference genome for annotation using CPGAVAS2 (Shi et al., 2019). Annotated data were manually adjusted and modified using Geneious software (Kearse et al., 2012). Physical maps of the four Cyclobalanopsis chloroplast genomes were drawn using OGDRAW software. Annotated data were submitted to NCBI with accession numbers ON303301, ON258628, ON258629, and ON258631 for *Q. ningangensis*, *Q. oxyodon*, *Q. gambleana*, and *Q. neglecta*, respectively.

1.4 Chloroplast Genome Comparative Analysis

Geneious and Editseq software were used to compile basic information including lengths of SSC, LSC, and IR regions and GC content for the four Cyclobalanopsis chloroplast genomes. Chloroplast genomes of *Q. tenuicupula* (MN199025), *Q. stewardiana* (MN199023), *Q. sichouensis* (NC_{036941}), and *Q. glauca* (NC_{036930}) were downloaded from NCBI for comparative analysis.

Non-repetitive sequences longer than 300 bp with ATG start codons were selected, yielding 52 CDS sequences for all species. CodonW software was used to calculate relative synonymous codon usage (RSCU), codon adaptation index

(CAI), effective number of codons (ENC), codon bias index (CBI), and frequency of optimal codons (FOP) for the four Cyclobalanopsis species and related taxa. EMBOSS online software was used to calculate GC1, GC2, and GC3 content for each CDS sequence. Neutrality plot analysis used the average of GC1 and GC2 (GC12) as the vertical axis and GC3 as the horizontal axis, with a y=x function line inserted. ENC-plot used GC3 and ENC values as horizontal and vertical axes, respectively, with a theoretical ENC curve (ENC=2+GC3+29/[$GC3^2$+(1-$GC3)^2$]). PR2-plot used G3/(G3+C3) as the horizontal axis and A3/(A3+T3) as the vertical axis to analyze base usage frequency and bias. All plots were generated using Origin software.

mVISTA software was used for visual comparative analysis of chloroplast genomes from the four Cyclobalanopsis species and related taxa. IRscope was used to draw IR boundaries of the four species and related taxa.

1.5 Phylogenetic Analysis

Eight publicly available Fagaceae chloroplast genome sequences were downloaded from NCBI: *Q. obovatifolia* (MG356785.1), *Q. aquifolioides* (KX911971.1), *Q. spinosa* (MG678038.1), *Q. engleriana* (MZ196209.1), *Q. variabilis* (NC031356.1), *Trigonobalanus doichangensis* (NC023959.1), *Fagus hayatae*, and *F. engleriana* (as outgroups). These were combined with the four newly assembled Cyclobalanopsis chloroplast genomes for phylogenetic tree construction. All sequences were aligned using MAFFT, and unaligned terminal regions were trimmed using MEGA 7. The best model (GTR+GI) for ML tree construction was identified using MEGA 7' s Models function. Maximum likelihood (ML) trees were constructed with 1,000 bootstrap replicates. Additionally, neighbor-joining (NJ) trees were constructed using the Maximum Composite Likelihood model with 10,000 bootstrap replicates.

2.1 Basic Characteristics of Chloroplast Genomes

The chloroplast genomes of *Q. ningangensis*, *Q. oxyodon*, *Q. gambleana*, and *Q. neglecta* all exhibited typical quadripartite structure comprising one LSC region, two IR regions, and one SSC region [Figure 1: see original paper]. Genome lengths ranged from 160,681 to 160,906 bp, with *Q. oxyodon* and *Q. gambleana* differing by only 1 bp (160,883 bp vs. 160,882 bp). LSC regions ranged from 90,245 to 90,360 bp, SSC regions from 18,891 to 18,929 bp, and IR regions from 25,816 to 25,840 bp. Total GC content (36.9%), IR region GC content (42.8%), and CDS region GC content (37.9%) were identical across the four species, with minimal variation in LSC and SSC region GC content (Table 2).

Annotation revealed all four species possessed 133 genes, including 37 tRNA genes, 8 rRNA genes, and 88 protein-coding genes (Table 2). Fifteen genes contained one intron (*rpoC1*, *ndhA*, *ndhB*, *rpl2*, *rpl16*, *atpF*, *rps16*, *trnL-UAA*, *trnK-UUU*, *trnI-GAU*, *trnA-UGC*, *trnG-GCC*, *trnV-UAC*, *petB*, *petD), while three genes contained two introns (*rps12, clpP, ycf3*) (Table 3).

2.2.1 Codon Composition Analysis

The four Cyclobalanopsis species contained 20,977-20,996 codons with effective number of codons (ENC) values of 49.81-49.91. Codon adaptation index (CAI) was 0.167 for all species. Codon bias index (CBI) was -0.099 for *Q. ningangensis*, *Q. gambleana*, and *Q. neglecta*, and -0.097 for *Q. oxyodon*. Frequency of optimal codons (FOP) was 0.355 for all four species, with GC content ranging from 37.93% to 37.95%. Similar codon preference indices across the four species indicated comparable codon usage patterns (Table 4). GC3 content ranged from 29.85% to 29.88%, demonstrating preference for A/T-ending codons.

Leucine (Leu), serine (Ser), and arginine (Arg) have six synonymous codons, tryptophan (Trp) and methionine (Met) have single codons, and other amino acids have two or more synonymous codons. The eight Cyclobalanopsis species showed similar codon preferences, with higher usage frequencies for ACU, UCU, UUA, GCU, UAU, GAU, and AGA, and lower frequencies for CUC, CUG, GCG, UAC, CAC, CAG, AAC, AAG, GAC, GAG, CGC, CGG, AGC, and GGC. In *Castanopsis carlesii* and *Lithocarpus longinux*, UAA showed lower usage frequency compared to other listed plants.

2.2.2 Neutrality Plot Analysis

Neutrality plot analysis of non-repetitive sequences (>300 bp, ATG start codon) from the four Cyclobalanopsis chloroplast genomes showed correlation coefficients between GC3 and GC12 of 0.00793, 0.05384, 0.005495, and 0.06, with regression coefficients of 0.11633, 0.33692, 0.34267, and 0.35694, respectively (all $P<0.01$). The weak correlation between GC12 and GC3 indicated significant differences between the first/second and third codon positions, suggesting natural selection has greater influence on codon usage preference than mutation pressure.

2.2.3 ENC-plot Analysis

ENC-plot analysis revealed most genes had actual ENC values below the standard curve, indicating stronger natural selection effects on codon usage preference for these genes. Some genes had ENC values near or above the curve, suggesting greater mutation influence. ENC ratio frequency distribution (Table 5) showed 18-19 genes (35-37%) in the -0.05 to 0.05 interval for each species, indicating significant mutation effects. However, 33-34 genes showed large differences between actual and theoretical ENC values, suggesting strong natural selection effects. Combined results indicate codon preference in these chloroplast genomes is influenced by both mutation and natural selection, with natural selection playing the dominant role.

2.2.4 PR2-plot Analysis

PR2-plot analysis [Figure 5: see original paper] showed uneven gene distribution, with most genes located in the lower right quadrant, indicating A>T and G>C usage at the third codon position. This demonstrates that codon preference in the four Cyclobalanopsis species is jointly influenced by mutation and natural selection.

2.3 Chloroplast Genome Comparative Analysis

Using *Q. ningangensis* as a reference, mVISTA analysis of *Q. oxyodon*, *Q. gambleana*, *Q. neglecta*, *Q. tenuicupula*, *Q. stewardiana*, *Q. sichouensis*, and *Q. glauca* chloroplast genomes revealed high conservation, with major variation in non-coding regions and substantial variation in the *ycf1* coding region [Figure 6: see original paper]. The LSC region showed greater variation than the more conserved IR regions, while SSC variation was intermediate. rRNA genes were highly conserved across all eight Cyclobalanopsis species.

2.4 IR Boundary Analysis

IRscope analysis of IR expansion/contraction and boundary positions [Figure 7: see original paper] showed consistent gene distribution near boundaries across eight Cyclobalanopsis species, though distances varied slightly. The *rps19* gene was consistently 11 bp from the SSC/IRb boundary. *trnH* distance from the IRa/LSC boundary was 16 bp in *Q. sichouensis* and *Q. glauca* but 1 bp in other species. In *Q. gambleana* and *Q. neglecta*, the *ndhF* gene had one base in the IRb region with the remaining 2,255 bp in SSC, while other species had the entire 2,255 bp *ndhF* gene in SSC. The *ycf1* gene, the second longest in higher plant chloroplast genomes, spanned the SSC/IRa boundary with 1,060 bp in IRa for *Q. ningangensis*, *Q. tenuicupula*, *Q. stewardiana*, *Q. sichouensis*, and *Q. glauca*, and 1,056 bp, 1,054 bp, and 1,062 bp in IRa for *Q. oxyodon*, *Q. gambleana*, and *Q. neglecta*, respectively. Corresponding *ycf1* fragments in the IRb region represent *ycf1* pseudogenes (*ycf1*) spanning the IRb/SSC boundary with 56-58 bp located upstream of IRb/SSC.

2.5 Phylogenetic Analysis

Phylogenetic trees were constructed using ML and NJ methods with 12 Fagaceae chloroplast genome sequences (including the four newly assembled Cyclobalanopsis genomes) and *Fagus hayatae* and *F. engleriana* as outgroups. Both methods yielded high support values and consistent topologies [Figure 8: see original paper]. The phylogeny showed *Fagus* and *Trigonobalanus* as early-diverging lineages at the base of Fagaceae. *Quercus spinosa*, *Q. engleriana*, and *Q. variabilis* clustered together, with interspersion between Ilex and Cerris groups. *Q. aquifolioides* showed closer affinity to Cyclobalanopsis than to other Quercus subg. Quercus species.

3.1 Chloroplast Genome Comparison and Codon Preference Analysis

Angiosperm chloroplast genomes typically range from 120,000 to 170,000 bp (Tangphatsornruang et al., 2010). The four Cyclobalanopsis species in this study (160,784-160,906 bp) showed highly conserved genome size, structure, gene number, total GC content, and region lengths/GC content. These characteristics are similar to other Fagaceae species, which exhibit greater gene density, stronger environmental adaptation, and higher stability compared to other families (Gao, 2020). Comparative analysis revealed high conservation across the four Cyclobalanopsis chloroplast genomes, with variation primarily in non-coding regions. Since Cyclobalanopsis identification relies mainly on cupule characteristics and many species are difficult to identify without fruits, highly variable chloroplast regions could be developed as DNA barcodes. This study identified $trnH\text{-}GUG_{trnQ\text{-}UUG,\ rpoC1}petN$, $trnT\text{-}UGC_{trnM\text{-}CAU,\ ycf4}psbJ$, and *ycf1* as potential DNA barcodes for Cyclobalanopsis identification. IR region expansion/contraction is common in plant evolution (Hansen et al., 2007; Davis & Soreng, 2010; Huang et al., 2014). Consistent gene content in IR regions across Cyclobalanopsis indicates conservative IR boundaries. All four species exhibited *ycf1* pseudogenes (*ycf1*), a phenomenon also observed in some Quercus subg. Quercus and *Castanea* species (Yang, 2018; Gao, 2020).

Codon usage bias is an important evolutionary phenomenon influenced by genome size, base mutation, genetic drift, natural selection, gene expression levels, tRNA abundance, GC content, and protein structure (Romero et al., 2000; Duret, 2000; Angellotti et al., 2007). The low GC3 content in this study' s Cyclobalanopsis species indicates preference for A/T-ending codons, a widespread phenomenon in angiosperms (Clegg et al., 1994; Tangphatsornruang et al., 2010; Delannoy et al., 2011). Similar RSCU patterns between Cyclobalanopsis and Quercus subg. Quercus suggest conserved codon preferences within Fagaceae, though intergeneric differences exist. The three plotting analyses demonstrated that codon preference in these four species is influenced by both selection pressure and mutation, with selection pressure being the dominant factor, consistent with findings in *Castanopsis carlesii* (Jiang et al., 2021), *Dalbergia odorifera* (Yuan et al., 2021), and *Erigeron breviscapus* (Li et al., 2021).

3.2 Phylogenetic and Interspecific Relationship Analysis

Phylogenetic results showing early divergence of *Fagus* and *Trigonobalanus* are consistent with previous fossil and GIS-based studies on Fagaceae origin (Zhou, 1999). Previous classifications based on pollen morphology and nuclear genes placed *Q. aquifolioides*, *Q. spinosa*, and *Q. engleriana* in the Ilex group and *Q. variabilis* in the Cerris group (Hubert et al., 2014). However, nuclear and plastid markers have shown Asian Ilex group species frequently interspersed within the Cerris group (Simeone et al., 2013; Hubert et al., 2014). Our study similarly

found that four Quercus subg. Quercus species did not form a monophyletic group, with *Q. spinosa* clustering with *Q. engleriana* and *Q. variabilis*, demonstrating interspersion between Ilex and Cerris groups consistent with previous molecular studies.

Flora of China, *Sylva Sinica*, *Flora Yunnanica*, and *Flora Guizhouensis* treat *Q. gambleana* and *Q. oxyodon* as separate species. However, Luo & Zhou (2001a) and Deng (2007) considered *Q. gambleana* a variety of *Q. oxyodon* based on stellate hairs on leaf abaxial surfaces, while Liu et al. (2008) suggested it should be a variety of *Q. longifolia* based on cluster analysis. Our phylogenetic analysis showed *Q. oxyodon* clustering with *Q. neglecta*, while *Q. gambleana* and *Q. oxyodon* did not form sister groups. Although *Q. gambleana* and *Q. oxyodon* share similar fruit and leaf morphology, they differ in stellate hair color and density on leaf abaxial surfaces, with substantial variation observed across different collection regions (Deng, 2007). Scanning electron microscopy reveals differences in pollen exine ornamentation: *Q. gambleana* shows granular-verrucate or aggregated-verrucate patterns, while *Q. oxyodon* exhibits granular ornamentation (Wang & Pu, 2004). Our *Q. gambleana* samples had dense yellow-brown stellate hairs, whereas *Q. oxyodon* had sparse white stellate hairs. Based on these combined molecular and morphological data, we support treating *Q. gambleana* and *Q. oxyodon* as separate species. The *Q. gambleana* sample was collected from Chongqing and *Q. oxyodon* from Guangxi, representing geographically distant populations. We hypothesize that substantial environmental differences and long-term restricted gene flow may have contributed to their divergence as distinct species.

References

AMIRYOUSEFI A, HYVÖNEN J, POCZAI P, 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes[J]. *Bioinformatics*, 34(17): 3030-3031.

ANGELLOTTI MC, BHUIYAN SB, CHEN GR et al., 2007. CodonO: codon usage bias analysis within and across genomes[J]. *Nucl Acid Res*, 35(suppl_2): W132-W136.

CHEN HY, HUANG CJ, 1998. *Flora Reipublicae Popularis Sinicae: Tomus 22* [M]. Beijing: Science Press: 263-334.

CLEGG MT, GAUT BS, LEARN GH et al., 1994. Rates and patterns of chloroplast DNA evolution[J]. *Proc Nat Acad Sci*, 91(15): 6795-6801.

DAVIS JI, SORENG RJ, 2010. Migration of endpoints of two genes relative to boundaries between regions of the plastid genome in the grass family (Poaceae)[J]. *Amer J Bot*, 97(5): 874-892.

DELANNOY E, FUJII S, COLAS DES FRANCS-SMALL C et al., 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes[J]. *Molec Biol Evol*, 28(7): 2077-2086.

DENG M, 2007. Anatomy, taxonomy, distribution & phylogeny of *Quercus* subg. *Cyclobalanopsis* (Oersted) Schneid (Fagaceae)[D]. Beijing: Chinese Academy of Sciences: 3-70.

DUAN MX, 1992. Studies on the Fagaceae wood-assortment in Taiwan[J]. *Fujian Sci Technol*, 19(4): 65-70.

DUAN MX, 1995. The comprehensive utilization of the resources of *Cyclobalanopsis* in China[J]. *J Beijing For Univ*, 17(2): 109-110.

DURET L, 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes[J]. *Trends Genet*, 16(7): 287-289.

FRODIN DG, GOVAERTS R, 1998. *World Checklist and Bibliography of Fagales* (Betulaceae, Corylaceae, Fagaceae and Ticodendraceae)[M]. London: the Royal Botanic Gardens: 201-394.

GAO XX, 2020. Comparative whole chloroplast genomics of *Castanea* and Fagales family[D]. Xi' an: Northwest University: 14-29.

HANSEN DR, DASTIDAR SG, CAI Z et al., 2007. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae)[J]. *Mol phylogenet evol*, 45(2): 547-563.

HUANG CJ, ZHANG YT, BRUCE B, 1999. *Flora of China* Vol.4[M]. Beijing: Science Press and Missouri Botanical Garden Press: 314–400.

HUANG H, SHI C, LIU Y et al., 2014. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships[J]. *BMC Evol biol*, 14(1): 1-17.

HUBERT F, GRIMM GW, JOUSSELIN E et al., 2014. Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*[J]. *Syst Biodivers*, 12(4): 405-423.

JIANG SZ, LIAN H, XIONG YF, et al., 2021. Analysis of codon bias in chloroplast genome of *Castanopsis carlesii*[J]. *Mol Plant Breed*, 26(7): 1-12.

JIN JJ, YU WB, YANG JB et al., 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes [J]. *Genome Biol*, 21(1): 1-31.

KATOH K, STANDLEY DM, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. *Mol biol evol*, 30(4): 772-780.

KEARSE M, MOIR R, WILSON A et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data[J]. *Bioinformatics*, 28(12): 1647-1649.

KORPELAINEN, 2004. The evolutionary processes of mitochondrial and chloroplast genomes differ from those of nuclear genomes [J]. *Naturwissenschaften*, 91(11): 505-518.

KUMAR S, STECHER G, TAMURA K, 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. *Mol biol evol*, 33(7): 1870-1874.

KUNMING INSTITUTE OF BOTANY, CHINESE ACADEMY OF SCIENCES, 1979. *Flora Yunnanica: Tomus 2* (spermatophyte)[M]. Beijing: Science Press: 322-325.

KWAK SY, LEW TTS, SWEENEY CJ et al., 2019. Chloroplast-selective gene delivery and expression in planta using chitosan-complexed single-walled carbon nanotube carriers[J]. *Nat nanotechnol*, 14(5): 447-455.

LAN KM, LI YK, 1982. *Guizhou botanical chronicles*[M]. Guiyang: Guizhou People' s Press: 103-105.

LIU LF, FANG ZJ, 1986. A study on pollen morphology of quercoideae oerst (fagaceae) in china[J]. *Guihaia*, 12(4): 243-251+323-327.

LIU LY, ZHANG ML, LI JQ, et al., 2008. A Numerical Taxonomic Study of the Genus *Cyclobalanopsis* Oersted from China[J]. *Wuhan Botan Res*, 26(5): 466-475.

LI XH, YANG SC, XIN YX, et al., 2021. Analysis of the codon usage bias of chloroplast genome in *Erigeron breviscapus* (Vant.) Hand-Mazz[J]. *J Yunnan Agr Univ*(Nat Sci Ed), 36(3): 384-392.

LUO Y, ZHOU ZK, 2001a. Leaf epidermis of *Quercus* subgen. *Cyclobalanopsis* (Oerst.) Schneid. (Fagaceae)[J]. *Acta Phytotaxon Sin*, 39(6): 489-501.

LUO Y, ZHOU ZK, 2001b. Phytogeography of *Quercus* subg. *Cyclobalanopsis*[J]. *Acta Bot Yunnanica*, 23(1): 1-16+28.

MAYOR C, BRUDNO M, SCHWARTZ JR et al., 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length[J]. *Bioinformatics*, 16(11): 1046-1047.

ROMERO H, ZAVALA A, MUSTO H, 2000. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces[J]. *Nucl acid res*, 28(10): 2084-2090.

SHARP PM, LI WH, 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications[J]. *Nucl acid res*, 15(3): 1281-1295.

SHI L, CHEN H, JIANG M et al., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer[J]. *Nucl acid Res*, 47(W1): W65-W73.

SIMEONE MC, PIREDDA R, PAPINI A et al., 2013. Application of plastid and nuclear markers to DNA barcoding of Euro-Mediterranean oaks (*Quercus*,

Fagaceae): problems, prospects and phylogenetic implications[J]. *Bot J Linn Soc*, 172(4): 478-499.

TANGPHATSORNRUANG S, SANGSRAKRU D, CHANPRASERT J et al., 2010. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships[J]. *DNA res*, 17(1): 11-22.

WANG PL, PU DF, 2004. *Pollen Morphology and Biogeography of Fagaceae*[M]. Guangzhou: Guangdong Sci Technol Press: 1-158.

XU YC, REN XW, 1976. The Classification and Distribution of Fagaceae of Yunnan Province (2)[J]. *J Syst Evol*, 14(2): 73-88.

YANG YC, 2018. Plastid genomes of *Quercus* in China and the key genera in Fagaceae: comparative and phylogenetic analyses[D]. Xi' an: Northwest University: 19-90.

YUAN XL, LI YQ, ZHANG JF, et al., 2021. Analysis of codon usage bias in the chloroplast genome of *Dalbergia odorifera*[J]. *Guihaia*, 41(4): 622-630.

ZHAO Z, WANG X, YU Y et al., 2018. Complete chloroplast genome sequences of *Dioscorea*: Characterization, genomic resources, and phylogenetic analyses[J]. *PeerJ*, 6(12): e6032.

ZHENG WJ, 1985. *Sylva Sinica*: Vol. 2 [M]. Beijing: China Forestry Publishing House: 2199-2356.

ZHOU ZK, 1999. Fossils of the Fagaceae and their implications in systematics and biogeography[J]. *Acta Phytotax Sin*, 37(4): 369-385.

ZUO LH, SHANG AQ, ZHANG S et al., 2017. The first complete chloroplast genome sequences of *Ulmus* species by de novo sequencing: Genome comparative and taxonomic position analysis[J]. *PLoS ONE*, 12(2): e0171264.

Note: Figure translations are in progress. See original paper for figures.