Comparative and Phylogenetic Analysis of the Complete Chloroplast Genomes of *Lithocarpus* Species (Fagaceae) in South China

Shi Shi , Ziyan Zhang , Xinhao Lin , Linjing Lu , Keyi Fu , Miaoxin He , Shiou Yih Lee , Hui Yin * , Jingwei Yu *

Posted Date: 28 April 2025

doi: 10.20944/preprints202504.2288.v1

Keywords: Chloroplast genomes; Comparative genomics; Lithocarpus; Phylogenetics analysis

Article

Comparative and Phylogenetic Analysis of the Complete Chloroplast Genomes of Lithocarpus Species (Fagaceae) in South China

Shi Shi [1,2], Ziyan Zhang [1], Xinhao Lin [2], Linjing Lu [2], Keyi Fu [2], Miaoxin He [1], Shiouyih Lee [3], Hui Yin [1,*] and Jingwei Yu [1,*]

[1] Guangdong Provincial Key Laboratory of Pharmaceutical Bioactive Substances and School of Basic Medical Sciences, Guangdong Pharmaceutical University, Guangzhou, China

[2] South China Limestone Plants Research Center and College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, China

[3] Faculty of Health and Life Sciences, INTI International University, Nilai, Malaysia

* Correspondence: huiyin0103@gdpu.edu.cn (H.Y.); yujingwei@gdpu.edu.cn (J.Y.)

Abstract: Background/Objectives: In South China, *Lithocarpus* species dominate mixed evergreen broadleaf forests, forming symbiotic relationships with ectomycorrhizal fungi and serving as food resources for diverse fauna, including frugivorous birds and mammals. The limited understanding of chloroplast genomes in this genus restricts our insights into its species diversity. This study investigates the chloroplast genome (cp genome) sequences from seven *Lithocarpus* species, aims to elucidates their structural variation, evolutionary relationships, and functional gene content, providing a foundation for future genetic conservation and breeding efforts. **Methods**: We isolated total DNA from fresh leaves and sequenced the complete cp genomes of these samples. Comparative cp genome analysis and phylogenetic analysis were conducted to develop genomic resources and elucidate the intraspecific phylogeny of *Lithocarpus*. **Results**: All studied species exhibited a conserved quadripartite chloroplast genome structure, with sizes ranging from 161,495 to 163,880 bp. Genome annotation revealed 130 functional genes and a GC content of 36.72–37.76%. Codon usage analysis showed a predominance of leucine-encoding codons. Our analysis identified 322 simple sequence repeats (SSRs), predominantly palindromic in structure (82.3%). These SSRs comprised 19 distinct categories, with similar type distributions observed across all examined species. Eight highly variable regions (ndhF, ycf1, ycf1, trnS-trnG-exon1, trnk(exon1)-rps16(exon2), rps16(exon2), rbcL-accD and ccsA-ndh) have been identified, which could serve as potential DNA markers for future population genetics or phylogeographic studies on this genus. The well-resolved plastome phylogeny tree provided critical insights into the evolutionary trajectory of Fagaceae, suggested that *Lithocarpus* was strongly supported as monophyletic, while Quercus was inferred to be polyphyletic, showed a significant cytonuclear discrepancy. **Conclusions**: We characterized and compared the chloroplast genome features across eight *Lithocarpus* species, followed by comprehensive phylogenetic analyses. These findings provide critical insights for resolving taxonomic uncertainties and advancing systematic research in this genus.

Keywords: chloroplast genomes; comparative genomics; *Lithocarpus*; phylogenetics analysis

1. Introduction

The genus *Lithocarpus* Blume, the second-largest genus in the family Fagaceae, comprising approximately 300–350 species distributed globally [1]. This genus is a critical component of montane and lowland ecosystems, where it plays a crucial role in shaping Northern Hemisphere forests, including temperate, subtropical, and tropical ecosystems[2,3]. Members such as *Quercus* (oaks), *Castanopsis* (chinquapins), and *Lithocarpus* (stone oaks) are dominant species in their habitats,

 doi:10.20944/preprints202504.2288.v1

reflecting their ecological adaptability and evolutionary success[4]. Among these, *Lithocarpus* is a key component of subtropical evergreen broad-leaved forests, yet its genomic background and evolutionary history remain uncleared compared to temperate relatives like *Quercus* and Fagus[5–7]. Beyond their ecological roles, several species hold significant economic value. For example, *L. litseifoliu* leaves are used to produce "sweet tea", a traditional beverage in south China, which has both medicinal and edible functions, exhibit potent antioxidant and anticancer activities[8]. Similarly, extracts from *L. polystachyus* rich in dihydrochalcones with demonstrated antidiabetic and anti-inflammatory propertie, underscoring the genus's potential in ethnopharmacology and modern drug discovery[9,10]. Despite their ecological and economic significance, the taxonomy of this genus remains challenging. Traditional identification based on morphological traits is not only unreliable but also time-consuming. In some cases, there are no obvious phenotypic differences among species, while morphological characteristics often show considerable intraspecific variations. Thus, the phylogenetic relationship of *Lithocarpus* still needs to be resolved.

Recent advances in molecular systematics and high-throughput sequencing technologies have significantly improved our understanding of the phylogeny of *Lithocarpus* [11,12]. Traditional morphological classifications have been challenged by the complex phenotypic variation within the genus, whereas molecular data have provided new insights into its evolutionary history. Studies have consistently supported the monophyly of *Lithocarpus* using plastid genes and nuclear markers, however, significant genetic heterogeneity observed in some widespread or ecologically diverse species suggests potential cryptic speciation or hybrid introgression events (Herliana 等 2023 年, Zhou, 2022, Yang, 2024; yang,2018,)[11–14].

Chloroplasts (cp) are a crucial organelle in plants responsible for photosynthesis, possessing a circular DNA (cpDNA) that is haploid and maternally inherited in angiosperms[15]. Chloroplast genome sequencing has revolutionized plant systematics and evolutionary biology by providing a robust molecular framework to resolve long-standing taxonomic ambiguities[16,17]. Unlike nuclear genomes, cpDNA exhibit a conserved quadripartite structure (large single-copy [LSC], small single-copy [SSC], and inverted repeat [IR] regions), low recombination rates, and maternal inheritance, making them ideal markers for reconstructing deep evolutionary divergences and recent radiations [18,19]. Comparative of cpDNA can discover lineage-specific structural variations, such as IR expansions/contractions, gene loss and inversions, which often correlate with ecological adaptations or historical biogeographic events. For instance, in Fagaceae, cpDNA analyses have clarified the polyphyletic nature of *Castanopsis* and *Lithocarpus*, revealing complex speciation driven by Pleistocene climatic fluctuations[20]. Moreover, codon usage bias analyses and selective pressure assessments (e.g., dN/dS ratios) in chloroplast genes (e.g., rbcL, matK, ndhF) have provided insights into adaptive evolution, particularly in response to light availability and temperature gradients[21]. These advancements highlight the dual utility of cpDNA: as phylogenetic anchors for resolving species boundaries and as functional genomic tools for detecting the molecular basis of ecological diversification.

In this study, we sequenced and analyzed the complete chloroplast genomes of eight ecologically divergent *Lithocarpus* species endemic to South China: *Lithocarpus uvarifolius*, *L. crassifolius*, *L. corneus*, *L. calophyllus*, *L. oleifolius*, *L. longipedicellatus*, *L. litseifolius*, and *L. longanoides*. Subsequently, we characterized the sequence structure and variation through a comparative assessment of these chloroplast (cp) DNA genomes. We analyzed codon usage bias and identified hotspots of high nucleotide diversity within the cp genomes. Additionally, we constructed a phylogenetic tree based on the cp genomes of 76 species to investigate the phylogeny and relationships within the *Lithocarpus* genus. The *Lithocarpus* species employed in this study are widely distributed across southern China, predominantly occupying low-altitude regions with similar ecological niches. Through these investigations, our study aims to elucidate the molecular-level variations among these species, deepen our understanding of their evolutionary relationships, and provide a foundation for the development and utilization of these plant resources.

 doi:10.20944/preprints202504.2288.v1

2. Materials and Methods

2.1. Collection and Preservation of Materials

Fresh leaves of Six *Lithocarpus* species (*L. uvarifolius*, *L. crassifolius*, *L. corneus*, *L. calophyllus*, *L. longanoides*, *L. oleifolius*, and *L. longipedicellatus*) were collected in July 2023 from natural populations in the Heishiding Mountain, Guangdong Province, China, while *L. longipedicellatus* was collected in August 2024 from Jianfengling Mountain, Hainan Province, China. All 7 specimens were collected across representative microhabitats within the reserve to ensure coverage of genetic variation. The plant materials were immediately silica-gel-dried and stored in a freezer at -80℃, and the voucher specimens were deposited in the Herbarium of South China Agricultural University (CANT) (Accession number: 32213~32220).

2.2. Chloroplast Genome Assembly and Comparative Analysis

2.2.1. DNA Extraction and Sequencing

Total DNA was extracted from the seven *Lithocarpus* species using a modified cetyltrimethylammonium bromide (CTAB) method. DNA integrity was assessed via 1% agarose gel electrophoresis, and concentrations were quantified using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Shanghai, China). Sequencing libraries were prepared with the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA; Catalog #E7370L), targeting an average fragment size of 350 bp. Paired-end sequencing (PE 150) was performed on the Illumina Novaseq Xplus platform, yielding about 6 GB of raw data per sample.

2.2.2. Chloroplast Genome Assembly and Annotation

Raw reads were preprocessed with Fastp v0.23.4 to remove adapters[22], trim low-quality bases ($Q < 20$), and discard reads with above 10% N content and the length shorter than 140 nt. Chloroplast genomes were assembled using GetOrganelle v1.7.7.1.[23] , utilizing the cp genome of L. longipedicellatus as the reference. The assembled chloroplast genome was annotated using online annotation software Geseq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) and CPGAVAS2 (http://47.96.249.172:16019/analyzer/annotate)[24,25]. These tools were employed to determine the start positions of the chloroplast genome and the IR regions, as well as to annotate the genes present. Finally, manual adjustments were employed to ensure the accuracy of the annotations and to identify any potential errors.

2.2.3. Comparative Analysis of Chloroplast Genomes

Considering that *L. litseifolius* is morphologically similar to the *Lithocarpus* species we collected and their distribution ranges overlap, we downloaded the complete chloroplast genome (NC_063927.1) of this specie for comparative analysis. mVISTA is a set of programs used to compare DNA sequences of millions of base pairs in length between two or more species, and to visualize these comparison results in conjunction with annotation information. Use the online tool mVISTA (https://genome.lbl.gov/vista/mvista/), we upload the aligned fasta file and each sample's separate annotation file at the same time for analysis[26]. Nucleotide Polymorphism (Pi) is a parameter that measures the level of polymorphism in a specific population, which refers to the average difference in nucleotides at each position between two randomly selected DNA sequences within the same population. This analysis uses Mega 7.0 for alignment and Dnasp5 software to calculate all sequences, with the window set to 400 bp and the step set to 200 bp[27,28].

2.2.3. IR Boundary Variation Analysis

The inverted repeat (IR) regions of chloroplast genomes are considered the most conserved regions; however, their boundary sequences can exhibit dynamic changes, including outward

 doi:10.20944/preprints202504.2288.v1

expansion or inward contraction. These variations can lead to alterations in the copy number of associated genes or the generation of pseudogenes in the boundary regions. Such phenomena are common during chloroplast genome evolution and are a primary cause of length variation among chloroplast genomes. In this study, we utilized the online software IRscope (https://irscope.shinyapps.io/irapp/) to analyze the genetic structure near the junctions connecting the IR regions with the short single-copy (SSC) and long single-copy (LSC) regions[29]. The analysis was performed by uploading the GenBank-annotated files of eight *Lithocarpus* species. This approach allowed us to visualize and compare the structural dynamics at the IR boundaries across the eight species, providing insights into the evolutionary mechanisms driving the divergence of chloroplast genome architecture within the genus *Lithocarpus*.

2.2.4. Repetitive Sequence Analysis

Repetitive sequences are known to play a crucial role in genomic rearrangements and recombination, and they may also exhibit phylogenetic information within certain populations. In this study, we employed the REPuter software (https://bibiserv.cebitec.uni-bielefeld.de/reputer) to quantify the number of long sequence repetitive fragments in each sample[30]. The REPuter software operates on the principle of detecting repetitive sequences within a genome using algorithms. It can identify various types of repeats, including forward (f), reverse (r), complement (c), and palindromic (p) repeats. For our analysis, we set the Hamming Distance to 3, the Maximum Computed Repeats to 50, and the Minimal Repeat Size to 30. Meanwhile, we utilized the MISA (Microsatellite Identification Tool) online platform [8] to analyze various data related to SSRs, including the number of repetitions, repeat units, and lengths. The primary principle of this tool is to scan the given genomic sequences to search for repeat units within specific length ranges. These repeat units can consist of sequences of 1 to 6 nucleotides, such as mononucleotides (e.g., repetitions of A, T, C, G) and dinucleotides (e.g., repetitions of AT, CG, etc.).

2.2.5. Codon Usage Bias Analysis

The relative probability of a specific codon among its synonymous codons reflects the degree of codon bias. Relative synonymous codon usage (RSCU) values were calculated to quantify codon bias[31]. This analysis is crucial for understanding species evolutionary pressures and further genetic research. In this analysis, we used the DAMBE software to calculate RSCU values for each sample individually. During the analysis, duplicate genes were removed, and the editing of non-ATG start codons for methionine was considered to ensure accurate results.

2.2.6. Phylogenetic Tree Construction

To analyze the phylogenetic relationships within Fagaceae, we conducted a plastid genome phylogenomic study using 71 complete chloroplast genomes. The dataset including 11 *Castanopsis* species, 28 *Quercus* species, 25 *Lithocarpus* plastomes, and 5 *Castanea* species, with *Carpinus laxiflora* and *Morella salicifolia* setted as the outgroup. 7 plastomes of *Lithocarpus* were newly sequenced in this study, and other 68 were sourced from NCBI (Accession numbers of the complete chloroplast genomes are pro-vided in Supplementary Table S1).

Sequence alignment was performed using MAFFT 7.490 with default parameters[32]. The optimal nucleotide substitution model (K3Pu+F+I+G4) was selected using ModelFinder. Maximum likelihood (ML) phylogenetic analysis was conducted in IQ-TREE v2.4.0 with 1000 ultrafast bootstrap replicates under default settings[33]. Alignments were trimmed in Geneious, and phylogenies were inferred using maximum likelihood (ML) [34]. Trees were visualized and edited in FigTree v1.4.3[35].

 doi:10.20944/preprints202504.2288.v1

3. Results

3.1. Chloroplast Genome Structure Characteristics

In this study, we sequenced and assembled the complete chloroplast genome of eight *Lithocarpus* species (*L. uvarifolius*, *L. crassifolius*, *L. corneus*, *L. calophyllus*, *L. oleifolius*, *L. longipedicellatus*, *L. litseifolius*, and *L. longanoides*), with raw sequencing data deposited in the NCBI GenBank database (Table1). The lengths of nucleotide sequences of eight species ranged from 162615 bp in *L. longipedicellatus* to 161212 bp in *L. oleifolius*. There are four distinct regions in the chloroplast genomes of each of the 8 species studied: a large single copy (LSC), a small single copy (SSC), and two inverted repeats (IRa and IRb). A representative circular chloroplast genome structure of L. longipedicellatus is shown in Figure 1. All eight species displayed similar gene content, with 130 functional genes identified: 86 protein-coding genes and 36 transfer RNA (tRNA) genes. The GC content showed minimal variation across species (36.72-36.76%), further supporting structural conservation (Table 2).

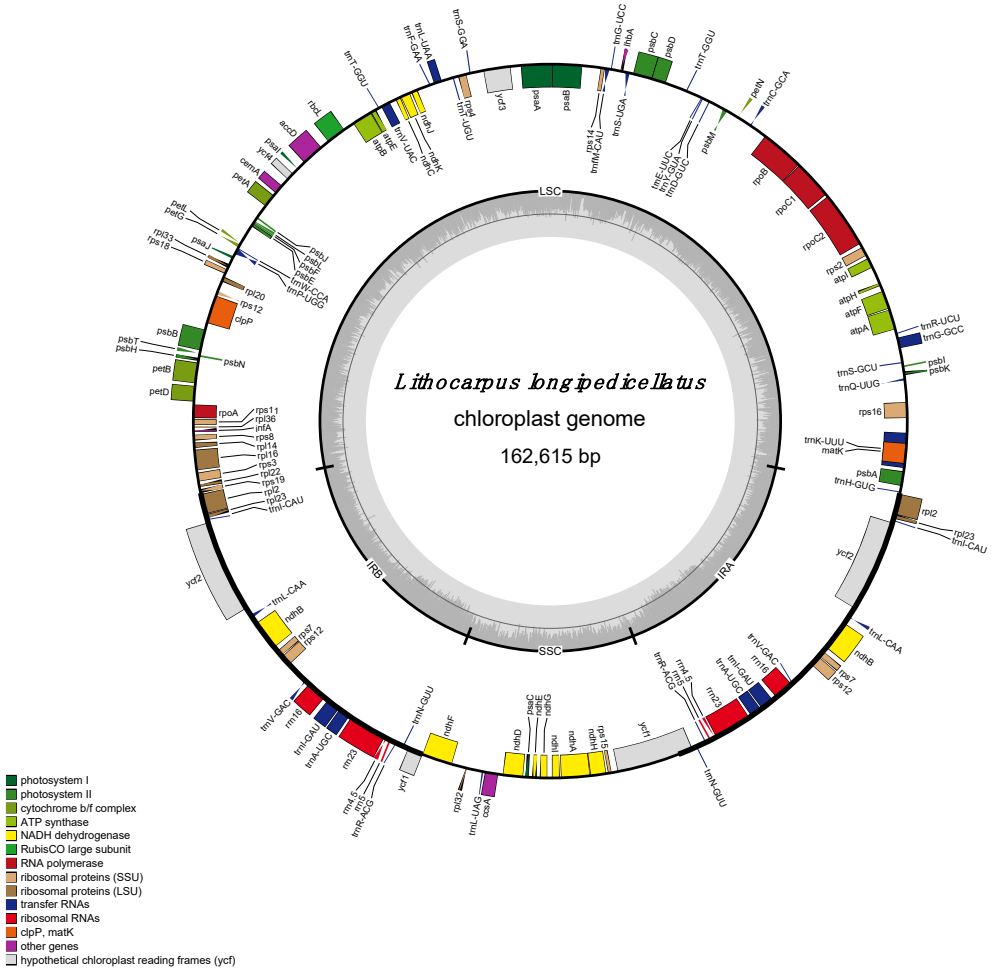

Figure 1. Genome map of *L. longipedicellatus* presents a typical chloroplast genome structure and content(a). Genes drawn in circles are transcribed clockwise, whereas genes outside the circle are transcribed counterclockwise. Different colors are annotated according to the different functions of the genes.

 doi:10.20944/preprints202504.2288.v1

Table 1. General characteristics of the 8 newly sequenced chloroplast genomes.

Species	Chloroplast Genome Size/bp	IR Length/bp	Overall GC Content %	Gene Number	tRNA Genes	Protein Coding Genes	GenBank Accession Number
Lithocarpus longipedicellatus	162615	25897	36.74	130	36	86	PP234611.1
Lithocarpus longanoides	161259	25881	36.75	130	36	86	PV191269
Lithocarpus oleifolius	161212	25883	36.74	130	36	86	OR805597.1
Lithocarpus litseifolius	161322	25897	36.73	130	36	86	NC_063927.1
Lithocarpus calophyllus	161225	25880	36.73	130	36	86	PP234612.1
Lithocarpus corneus	161265	25897	36.73	130	36	86	PP234614.1
Lithocarpus crassifolius	161331	25906	36.72	130	36	86	PP234613.1
Lithocarpus uvariifolius	161396	25899	36.76	130	36	86	PP234615.1

Table 2. Gene composition of the 7 newly sequenced chloroplast genomes.

Category	Gene group	Genes
Photosynthesis	Subunits of photosystem I	psaA, psaB, psaC, psaI, psaJ
	Subunits of photosystem II	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT
	Subunits of NADH dehydrogenase	ndhA*, ndhB*(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK
	Subunits of cytochrome b/f complex	petA, petB*, petD*, petG, petL, petN
	Subunits of ATP synthase	atpA, atpB, atpE, atpF*, atpH, atpI
	Large subunit of rubisco	rbcL
	Subunits photochlorophyllide reductase	-
Self-replication	Proteins of large ribosomal subunit	rpl14, rpl16*, rpl2*(2), rpl20, rpl22, rpl23(2), rpl32, rpl33, rpl36
	Proteins of small ribosomal subunit	rps11, rps12**(2), rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7(2), rps8
	Subunits of RNA polymerase	rpoA, rpoB, rpoC1*, rpoC2
	Ribosomal RNAs	rrn16(2), rrn23(2), rrn4.5(2), rrn5(2)
	Transfer RNAs	trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC*, trnG-UCC, trnH-GUG, trnI-CAU(2), trnI-GAU*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU(2), trnT-UGU, trnV-GAC(2), trnV-UAC*, trnW-CCA, trnY-GUA, trnfM-CAU
Other genes	Maturase	matK
	Protease	clpP**

 doi:10.20944/preprints202504.2288.v1

	Envelope membrane protein	cemA
	Acetyl-CoA carboxylase	accD
	c-type cytochrome synthesis gene	ccsA
	Translation initiation factor	infA
	other	-
Genes of unknown function	Conserved hypothetical chloroplast ORF	lhbA, ycf1(2), ycf2(2), ycf3**, ycf4

Notes: Gene *: Gene with one introns; Gene**: Gene with two introns; #Gene: Pseudo gene; Gene(2): Number of copies of multi-copy genes.

3.2. Comparative Analysis of Chloroplast Genomes

To accurately estimate sequence variation, we conducted multiple comparative analyses of the chloroplast genomes of the eight *Lithocarpus* species previously mentioned. The analysis results showed that low degree of variation in the chloroplast genome sequences for the eight species; however, the variation of non-coding region was higher than that of coding region (Figure 2).

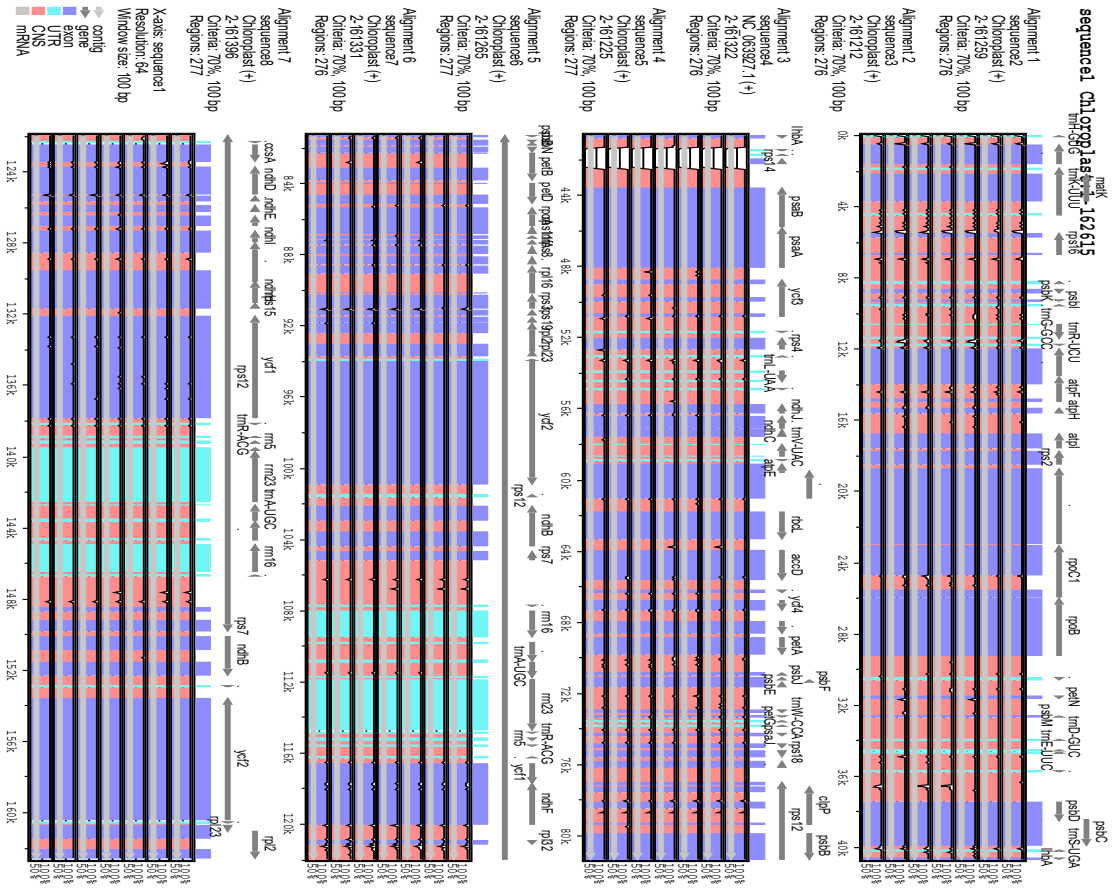

Figure 2. Subsequent sequence identity plots depict the chloroplast genome sequences of 8 *Lithocarpus* species. The orientation of the genes is indicated by the grey arrows; non-coding sequences are represented by red bars, exons by purple bars, and introns by blue bars. The vertical scales indicate percentage identity in the range of 50% to 100%.

Nucleotide polymorphism (Pi) is a parameter used to measure the level of polymorphism within a specific population, defined as the mean of nucleotide differences at each position between two randomly selected DNA sequences from the population. Nucleotide polymorphism (Pi) can reveal the extent of nucleic acid sequence variation among different species, and regions with higher

 doi:10.20944/preprints202504.2288.v1

variation may provide potential molecular markers for population genetics research. Aims to quantify the nucleotide polymorphism levels of these 8 chloroplast genomes more precisely and identify the mutational hotspot regions among them, we estimated the Pi values of these genomes in the aligned matrix with a window of 400 bp, which ranged from 0 to 0.02312.Further filtering the window of the top 5% of Pi values, we identified the lowest Pi value of 0.01161 as a threshold and accordingly detected five potential molecular marker regions that were considered as mutational hotspots due to having high Pi values (>0.01161)(Figure3). Eight regions of the chloroplast genes and gene spacer regions of the eight *Lithocarpus* species were identified as highly variable in the LSC region (Pi > 0.01161), namely ndhF (0.01768), ycf1(0.01625), ycf1(0.01161), trnS-trnG-exon1(0.01366), trnk(exon1)-rps16(exon2) (0.01911), rps16(exon2) (0.0183), rbcL-accD (0.02312) and ccsA-ndhD (0.1312).

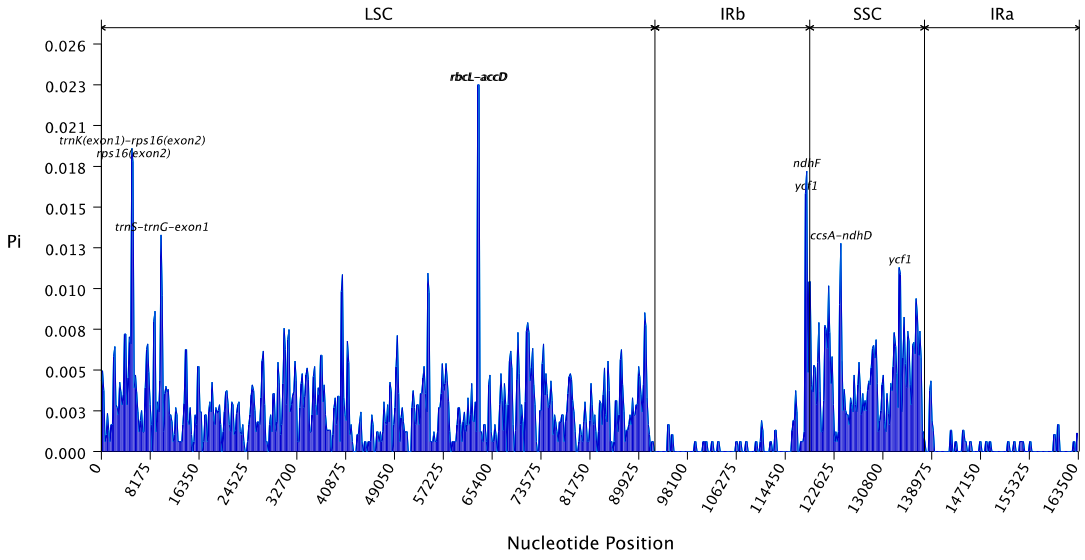

Figure 3. Nucleotide polymorphisms (Pi) of the 8 *Lithocarpus* chloroplast genomes. Each window has its nucleotide diversity shown by the y-axis, whereas each window has a position represented by the x-axis of the sliding window.

3.3. IR Region Contraction and Expansion (Analysis of IR Boundary Variation)

The regulation of the inverted repeat (IR) region size through contraction and expansion processes is fundamental to the dynamics of chloroplast genome architecture and serves as the primary factor influencing its size variability. Comparative analysis of the chloroplast genomes of eight *Lithocarpus* species revealed size variations ranging from 161,212 bp (*L. oleifolius)* to 162,615 bp (*L. longipedicellatus*), with the latter exhibiting the largest IR region (51,794 bp). Structural alignment identified species-specific shifts in the IR/SSC junction positions, suggesting potential adaptive divergence in these regions during evolutionary diversification. Using IRscope technology, we meticulously documented the dynamic changes in chloroplast genome sizes across the eight *Lithocarpus* species (Figure 4). The analysis indicated that the IR region lengths were relatively conserved, showing no significant contraction or expansion. In contrast, the SSC (small single-copy) and LSC (large single-copy) regions exhibited length variations ranging from 17 to 161 bp, with most differences limited to a few base pairs.

Our analysis discovered that five genes-rps19, rpl2, ndhF, ycf1, and trnH-located nearby the LSC-IRb, SSC-IRb, SSC-IRa, and LSC-IRa boundaries. Notably, rpl19 and trnH were entirely confined to the LSC region, positioned adjacent to the LSC-IRb and LSC-IRa boundaries without crossing into the IR regions. The ndhF gene, mainly located in the SSC region, was found to span the IRb-SSC

 doi:10.20944/preprints202504.2288.v1

boundary in *L. crassifolius*, *L. litseifolius*, *L. calophyllus*, *L. oleifolius*, and *L. longanoides*, with a minimal extension of 2-7 base pairs extended into the IRb region. The ycf1 gene was identified to span both the IRb-SSC and SSC-IRa boundaries among all eight *Lithocarpus* species. At the IRb-SSC boundary, this gene predominantly resided within the IRb region, exhibiting minimal extension (5-23 bp) into the SSC region. In contrast, a more substantial genomic span was observed at the SSC-IRa boundary. The SSC-region segment of ycf1 demonstrated high sequence conservation, maintaining a consistent length ranging from 4,579 to 4,599 bp. However, its IRa-region counterpart displayed marked length variation, spanning 875-1,115 bp.

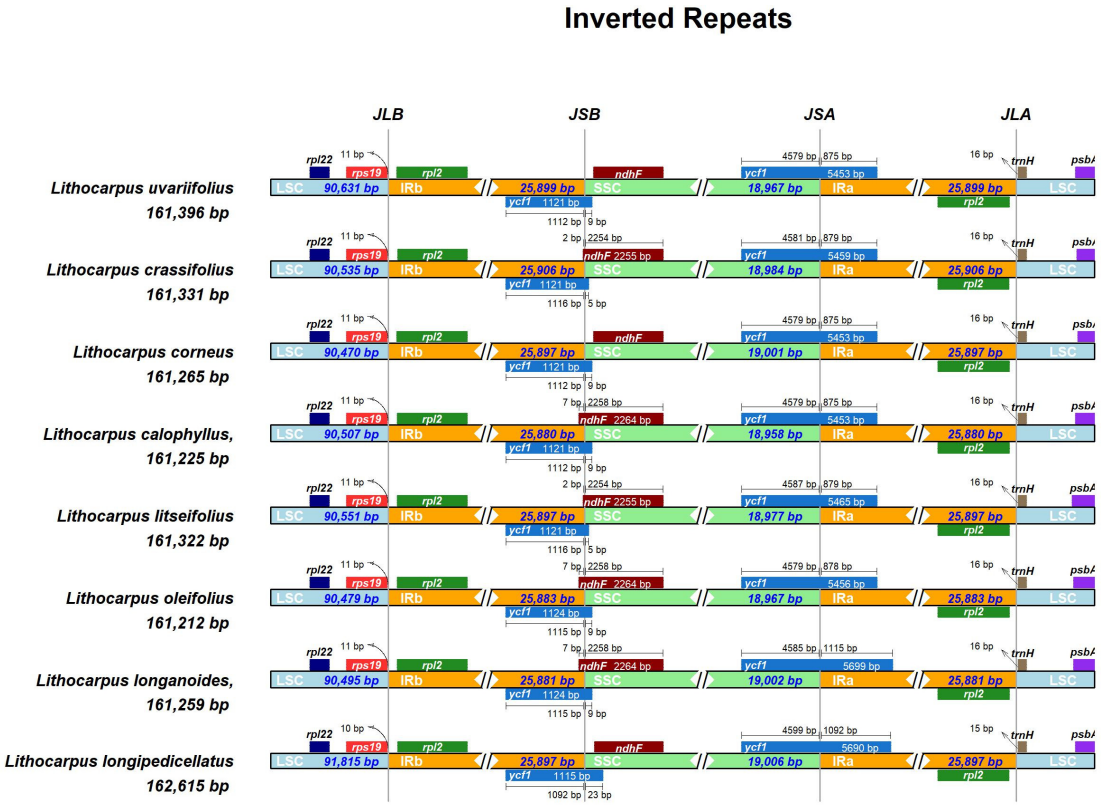

Figure 4. Comparison of the Comparison of the boundaries of large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions in the *Lithocarpus* cpDNAs. Note: Genes are illustrated as horizontal bars in schematic diagrams, and the intervals and boundaries between genes are represented by base pair length. Structural extensions such as exon elongations or regulatory regions should be clearly marked above the corresponding bars.

3.4. Repeat Analysis

Simple sequence repeats (SSRs) are abundant, highly polymorphic, uniformly distributed across the entire genome, co-dominantly inherited, and easy to detect, making them the second-generation molecular markers widely used in genetic map construction, target gene localization, genetic diversity research, molecular assisted breeding, and germplasm resource identification. Previous studies have emphasized the presence of simple repetitive motifs in the chloroplast genome, which are associated with various genomic rearrangements, recombination events, and large inversions. In this study, we compared eight species of the genus *Lithocarpus* and identified a total of 19 SSRs. The distribution of SSR types in the chloroplast genome of each species was similar, ranging from 14 to

 doi:10.20944/preprints202504.2288.v1

15 types. Mononucleotide repeats (A/T and C/G) accounted for the largest proportion (80.7%-83.7%), followed by dinucleotide repeats (AG/CT and AT/AT) (5.2%-7.9%). The longest SSR identified was a hexanucleotide repeat (AAATAT/ATATTT, AATATT/AATATT, AAAAAT/ATTTTT, AATACT/AGTATT, and AAAGAT/ATCTTT) (Figure 5).

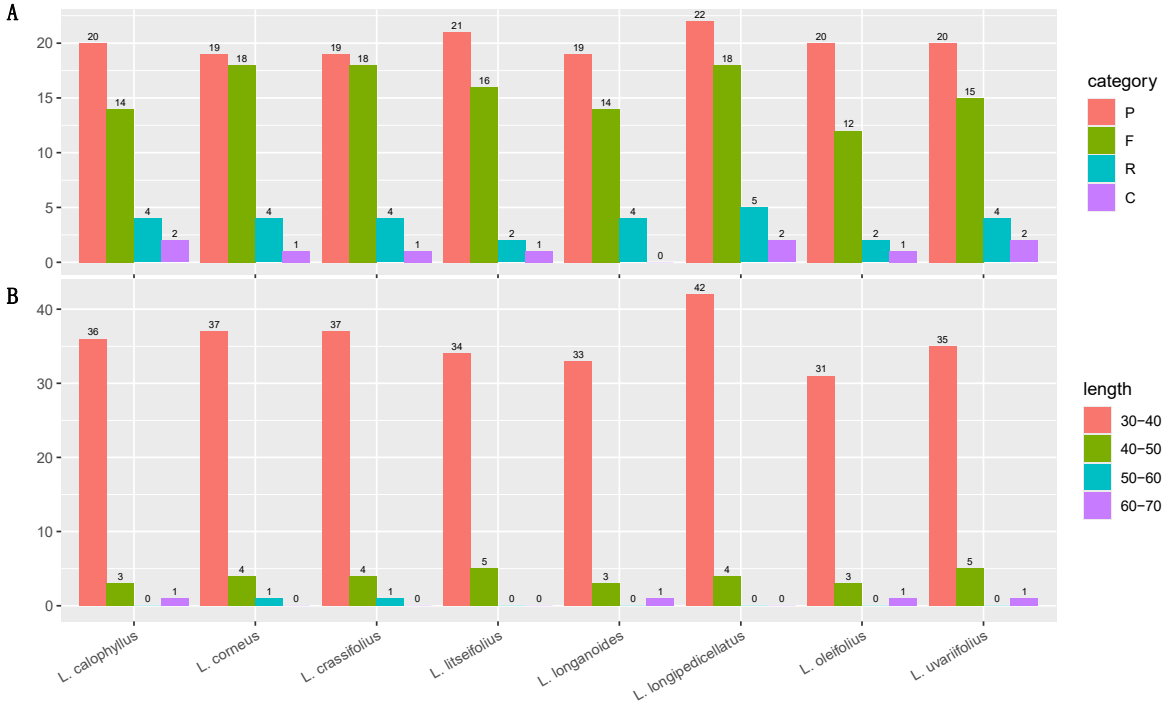

Figure 5. Quantitative analysis of long repeat sequences in eight *Lithocarpus* species. (A) Proportion of four types of long repeat sequences in each species: p represents palindromic sequences, F represents forward repeat sequences, R represents reverse repeat sequences, and C represents complementary sequences. (B) Proportion of repeat sequences of different lengths in each species.

Meanwhile, the type, amount, and length of the long repeat sequences were analyzed in eight *Lithocarpus* species. The analysis discovered a total of 322 significant repeat sequences, comprising 160 palindromic sequences, 125 forward repeats, 27 reverse repeats, and 10 complementary repeats. Among these investigated species, the chloroplast genome of *L. oleifolius* possesses the fewest forward repeats, with only 12 repeats identified, whereas the seven other *Lithocarpus* species exhibit significantly higher repeats numbers ranging from 14 to 18. *L. oleifolius* cpDNA also exhibits the minimal number of reverse tandem repeats, with only 2 repeats identified, whereas other *Lithocarpus* species display significantly higher counts ranging from 4 to 5. Notably, the quantity of palindromic repeats remains relatively conserved across species, approximately 19-22 copies, although *L. longanoides* lacks complement repeats entirely, in contrast to the 1-2 copies observed in other species (Figure 6A). The long repeat sequences were predominantly distributed within 30-40 bp, with 285 instances constituting 88.51% of the total. Notably, *L. corneus* and *L. crassifolius* exclusively exhibited repetitive elements within 50-60 bp, a distinct pattern diverging from the general trend (Figure 6B).

 doi:10.20944/preprints202504.2288.v1

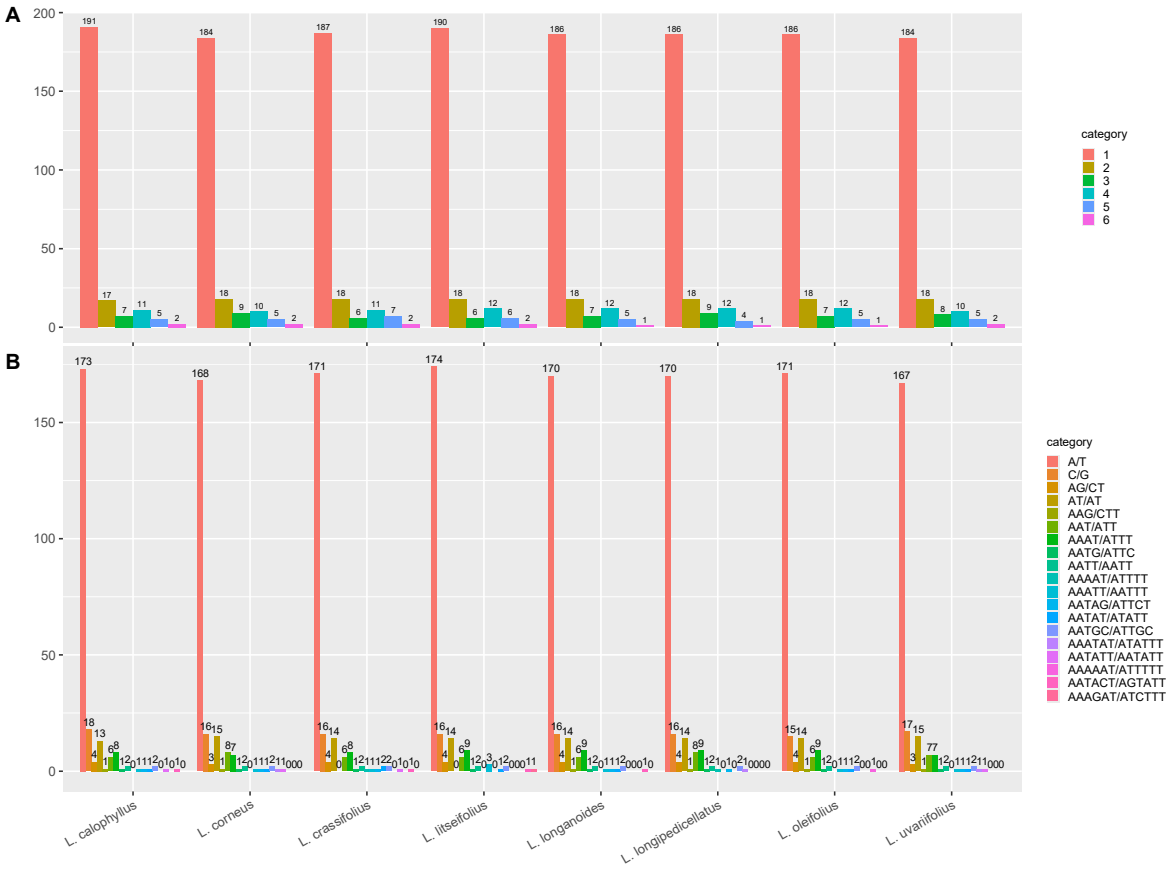

Figure 6. Quantitative analysis of SSRs in eight *Lithocarpus* species. (A) Proportion of SSRs with different repeat unit lengths in each species. (B) Frequency and types of identified SSRs.

3.5. Code Usage Bias Analysis

The relative probability of a specific codon being used among synonymous codons encoding the same amino acid reflects the degree of codon usage bias. This bias can be quantified by calculating the Relative Synonymous Codon Usage (RSCU), which measures the deviation from equal usage of synonymous codons. Investigating codon usage patterns is significant important for elucidating species-specific evolutionary pressures and advancing genetic research. In this study, we employed DAMBE to compute RSCU values on a sample-by-sample basis, followed by aggregation of the results. The analysis excluded redundant genes to avoid bias and accounted for non-ATG start codons encoding methionine (M) during sequence annotation. This approach ensures robustness in identifying evolutionary constraints and translational efficiency patterns across the studied samples. Our analysis revealed that the chloroplast genomes of the eight investigated *Lithocarpus* species collectively comprised 63 synonymous codons responsible for encoding 20 distinct amino acids (Figure 7). Among these amino acids, Leucine was predominantly encoded by the UUA codon, exhibiting a high RSCU value exceeding 1.90 (range:1.9725-2.016), and Arginine (Arg) showed a preference for the AGA codon, with the RSCU value ranging from 1.8373 to 1.8584. In contrast, Tryptophan (Trp) had RSCU values precisely at 1.00. Notably, these *Lithocarpus* species exhibited similar codon usage patterns, this similarity provides valuable reference data for further phylogenetic analysis.

 doi:10.20944/preprints202504.2288.v1

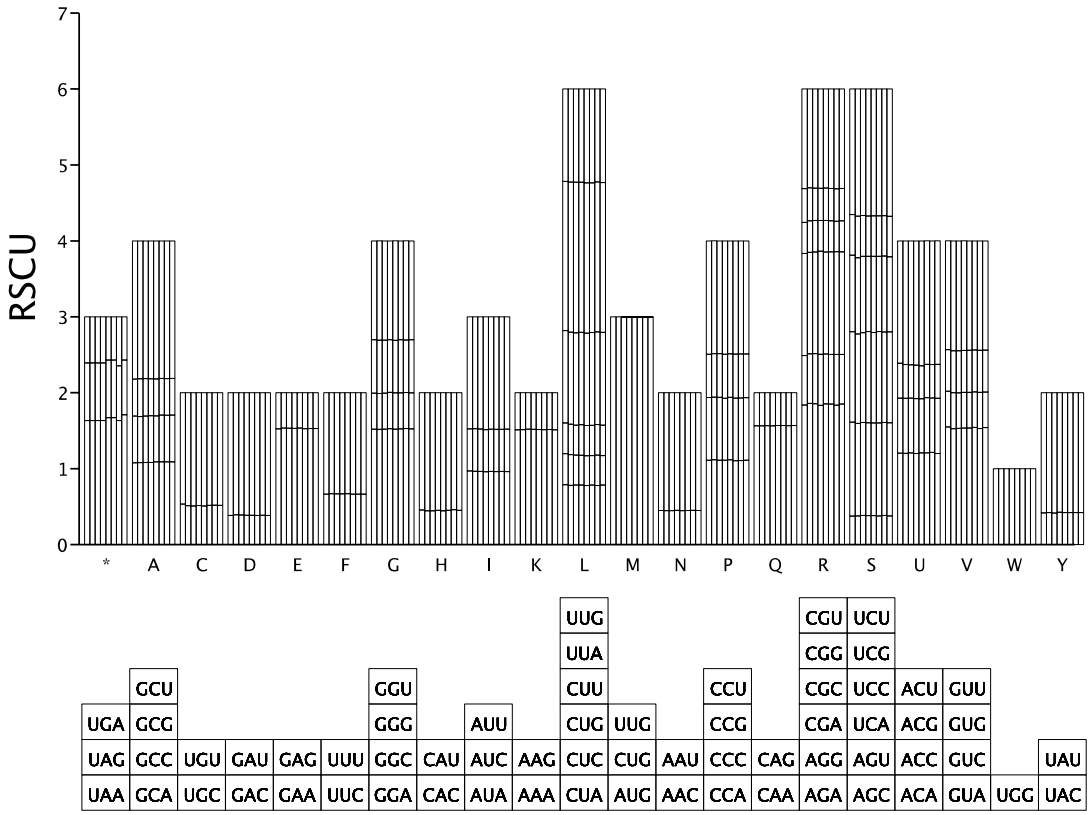

Figure 7. Bar chart illustrating codon usage bias across eight *Lithocarpus* species. The x-axis denotes the 20 essential amino acids and termination codon (*), while the y-axis represents the codons employed for each amino acid and their corresponding usage frequencies within the analyzed dataset.

3.6. Phylogenetic Analysis

Phylogenetic analysis based on maximum likelihood (ML) methods yielded a highly supported tree topology (Figure 8). All Fagaceae genera except *Quercus* demonstrated strong support of monophyly. *Castanopsis* and *Castanea* formed a well-supported sister clade (BS =100, PP= 1.00), which clustered with the *Cerris* subgenus of *Quercus* encompassing sections *Cyclobalanopsis* and *Ilex*. In contrast, the remaining *Quercus* species resolved as a grade in three successive divergences, and *Quercus* subgenus *Quercus* occupied the basalmost position in the entire Fagaceae phylogeny (BS = 100, PP = 1.00), representing the earliest-diverging lineage of the family. *Lithocarpus* was resolved into two clades (Clade I and Clade II). Clade I was further divided into two subclades: Subclade I-a, mainly consisting of species from Southwest China, including *L. longipedicellatus*, while Subclade I-b, comprising species distributed in central-south China. Within clade I-b, three distinct sister-taxa pairs were identified: *L. oleifolius* and *L. longanoides*, *L. litseifolius* and *L. crassifolius*, *L. corneus* and *L. uvariifolius*. These groupings exhibited high levels of phylogenetic support (BS >94.9, PP= 1.00). Clade II comprised *L. dealbatus* and *L. cleistocarpus* (BS = 100, PP = 1.00), forming a well-supported monophyletic group.

 doi:10.20944/preprints202504.2288.v1

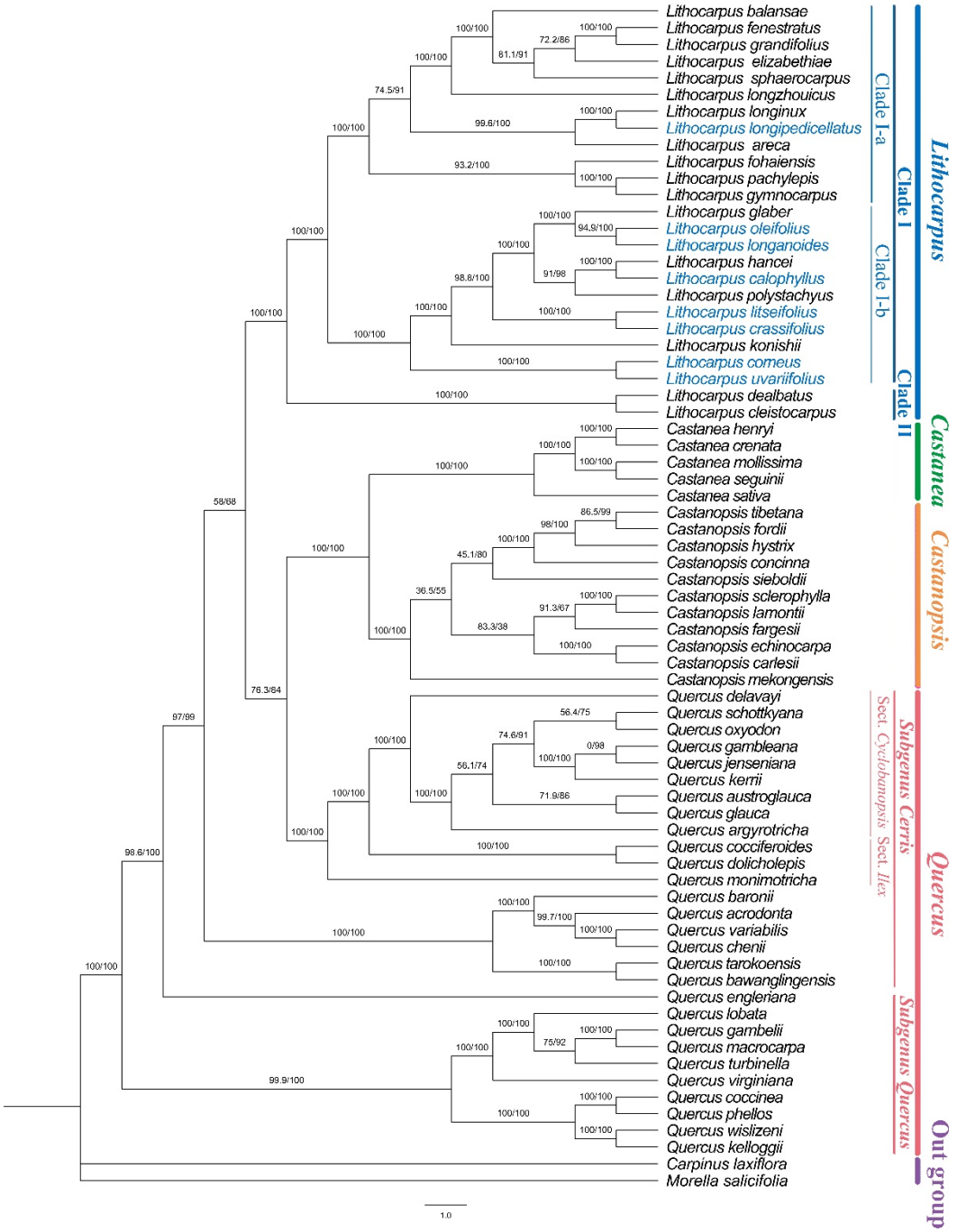

Figure 8. Phylogenetic analysis of chloroplast genomes from 71 *Fagaceae* species. Blue bold font indicates the eight investigated *lithocarpus* species.

4. Discussion

The comparative analysis of chloroplast (cp) genomes in eight *Lithocarpus* species revealed critical insights into their structural conservation, evolutionary dynamics, and potential molecular markers, aligning with recent advances in cp genomics and phylogenomics. Our findings demonstrate that *Lithocarpus* cp genomes exhibit remarkable structural stability, consistent with trends observed in other *Fagaceae* genera such as *Quercus* and *Castanopsis* [36]. This level of conservation is vital for preserving genomic stability and ensuring proper chloroplast functionality. The conserved gene content (130 functional genes) and minimal GC content variation (36.72–36.76%) across species underscore the evolutionary constraints on core photosynthetic and metabolic functions, a pattern similarly documented in stress-adapted angiosperms (Wang et al., 2023).

 doi:10.20944/preprints202504.2288.v1

Meanwhile, this conservation aligns with prior studies on *L. dealbatus* and *L. hancei*, where IR regions demonstrated higher GC content (42.7%) than single-copy regions due to rRNA genes[37]. Notably, the IR/SSC boundary shifts in ycf1 and ndhF genes—observed in five species—may reflect adaptive divergence, as similar junction dynamics have been linked to ecological speciation in *Fagaceae*[38]. However, the quadripartite chloroplast structure (LSC/SSC/IRa/IRb) exhibited minimal length variation (161,212–162,615 bp) and near-identical GC content (36.72–36.76%), this data suggests species-specific adaptations, potentially linked to ecological niche differentiation in *Lithocarpus*, which dominates subtropical evergreen forests under different light conditions[39,40].

Nucleotide polymorphism (Pi) is a parameter used to measure the level of polymorphism within a specific population, defined as the mean of nucleotide differences at each position between two randomly selected DNA sequences from the population. Nucleotide polymorphism (Pi) can reveal the extent of nucleic acid sequence variation among different species, and regions with higher variation may provide potential molecular markers for species identification research. In our study, five hypervariable regions (e.g., ndhF, ycf1, and rbcL-accD) with Pi values >0.01161 were identified as potential targets for molecular marker development. Notably, ycf1, which implicated in oxidative stress tolerance and seem as the most promising DNA barcode of land plants[41,42], exhibited dual high-Pi value (0.01625 and 0.01161), suggesting its role in adaptive divergence among *Lithocarpus* species. These results corroborate recent studies highlighting ycf1 as a hotspot for cp genome evolution in *Fagaceae*[39]. Furthermore, the rbcL-accD spacer (Pi = 0.02312) emerged as the most polymorphic region, consistent with its recognized utility in species-level phylogenetics[43]. These markers resolve taxonomic uncertainties, such as the sister relationship between *L. oleifolius* and *L. longanoides*, contradicting morphological classifications that grouped *L. oleifolius* with *L. litseifolius*. These markers might provide clues to resolve taxonomic ambiguities in *Lithocarpus*, particularly for morphologically cryptic species, a persistent challenge in Fagaceae systematics.

The regulation of the inverted repeat (IR) region size through contraction and expansion processes is fundamental to the dynamics of chloroplast genome architecture and serves as the primary factor influencing its size variability. IR boundary dynamics, particularly the extension of ycf1 into the IRa region (up to 240 bp), reflect lineage-specific expansions, a phenomenon increasingly reported in Fagaceae. Such expansions may stabilize cp genome architecture against deleterious mutations, as IR regions are known to buffer structural variations. The observed ndhF overlap with IRb (2–7 bp) further supports the hypothesis that IR boundary shifts drive cp genome diversification in closely related species.

Simple sequence repeats (SSRs) are abundant, highly polymorphic, uniformly distributed across the entire genome, co-dominantly inherited, and easy to detect, making them the second-generation molecular markers widely used in genetic map construction, target gene localization, genetic diversity research, molecular assisted breeding, and germplasm resource identification. Previous studies have emphasized the presence of simple repetitive motifs in the chloroplast genome, which are associated with various genomic rearrangements, recombination events, and large inversions. The predominance of mononucleotide SSRs (A/T, 80.7–83.7%) aligns with cp genome trends in angiosperm chloroplast genomes. These SSRs, coupled with abundant palindromic repeats (160/322), may facilitate genomic rearrangements and adaptive evolution. However, the functional implications of these repeats in *Lithocarpus* require experimental validation, such as assessing their correlation with phenotypic traits under environmental stress. SSR and repeat sequence analyses further highlighted evolutionary constraints. The reduced repeat numbers in *L. oleifolius* (12 forward vs. 14–18 in others) may indicate lineage-specific genomic streamlining. Codon usage bias favored UUA (Leucine, RSCU >1.90) and AGA (Arginine), consistent with *Lithocarpus* codon adaptation patterns, potentially reflecting translational efficiency optimization.

Our maximum likelihood phylogenetic reconstruction provided critical insights into the evolutionary trajectory of Fagaceae. As shown by the phylogenetic tree, *Lithocarpus*, *Castanopsis*, and *Castanea* were resolved as monophyletic groups, consistent with previous studies [11,13]. *Quercus* has consistently been resolved as a monophyletic group in earlier phylogenetic

 doi:10.20944/preprints202504.2288.v1

studies based on nuclear genes [44–46]. Notably, in this study, the infrageneric taxa of *Quercus* based on chloroplast genomes was inferred to be polyphyletic, revealing a striking nuclear-cytoplasmic phylogenetic incongruence. This cytonuclear discrepancy aligns with findings from Zhou et al.[12], who proposed that the cooling climate of the Miocene epoch, accompanied by a sea-level drop, led to the re-emergence of land bridges (such as the Bering Strait), facilitated interspecific hybridization of Quercus between Eurasian and North American, ultimately driving divergence in the genetic histories of nuclear and plastid genes.

In the basal branch of subgenus *Cerris*, species from the *Ilex* section (e.g., *Q. tarokoensis*, *Q. bawanglingensis*) from southern Chinese islands (Hainan and Taiwan) were nested within species from the *Cerris* section (e.g., *Q. chenii*, *Q.variabilis*, and *Q. acrodonta*) widely distributed in mainland China. This pattern, consistent with Hubert et al. [44], may reflect incomplete lineage sorting or gene introgression.

This study analyzed the maternal evolutionary history of *Lithocarpus* using chloroplast genome data. A phylogenomic tree based on chloroplast genomes robustly resolved *Lithocarpus* into two major clades. One clade comprised two species from southwestern China-India and central China. Another clade included two subclades corresponding to southwestern China (subclade I-a) and central-southern China (subclade I-b), indicating potential phylogeographic structure in the chloroplast genomes of *Lithocarpus*, consistent with previous studies [11,13].

Species collected in Guangdong in this study clustered within subclade I-b. *Lithocarpus corneus* and *Lithocarpus uvariifolius* were clustered with high support, forming a distinct monophyletic group. These two species frequently share overlapping habitats in Guangdong, where they act as ecologically associated species, and share similar morphological traits, including deeply cupulate cupules that envelop over half of their large nuts. In Guangdong, their fruits are harvested as the shared medicinal name "Fengliu fruit" in local markets. They are distinguished by *L. corneus* having larger fruits and more elongated leaves.

Numerous studies have shown significant topological incongruence between nuclear and chloroplast genome datasets in *Lithocarpus*. This incongruence may arise from factors such as chloroplast genome convergence, introgression, incomplete lineage sorting, or differential rates of gene flow mediated by pollen and seeds[44–46].

5. Conclusions

In present study, the complete chloroplast genomes of seven *Lithocarpus* species were sequenced and assembled，revealing that highly conserved genomic architecture with minimal variation in size (161,495-163,880 bp), GC content (36.72-37.76%), and repeat element distribution. Comparative genomic analyses with closely related species identified both shared ancestral features and genus-specific characteristics. Phylogenetic reconstruction based on these chloroplast genomes provides robust support for interspecific relationships within *Lithocarpus*. These findings significantly expand the genomic resources for Fagaceae systematics while offering new insights into the evolutionary patterns and diversification mechanisms within this ecologically important genus.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Table S1: Shows the accession numbers and species information of the chloroplast genomes that were used in the phylogenetic analysis.

Author Contributions: Conceptualization, J.Y. and S.S.; methodology, Z.Z. and M.H.; software, X.L.; data curation, L.L.; writing—original draft preparation, S.L. and S.S.; visualization, K.F.; supervision, H.Y. and J.Y; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Investigation of forest germplasm resources in Guangdong Province, China, Funding number GDZZDC20228703.

Institutional Review Board Statement: Not applicable.

 doi:10.20944/preprints202504.2288.v1

Informed Consent Statement: Not applicable.

Data Availability Statement: The complete chloroplast genomes generated during the current study were deposited in GenBank with the accession numbers PV191269, OR805597.1, and PP234611~PP233615. The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. POWO. Plants of the world online. Facilitated by the Royal Botanic Gardens, Kew. 2024. Aug 2024. Available online: (accessed on
2. Webb, C.; Cannon, C.; Davies, S. Ecological organization, biogeography, and the phylogenetic structure of tropical forest tree communities. **2008**.
3. Petit, R.J.; Carlson, J.; Curtu, A.L.; Loustau, M.L.; Plomion, C.; González-Rodríguez, A.; Sork, V.; Ducousso, A. Fagaceae trees as models to integrate ecology, evolution and genomics. *New Phytol* **2013**, *197*, 369-371, doi:10.1111/nph.12089.
4. XY; JX; Xia; XF; Zhao; LC. Cupules and fruits of Lithocarpus (Fagaceae) from the Miocene of Yunnan, southwestern China. *TAXON* **2015**, *2015,64(4)*, 795-808.
5. Cannon, C.H.; Manos, P.S. Phylogeography of the Southeast Asian stone oaks (Lithocarpus). *J Biogeogr* **2003**, *30*, 211-226, doi:DOI 10.1046/j.1365-2699.2003.00829.x.
6. Liu, X.Y.; Song, H.Z.; Jin, J.H. Diversity of Fagaceae on Hainan Island of South China During the Middle Eocene: Implications for Phytogeography and Paleoecology. *Front Ecol Evol* **2020**, *8*, doi:ARTN 25510.3389/fevo.2020.00255.
7. Wei, Z.; Baocheng, W.; Chunfeng, S.; Qixin, L. Resources and Exploitation of Lithocarpus(Fagaceae) in China. *Chinese Wild Plant Resources* **2016**.
8. Guo, H.; Fu, M.X.; Zhao, Y.X.; Li, H.; Li, H.B.; Wu, D.T.; Gan, R.Y. The Chemical, Structural, and Biological Properties of Crude Polysaccharides from Sweet Tea (Lithocarpus litseifolius (Hance) Chun) Based on Different Extraction Technologies. *Foods* **2021**, *10*, doi:10.3390/foods10081779.
9. Hou, S.Z.; Chen, S.X.; Huang, S.; Jiang, D.X.; Zhou, C.J.; Chen, C.Q.; Liang, Y.M.; Lai, X.P. The hypoglycemic activity of Lithocarpus polystachyus Rehd. leaves in the experimental hyperglycemic rats. *Journal of Ethnopharmacology* **2011**, *138*, 142-149.
10. Liu; Zhengbo; Sun; Yinshi; Li; Wei. Preparative isolation, quantification and antioxidant activity of dihydrochalcones from Sweet Tea (Lithocarpus polystachyus Rehd.). *Journal of Chromatography B Analytical Technologies in the Biomedical & Life Sciences* **2015**.
11. Yang, L.; Zhang, S.; Wu, C.; Jiang, X.; Deng, M. Plastome characterization and its phylogenetic implications on Lithocarpus (Fagaceae). *BMC Plant Biol* **2024**, *24*, 1277, doi:10.1186/s12870-024-05874-z.
12. Zhou, B.F.; Yuan, S.; Crowl, A.A.; Liang, Y.Y.; Shi, Y.; Chen, X.Y.; An, Q.Q.; Kang, M.; Manos, P.S.; Wang, B. Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. *Nat Commun* **2022**, *13*, 1320, doi:10.1038/s41467-022-28917-1.
13. Yang, Y.; Zhu, J.; Feng, L.; Zhou, T.; Bai, G.; Yang, J.; Zhao, G. Plastid Genome Comparative and Phylogenetic Analyses of the Key Genera in Fagaceae: Highlighting the Effect of Codon Composition Bias in Phylogenetic Inference. *Front Plant Sci* **2018**, *9*, 82, doi:10.3389/fpls.2018.00082.
14. Herliana, L.; Iryanto, S.; Fauzan, Y.; Robiansyah, I. Comparative analysis of Lithocarpus chloroplast genomes reveals candidate DNA barcoding loci. In Proceedings of the IOP Conference Series: Earth and Environmental Science, 2023; p. 012083.
15. Zhao, C.C.; Wang, Y.Y.; Chan, K.X.; Marchant, D.B.; Franks, P.J.; Randall, D.; Tee, E.E.; Chen, G.; Ramesh, S.; Phua, S.Y.; et al. Evolution of chloroplast retrograde signaling facilitates green plant adaptation to land. *P Natl Acad Sci USA* **2019**, *116*, 5015-5020, doi:10.1073/pnas.1812092116.
16. Lu, R.; Hu, K.; Sun, X.; Chen, M. Low-coverage whole genome sequencing of diverse Dioscorea bulbifera accessions for plastome resource development, polymorphic nuclear SSR identification, and phylogenetic analyses. *Front Plant Sci* **2024**, *15*, 1373297, doi:10.3389/fpls.2024.1373297.

17. Daniell, H.; Jin, S.; Zhu, X.G.; Gitzendanner, M.A.; Soltis, D.E.; Soltis, P.S. Green giant-a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny. *Plant Biotechnol J* **2021**, *19*, 430-447, doi:10.1111/pbi.13556.
18. Gitzendanner, M.A.; §, P.S.S.; Yi, T.S.; Li, D.Z.; §, D.E.S. Plastome Phylogenetics: 30 Years of Inferences Into Plant Evolution - ScienceDirect. *Advances in Botanical Research* **2018**, *85*, 293-313.
19. Yang, Y.Y.; Qu, X.J.; Zhang, R.; Stull, G.W.; Yi, T.S. Plastid phylogenomic analyses of Fagales reveal signatures of conflict and ancient chloroplast capture. *Mol Phylogenet Evol* **2021**, *163*, 107232, doi:10.1016/j.ympev.2021.107232.
20. Chen, S.; Chen, R.; Zeng, X.; Chen, X.; Qin, X.; Zhang, Z.; Sun, Y. Genetic Diversity, Population Structure, and Conservation Units of Castanopsis sclerophylla (Fagaceae). *Forests (19994907)* **2022**, *13*.
21. Liu, S.Y.; Yang, Y.Y.; Tian, Q.; Yang, Z.Y.; Li, S.F.; Valdes, P.J.; Farnsworth, A.; Kates, H.R.; Siniscalchi, C.M.; Guralnick, R.P.; et al. An integrative framework reveals widespread gene flow during the early radiation of oaks and relatives in Quercoideae (Fagaceae). *J Integr Plant Biol* **2024**, doi:10.1111/jipb.13773.
22. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884-i890, doi:10.1093/bioinformatics/bty560.
23. Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; dePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* **2020**, *21*, 241, doi:10.1186/s13059-020-02154-5.
24. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* **2017**, *45*, W6-W11, doi:10.1093/nar/gkx391.
25. Shi, L.C.; Chen, H.M.; Jiang, M.; Wang, L.Q.; Wu, X.; Huang, L.F.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res* **2019**, *47*, W65-W73, doi:10.1093/nar/gkz345.
26. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **2004**, *32*, W273-W279, doi:10.1093/nar/gkh458.
27. Librado, P.; Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451-1452, doi:10.1093/bioinformatics/btp187.
28. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **2018**, *35*, 1547-1549, doi:10.1093/molbev/msy096.
29. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030-3031, doi:10.1093/bioinformatics/bty220.
30. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* **2001**, *29*, 4633-4642, doi:10.1093/nar/29.22.4633.
31. Sharp, P.M.; Li, W.H. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **1987**, *15*, 1281-1295, doi:10.1093/nar/15.3.1281.
32. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **2017**, *14*, 587-589, doi:10.1038/nmeth.4285.
33. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **2015**, *32*, 268-274, doi:10.1093/molbev/msu300.
34. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology & Evolution* **2011**, *28*, 2731.
35. Jovanovic, N.; Mikheyev, A.S. Interactive web-based visualization and sharing of phylogenetic trees using phylogeny. IO. *Nucleic Acids Res* **2019**, *47*, W266-W269.
36. Yang, Y.; Zhou, T.; Duan, D.; Yang, J.; Feng, L.; Zhao, G. Comparative Analysis of the Complete Chloroplast Genomes of Five Quercus Species. *Front Plant Sci* **2016**, *7*, 959, doi:10.3389/fpls.2016.00959.
37. Shelke, R.G.; Banerjee, R.P.; Joshi, B.; Singh, P.P.; Tiwari, G.J.; Adhikari, D.; Jena, S.N.; Barik, S.K. Chloroplast Genome of Lithocarpus dealbatus (Hook.f. & Thomson ex Miq.) Rehder Establishes Monophyletic Origin of the Species and Reveals Mutational Hotspots with Taxon Delimitation Potential. *Life (Basel)* **2022**, *12*, doi:10.3390/life12060828.

 doi:10.20944/preprints202504.2288.v1

38. Ma, C.X.; Yan, H.F.; Ge, X.J. The complete chloroplast genome of Lithocarpus hancei (Benth.) Rehd (Fagaceae) from Zhejiang, China. *Mitochondrial DNA B Resour* **2021**, *6*, 2022-2023, doi:10.1080/23802359.2021.1935357.
39. Liang, D.; Wang, H.; Zhang, J.; Zhao, Y.; Wu, F. Complete Chloroplast Genome Sequence of Fagus longipetiolata Seemen (Fagaceae): Genome Structure, Adaptive Evolution, and Phylogenetic Relationships. *Life (Basel)* **2022**, *12*, doi:10.3390/life12010092.
40. Yu, R.; Huang, J.; Xu, Y.; Ding, Y.; Zang, R. Plant Functional Niches in Forests Across Four Climatic Zones: Exploring the Periodic Table of Niches Based on Plant Functional Traits. *Front Plant Sci* **2020**, *11*, 841, doi:10.3389/fpls.2020.00841.
41. Seol, M.-A.; Kim, S.S.; Lee, E.S.; Nam, K.-H.; Chun, S.-J.; Lee, J.-W. Yeast cadmium factor 1-enhanced oxidative stress tolerance in Arabidopsis thaliana. *Journal of Plant Biotechnology* **2024**, *51*, 337-343.
42. Dong, W.P.; Xu, C.; Li, C.H.; Sun, J.H.; Zuo, Y.J.; Shi, S.; Cheng, T.; Guo, J.J.; Zhou, S.L. Ycf1, the most promising plastid DNA barcode of land plants. *Sci Rep-Uk* **2015**, *5*, doi:ARTN 834810.1038/srep08348.
43. Wang, L.L.; Li, Y.; Zheng, S.S.; Kozlowski, G.; Xu, J.; Song, Y.G. Complete Chloroplast Genomes of Four Oaks from the Section Cyclobalanopsis Improve the Phylogenetic Analysis and Understanding of Evolutionary Processes in the Genus Quercus. *Genes (Basel)* **2024**, *15*, doi:10.3390/genes15020230.
44. Hubert, F.O.; Grimm, G.W.; Jousselin, E.; Berry, V.; Franc, A.; Kremer, A. Multiple nuclear genes stabilize the phylogenetic backbone of the genus Quercus. *Systematics & Biodiversity* **2014**, *12*, 405-423.
45. Oh, S.H.; Manos, P.S. Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences. *Taxon* **2008**, *57*.
46. Grimm, D.G.W. The oaks of western Eurasia: Traditional classifications and evidence from two nuclear markers. *Taxon* **2010**, *59*, -.