

## Article

# Complete Chloroplast Genome of an Endangered Species *Quercus litseoides*, and Its Comparative, Evolutionary, and Phylogenetic Study with Other *Quercus* Section *Cyclobalanopsis* Species

Yu Li <sup>1,2,†</sup>, Tian-Rui Wang <sup>2,†</sup>, Gregor Kozłowski <sup>2,3,4</sup>, Mei-Hua Liu <sup>1</sup>, Li-Ta Yi <sup>1,\*</sup> and Yi-Gang Song <sup>1,2,\*</sup><sup>1</sup> College of Forestry and Biotechnology, Zhejiang A&F University, Lin'an, Hangzhou 311300, China;

liyulu980625@163.com (Y.L.); mhliu@zafu.edu.cn (M.-H.L.)

<sup>2</sup> Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China; wtianrui@163.com (T.-R.W.); gregor.kozłowski@unifr.ch (G.K.)<sup>3</sup> Department of Biology and Botanic Garden, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland<sup>4</sup> Natural History Museum Fribourg, Chemin du Musée 6, 1700 Fribourg, Switzerland

\* Correspondence: yilita@zafu.edu.cn (L.-T.Y.); ygsong@cemps.ac.cn (Y.-G.S.)

† These authors contributed equally to this work.

**Abstract:** *Quercus litseoides*, an endangered montane cloud forest species, is endemic to southern China. To understand the genomic features, phylogenetic relationships, and molecular evolution of *Q. litseoides*, the complete chloroplast (cp) genome was analyzed and compared in *Quercus* section *Cyclobalanopsis*. The cp genome of *Q. litseoides* was 160,782 bp in length, with an overall guanine and cytosine (GC) content of 36.9%. It contained 131 genes, including 86 protein-coding genes, eight ribosomal RNA genes, and 37 transfer RNA genes. A total of 165 simple sequence repeats (SSRs) and 48 long sequence repeats with A/T bias were identified in the *Q. litseoides* cp genome, which were mainly distributed in the large single copy region (LSC) and intergenic spacer regions. The *Q. litseoides* cp genome was similar in size, gene composition, and linearity of the structural region to those of *Quercus* species. The non-coding regions were more divergent than the coding regions, and the LSC region and small single copy region (SSC) were more divergent than the inverted repeat regions (IRs). Among the 13 divergent regions, 11 were in the LSC region, and only two were in the SSC region. Moreover, the coding sequence (CDS) of the six protein-coding genes (*rps12*, *matK*, *atpF*, *rpoC2*, *rpoC1*, and *ndhK*) were subjected to positive selection pressure when pairwise comparison of 16 species of *Quercus* section *Cyclobalanopsis*. A close relationship between *Q. litseoides* and *Quercus edithiae* was found in the phylogenetic analysis of cp genomes. Our study provided highly effective molecular markers for subsequent phylogenetic analysis, species identification, and biogeographic analysis of *Quercus*.

**Keywords:** Fagaceae; montane cloud forests; *Cyclobalanopsis litseoides*; repeat sequences; plastome

**Citation:** Li, Y.; Wang, T.-R.; Kozłowski, G.; Liu, M.-H.; Yi, L.-T.; Song, Y.-G. Complete Chloroplast Genome of an Endangered Species *Quercus litseoides*, and Its Comparative, Evolutionary, and Phylogenetic Study with Other *Quercus* Section *Cyclobalanopsis* Species. *Genes* **2022**, *13*, 1184. <https://doi.org/10.3390/genes13071184>

Academic Editor: Zhiqiang Wu

Received: 29 May 2022

Accepted: 29 June 2022

Published: 1 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Trees provide habitat for half the world's known terrestrial plant and animal species and are highly significant components of biodiversity and carbon storage in many ecosystems [1]. Recently, the vital importance of trees has received increasing attention ecologically, culturally, and economically [2,3]. Globally, more than 30% of tree species were classified as threatened in the State of the World's Trees report [1].

*Quercus* (oaks) is the largest genus (ca. 430 species) in Fagaceae and is one of the most important, species-rich, and entirely woody eudicot families [2]. There are two subgenera (*Quercus* and *Cerris*) with eight sections in the genus [4]. The taxonomy, in which section *Cyclobalanopsis* is nested in *Quercus*, has been largely investigated for their phylogenetic

relationships [5–7]. It has been treated as a separate genus in Flora of China [8]. *Quercus* is predominantly found in the temperate and subtropical forest ecosystems of the Northern Hemisphere. According to the IUCN's method for calculating threatened proportions incorporating data deficient species, 31% of oaks are threatened with extinction [9–11].

*Quercus litseoides* Dunn, an evergreen vulnerable tree/shrub species, forms small fragments of pure forests or sparse forests near the top of mountains with elevations between 700 and 1000 m [8]. It belongs to the *Quercus* section *Cyclobalanopsis*, with only five known populations endemic to South Guangdong and Hong Kong, China [8,12]. Thus, this species showed an island distribution pattern that is known as “sky islands”. The limited populations of *Q. litseoides* are especially threatened by habitat destruction, soil erosion, and climate change [9,13]. Assessment of rough data has finally led to *Q. litseoides* being classified as vulnerable [9].

Morphologically, the obovate-oblongate to narrowly elliptic leaf blade of *Q. litseoides* resembles another sky island species, *Quercus arbutifolia*. Leaf epidermal features of *Q. litseoides* (uniseriate trichome with single-celled trichome base (STB)) showed that it systematically belongs to the STB group [14]. The molecular phylogenetic work on *Quercus* always lacks the *Q. litseoides* species, owing to its limited distribution and the difficulty in acquiring materials [5–7]. Thus, few studies focused on its phylogenetic and phylogenomic analysis, and genetic and structural diversity, and so on.

The chloroplast (cp) is an essential maternal hereditary organelle in green plant cells with an independent circular genome and plays a critical role in photosynthesis and carbon fixation [15–17]. Owing to its conserved genome structure, gene composition, and variation ratio, the cp genome is used for comparative genomics, species identification, plant evolutionary studies, tracking seed dispersal, and studying the structural diversity and evolution of organellar genomes [7,18–20].

Owing to the benefits of next-generation sequencing technologies, whole cp genome sequencing is affordable and more efficient than ever before. Fifty complete cp genomes of *Quercus* have been published in the National Center for Biotechnology Information (NCBI) database, including 14 from the *Quercus* section *Cyclobalanopsis*. Thus, we have opportunities to (1) elaborate on the typical structural characteristics of *Q. litseoides*, (2) examine abundant, simple sequence repeats (SSRs) and repeat structures in the whole cp genome of *Q. litseoides* to provide markers for phylogenetic and genetic studies, (3) identify evolutionary selection pressure of the coding sequence (CDS) of the 78 shared protein-coding genes (PCGs), and (4) explore the phylogenomic position of *Q. litseoides*.

## 2. Materials and Methods

### 2.1. Plant Material, DNA Extraction, and Sequencing

Fresh, healthy leaf samples of *Q. litseoides* were collected from the Wu-Tong Mountain in Shenzhen, Guangdong Province (113°17' E, 22°23' N; Alt. 944 m). The leaves were desiccated in silica gel and deposited at the herbarium of the Shanghai Chenshan Botanical Garden. Total genomic DNA was extracted and purified from leaf tissues using a modified cetyl trimethyl ammonium bromide (CTAB) protocol [21]. DNA was fragmented by ultrasonic mechanical interruption, and the DNA fragments were purified. A DNA library with an average insert size of 350 bp was constructed using the whole genome shotgun strategy, and the quantified DNA was then double-terminally sequenced based on the Illumina NovaSeq6000 (Illumina, San Diego, CA, USA) technology platform in accordance with the manufacturer's manual at Wuhan Benagen Technology Co., Ltd. (Wuhan, China) [22].

### 2.2. Genome Assembly and Annotations

Raw reads from the sample produced at least five Gb with 150 bp pair-end read lengths by base calling analysis [23]. Clean reads were obtained by filtering low-quality sequences (quality value of  $Q \leq 5$  and N bases  $> 5\%$ ) using SOAPnuke Toolkit v.1.3.0 [24]. Genome assembly was performed using SPAdes v.3.13.0, with default parameters [25]. Prediction of coding genes and non-coding RNA annotations (ribosomal RNA (rRNA) and transfer

RNA (tRNA)) was performed using CPGAVAS2 [26]. The fully annotated cp genome of the circular diagram was drawn using the online program Organellar GenomeDRAW (OGDRAW) [27].

### 2.3. Repeated Sequence Analysis

SSRs loci were identified using the online program MicroSatellite (MISA) [28]. The minimum number of SSRs was set to ten for mono-nucleotide; four for di- and tri-nucleotide; three for tetra-, penta-, and hexa-nucleotide SSR motifs. Composite microsatellites were identified by setting the minimum distance between the two SSRs to be less than 100 bp. Minisatellite sequence repeats (M) of at least 10 bp in length were identified using the program Tandem Repeats Finder (TRF) [29]. The alignment parameters for match, mismatch, and indels were set to be two, seven, and seven, respectively. The minimum alignment score and maximum period size were set to 80 and 500, respectively. Additionally, REPuter was applied to predict forward repeat sequences (F), reverse repeat sequences (R), complementary repeat sequences (C), and palindromic repeat sequences (P) using the following settings: minimum repeat sequence of 30 bp, Hamming distance of three, and sequence identity between the two repeats of more than 90% [30,31].

### 2.4. Genome Structure Comparisons and Sequence Divergence Analysis

The boundaries and gene rearrangement between the large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions among the 16 species (Section *Cyclobalanopsis*) were horizontally visualized using the online tool IRscope [32]. Using the cp genome of *Q. litseoides* as the reference sequence, the program mVISTA was used to identify interspecific variations across the complete cp genomes of 16 species of section *Cyclobalanopsis* in Shuffle-LAGAN mode [33,34]. The cp genome sequences of the *Quercus* section *Cyclobalanopsis* were aligned using MAFFT v.7.847 [35]. After manual adjustment using BioEdit v.7.2.5 [36], single nucleotide polymorphisms and indel sites were counted using DnaSP v.6.12.03 [37]. A sliding window analysis was further conducted to calculate hotspots of nucleotide variability (Pi) values between cp genomes following a window length of 600 base pairs and a step size of 200 base pairs [38].

### 2.5. Evolutionary Selection Pressure Analysis

To identify the evolutionary selection pressure in the cp genomes of the *Quercus* section *Cyclobalanopsis* [39], the CDS of the 78 shared PCGs of *Cyclobalanopsis* were extracted and aligned using MAFFT v.7.847. The synonymous substitution rate (Ks), nonsynonymous substitution rate (Ka), and Ka/Ks ratio ( $\omega$ ) were calculated using DnaSP v.6.12.03. This value made sense if Ks is not equal to zero. Based on Ka/Ks ratio, the evolutionary selection of CDS was classified as positive selection ( $Ka/Ks > 1$ ), neutral selection ( $Ka/Ks = 1$ ), or purifying selection ( $Ka/Ks < 1$ ).

### 2.6. Phylogenetic Analyses

Thirty-four complete cp genome sequences, comprising one new cp genome sequence and 33 cp genome sequences of five sections of *Quercus* species from the GenBank database, were used to reconstruct phylogenetic relationships. *Trigonobalanus doichangensis*, *Fagus engleriana*, and *Juglans mandshurica* were used as outgroup species. The GenBank accession numbers for each taxon used are shown in Table S1. Homblock v1.0 was used to first screen out homologous sequences of the whole cp genomes of these 37 species [40], and the online software Circoletto was then used to visualize the alignment sequence of *Q. litseoides* and the homologous sequences of all species [41]. Next, the phylogenetic trees were reconstructed using two methods: Maximum Likelihood (ML) and Bayesian Inference (BI), using IQtree v.1.6.12 and MrBayes v3.2.7 [42,43], respectively. The ML tree adopted TVM + F + R2 as the best nucleotide replacement model with 1000 bootstrap replicates. The BI tree was set as follows: Markov chain Monte Carlo simulations (MCMC) algorithm for 5,000,000 generations with four incrementally heated chains, starting from random

trees, and sampling one out of every 100 generations. The first 25% of trees were discarded as burn-in. The constructed phylogenetic trees were further edited and visualized using FigTree v.1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) (accessed on 29 May 2022).

### 3. Results

#### 3.1. Chloroplast Genome Assembly and Annotation of *Q. litseoides*

Five Gb clean reads were generated in total from the genomic DNA of *Q. litseoides* using the Illumina sequencing. The complete cp genome of *Q. litseoides* has a quadripartite structure comprising 160,782 bp, including an LSC region of 90,235 bp and an SSC region of 18,867 bp, which were separated by a pair of IR regions (IRa and IRb) of 25,840 bp (Table 1 and Figure 1). The overall guanine and cytosine (GC) content of the cp genome of *Q. litseoides* was 36.90%, and the corresponding values in the LSC, SSC, and IR regions were 34.74%, 31.13%, and 42.77%, respectively (Table 1).

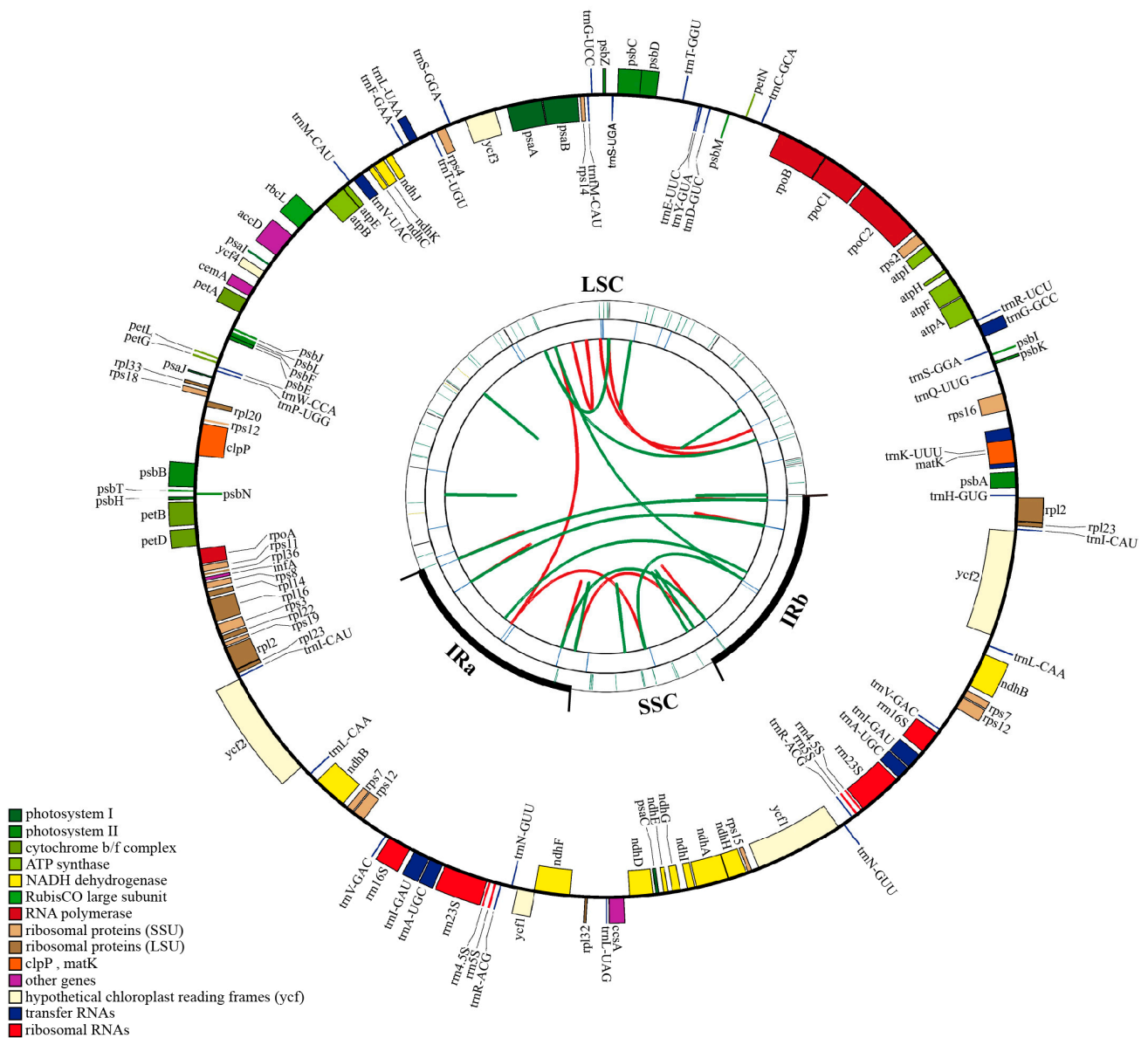
**Table 1.** Chloroplast genome structure and feature of *Q. litseoides*. Abbreviations: LSC (Large Single Copy), SSC (Small Single Copy), IR (Inverted Repeat), PCGs (protein-coding genes), tRNA (Transfer RNA genes), and rRNA (Ribosomal RNA genes).

Genome Feature		Length (bp)/Numbers	GC Content (%)
Structure length	Total	160,782	36.9
	LSC region	90,235	34.74
	SSC region	18,867	31.13
	IR (a/b) region	25,840	42.77
Gene numbers of different categories	Genes	131	39.5
	PCGs	86	37.88
	tRNA	37	53.2
	rRNA	8	55.49
Gene numbers of different regions	LSC region	61 (PCGs) and 22 (tRNA)	No information
	SSC region	11 (PCGs) and 1 (tRNA)	No information
	IR regions	14 (PCGs), 14 (tRNA) and 8 (rRNA)	No information

A total of 131 predicted genes in the cp genome of *Q. litseoides* were assigned to three groups based on their functions: 86 PCGs, 37 tRNA genes, and eight rRNA genes (Table 1). The GC content of all the genes was 39.5%, with 37.88% for PCGs, 53.2% for tRNA genes, and 55.49% for rRNA genes (Table 1). There were 83 genes (61 PCGs and 22 tRNA genes) in the LSC region, 12 genes (11 PCGs and one tRNA gene) in the SSC region, and 36 genes (seven PCGs, seven tRNA, and four rRNA genes duplicated) in the IR regions (Figure 1 and Table 1). In addition, the *rps12* and *ycf1* genes span two regions between IRs and LSC/SSC, respectively. Moreover, the *rps12* gene was recognized as a trans-spliced gene, with exons located in the IR and LSC regions (Figure 1).

Based on the cp genome annotation of *Q. litseoides*, 113 unique genes were divided into four functional categories with 18 groups. Among the 113 unique genes, there were 60 genes related to transcription and translation, 44 genes related to photosynthesis, five genes related to biosynthesis, and four genes whose functions were unknown. A total of 18 genes in the cp genome of *Q. litseoides* contained introns (12 PCGs and six tRNA genes), of which 15 genes (*trnK-UUU*, *trnG-GCC*, *trnL-UAA*, *trnV-UAC*, *trnI-GAU*, *trnA-UGC*, *rps16*, *rpl16*, *rpl2*, *rpoC1*, *atpF*, *ndhA*, *ndhB*, *petB*, and *petD*) contained one intron, whereas the other three genes (*ycf3*, *clpP*, and *rps12*) contained two introns (Table 2).





**Figure 1.** Gene map of the chloroplast genome of *Q. litseoides*. The chloroplast genome map has four circles. Outward from the center, the first circle shows forward and reverse repeats connected by red and green arcs, respectively. The second circle shows tandem repeats marked with a short bar. The third circle is the SSRs identified by MISA. The fourth circle is drawn with drawgenemap to display the gene structure on the chloroplast genome. The genes shown outside of the circle are transcribed clockwise, while those inside of the circle are transcribed counterclockwise. Genes with different functional groups are identified by different colors.

### 3.2. Repeat Sequences in the Chloroplast Genome of *Q. litseoides*

The SSRs, minisatellite sequences, dispersed repeat sequences, and palindromic repeat sequences were analyzed in the cp genome of *Q. litseoides*. A total of 165 SSRs were classified into five types (mono-, di-, tri-, tetra-, penta-nucleotide repeats). The first two types (mono- and di-nucleotide repeats) accounted for 87.28% of SSRs, and the proportions of the other three types were 4.25%, 6.67%, and 1.82%, respectively. It is worth noting that 42 composite microsatellites were identified because the minimum distance between the two SSRs was less than 100 bp. Most SSRs were composed of A and T, indicating a strong A/T bias. At the same time, the distribution of SSRs in the LSC region (66.7%) was higher than that in

the IR (20.6%) and SSC regions (12.7%). More than 64.8% of the SSRs were located in the intergenic spacer regions (IGS) and 35.2% in the gene regions (Table 3).

**Table 2.** Genetic classification of the chloroplast genome of *Q. litseoides*. Genes marked with the \* or \*\* sign are the gene with single or double introns, respectively. The duplicated genes located in IR regions were marked as (×2).

Category	Group	Name
Transcription and translation	Translational initiation factor	<i>infA</i>
	Ribosomal RNAs	<i>rrn16S</i> (×2), <i>rrn4.5S</i> (×2), <i>rrn23S</i> (×2), <i>rrn5S</i> (×2) <i>trnR-UCU</i> , <i>trnM-CAU</i> , <i>trnD-GUC</i> , <i>trnH-GUG</i> , <i>trnM-CAU</i> , <i>trnE-UUC</i> , <i>trnS-GCU</i> , <i>trnF-GAA</i> , <i>trnP-UGG</i> , <i>trnT-UGU</i> , <i>trnG-UCC</i> , <i>trnQ-UUG</i> , <i>trnY-GUA</i> , <i>trnW-CCA</i> , <i>trnS-UGA</i> , <i>trnC-GCA</i> , <i>trnT-GGU</i> , <i>trnL-UAG</i> , <i>trnS-GGA</i> , <i>trnK-UUU</i> *, <i>trnV-UAC</i> *, <i>trnL-UAA</i> *, <i>trnG-GCC</i> *, <i>trnA-UGC</i> *(×2), <i>trnI-GAU</i> *(×2), <i>trnL-CAA</i> (×2), <i>trnI-CAU</i> (×2), <i>trnN-GUU</i> (×2), <i>trnV-GAC</i> (×2), <i>trnR-ACG</i> (×2)
	Transfer RNAs	
	Small subunit of ribosome (SSU)	<i>rps2</i> , <i>rps11</i> , <i>rps19</i> , <i>rps14</i> , <i>rps4</i> , <i>rps15</i> , <i>rps16</i> *, <i>rps8</i> , <i>rps18</i> , <i>rps3</i> , <i>rps12</i> ** (×2), <i>rps7</i> (×2)
	Large subunit of ribosome (LSU)	<i>rpl14</i> , <i>rpl20</i> , <i>rpl36</i> , <i>rpl33</i> , <i>rpl16</i> *, <i>rpl32</i> , <i>rpl22</i> , <i>rpl2</i> *(×2), <i>rpl23</i> (×2)
	DNA-dependent RNA polymerase	<i>rpoC2</i> , <i>rpoB</i> , <i>rpoC1</i> *, <i>rpoA</i>
	Photosynthesis	Photosystem I
Photosystem II		<i>psbA</i> , <i>psbC</i> , <i>psbH</i> , <i>psbZ</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbF</i> , <i>psbD</i> , <i>psbT</i> , <i>psbN</i> , <i>psbL</i> , <i>psbM</i> , <i>psbE</i> , <i>psbB</i>
Subunit of cytochrome		<i>petB</i> *, <i>petN</i> , <i>petL</i> , <i>petG</i> , <i>petD</i> *, <i>petA</i>
ATP synthase		<i>atpA</i> , <i>atpI</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> *, <i>atpH</i>
RubisCO large subunit		<i>rbcL</i>
NADH dehydrogenase		<i>ndhG</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhK</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhF</i> , <i>ndhA</i> *, <i>ndhJ</i> , <i>ndhC</i> , <i>ndhB</i> *(×2)
Biosynthesis	Maturase	<i>matK</i>
	ATP-dependent Protease	<i>clpP</i> **
	Acetyl-CoA-carboxylase	<i>accD</i>
	Envelop membrane protein	<i>cemA</i>
	C-Type cytochrome synthesis	<i>ccsA</i>
Unknown	Hypothetical chloroplast reading frames( <i>ycf</i> )	<i>ycf4</i> , <i>ycf3</i> **, <i>ycf1</i> (×2), <i>ycf2</i> (×2)

**Table 3.** Simple sequence repeats (SSRs) number in the chloroplast genome of *Q. litseoides*. Abbreviations: LSC (Large Single Copy), SSC (Small Single Copy), IRs (Inverted Repeats), IGS (Intergenic Spacer Regions), GR (Gene Regions).

Repeat Type	Repeat Unit	Number (Proportion) of SSRs	Region			Location	
			LSC	SSC	IRs	IGS	GR
Mononucleotides	A/T	77 (46.67%)	60	11	6	59	18
	C/G	5 (3.03%)	5	0	0	3	2
Dinucleotides	AG/CT	19 (11.52%)	2	1	16	5	14
	AT/AT	43 (26.06%)	29	4	10	28	15
Trinucleotides	AAG/CTT	1 (0.61%)	0	1	0	0	1
	AAT/ATT	6 (3.64%)	4	2	0	3	3
Tetranucleotides	AAAT/ATTT	8 (4.85%)	7	1	0	5	3
	AATG/ATTC	1 (0.61%)	1	0	0	1	0
	AATT/AATT	2 (1.21%)	2	0	0	1	1
Pentanucleotides	AAAAT/ATTTT	1 (0.61%)	0	1	0	0	1
	AATGC/ATTGC	2 (1.21%)	0	0	2	2	0
Total		165 (100%)	110 (66.7%)	21 (12.7%)	34 (20.6%)	107 (64.8%)	58 (35.2%)

For the long repeat sequences, we detected seven M, 14 F, four R, two C, and 21 P. The length of the repeat units ranged from 19 to 56 bp (mainly between 30 and 40 bp),

and the repeat sequences had two to four repeats. Most of the long repeat sequences were distributed in the LSC region, especially all reverse and complementary repeat sequences. Six of the seven M were located in IR regions. Meanwhile, there were four sequences distributed in the regions between LSC and IRs, four between SSC and IRs, and seven between IRa and IRb. In contrast, 26 sequences belonged to the gene regions, while the others were located in the intergenic spacer regions. The long repeat sequences were mainly distributed in the following gene regions: *trnG-GCC*, *trnS-GGA*, *trnS-UGA*, *rpl2*, *rpl12*, *psaB*, *ndhA*, *clpP*, *ycf1*, *ycf2*, and *ycf3* (Tables 4, S2 and S3).

**Table 4.** The long repeat sequences, including minisatellite sequences (M), forward repeat sequences (F), reverse repeat sequences (R), complementary repeat sequences (C), and palindromic repeat sequences (P), in the cp genome of *Q. litseoides*.

No.	Repeat Type	Repeat Length (bp)	Region	Location	No.	Repeat Type	Repeat Length (bp)	Region	Location
1	M	19	LSC	IGS ( <i>trnF-GAA</i> , <i>ndhI</i> )	25	R	33	LSC, LSC	<i>clpP</i>
2	M	20	IRa	<i>rpl12</i>	26	C	34	LSC, LSC	IGS ( <i>rps16</i> , <i>trnQ-UUG</i> )
3	M	21	IRa	<i>ycf2</i>	27	C	30	LSC, LSC	IGS ( <i>petA</i> , <i>psbI</i> )
4	M	31	IRa	IGS ( <i>rrn4.5S</i> , <i>rrn5S</i> )	28	P	56	SSC, SSC	IGS ( <i>ndhD</i> , <i>psaC</i> )
5	M	31	IRb	IGS ( <i>rrn5S</i> , <i>rrn4.5S</i> )	29	P	44	LSC, LSC	IGS ( <i>psbT</i> , <i>psbN</i> )
6	M	21	IRb	<i>ycf2</i>	30	P	40	IRa, IRb	<i>rpl12</i>
7	M	20	IRb	<i>rpl12</i>	31	P	40	IRa, IRb	<i>rpl12</i>
8	F	40	IRa, IRa	<i>rpl2</i>	32	P	38	LSC, LSC	IGS ( <i>atpF</i> , <i>atpH</i> )
9	F	40	IRb, IRb	<i>rpl2</i>	33	P	34	SSC, SSC	<i>ycf1</i>
10	F	39	LSC, IRa	<i>ycf3</i>	34	P	39	LSC, IRb	<i>ycf3</i>
11	F	40	IRa, SSC	IGS ( <i>rps12</i> , <i>trnV-GAC</i> )	35	P	40	SSC, IRb	<i>ndhA</i>
12	F	30	IRa, IRa	IGS ( <i>rrn4.5S</i> , <i>rrn5S</i> )	36	P	39	LSC, LSC	IGS ( <i>trnT-GGU</i> , <i>psbD</i> )
13	F	30	IRb, IRb	IGS ( <i>rrn5S</i> , <i>rrn4.5S</i> )	37	P	30	LSC, LSC	<i>trnS-GGA</i>
14	F	30	LSC, LSC	<i>psaB</i>	38	P	30	IRa, IRb	IGS ( <i>rrn4.5S</i> , <i>rrn5S</i> )
15	F	30	LSC, IRa	<i>ycf3</i>	39	P	30	IRa, IRb	IGS ( <i>rrn4.5S</i> , <i>rrn5S</i> )
16	F	30	IRa, SSC	IGS ( <i>rps12</i> , <i>trnV-GAC</i> )	40	P	32	LSC, LSC	IGS ( <i>trnH-GUG</i> , <i>psbA</i> )
17	F	30	IRa, IRb	<i>ycf1</i>	41	P	30	LSC, IRb	<i>ycf3</i>
18	F	32	IRa, IRa	<i>ycf2</i>	42	P	30	IRa, IRa	<i>ycf1</i>
19	F	32	IRb, IRb	<i>ycf2</i>	43	P	30	SSC, IRb	<i>ndhA</i>
20	F	30	LSC, LSC	<i>trnS-GGA</i>	44	P	30	IRb, IRb	<i>ycf1</i>
21	F	30	LSC, LSC	<i>trnG-GCC</i>	45	P	32	LSC, LSC	IGS ( <i>rbcL</i> , <i>accD</i> )
22	R	31	LSC, LSC	IGS ( <i>trnR-UCU</i> , <i>atpA</i> )	46	P	32	IRa, IRb	<i>ycf2</i>
23	R	31	LSC, LSC	<i>clpP</i>	47	P	32	IRa, IRb	<i>ycf2</i>
24	R	31	LSC, LSC	IGS ( <i>atpA</i> , <i>atpF</i> )	48	P	30	LSC, LSC	<i>trnS-UGA</i>

### 3.3. Genome Structure Comparisons and Sequence Divergence of *Quercus* Section *Cyclobalanopsis*

The length of the cp genome in section *Cyclobalanopsis* changed little, with 445 bp, ranging from 160,533 bp (*Quercus acuta*) to 160,978 bp (*Quercus edithiae*). All cp genomes in

section *Cyclobalanopsis* were shorter than those in the other sections of *Quercus*. The length of the IR regions among different species in this section, *Cyclobalanopsis*, varied by 31 bp (Table S1).

The junction region between LSC and IRb (JLB) lies in the IGS between *rps19* and *rpl2* genes, and the junction region between LSC and IRa (JLA) is located between the *rpl2* and *trnH* genes. Most of the section *Cyclobalanopsis* species had 11 bp shifted away from the boundary for *rps19* gene in JLB, except *Q. edithiae* and *Quercus multinervis*, which had one bp shift. The *trnH* gene has 14 to 16 bp shifted from JLA. The *ycf1* gene crossed the junction regions between IRa/IRb and SSC (located in JSA and JSB), with the exception of the JSB of *Quercus chungii*, *Quercus acuta*, *Quercus saravanensis*, *Quercus schottkyana*, and *Quercus multinervis*. The *ycf1* gene (located in JSA) has 1045–1070 bp in the IRa region and 4612–4628 bp in the SSC region. However, the *ycf1* gene (located in JSB) has 1045–1060 bp in the IRb region and only 58–68 bp in the SSC region. The *ndhF* gene, located in the SSC region just beside the JSB, has a one or 11 bp shift in the species without *ycf1* gene (Figure 2).

To further investigate the divergence of cp genomes among related species, the evolution of cp genomes was explored in the *Quercus* section *Cyclobalanopsis* using the annotated cp genome of *Q. litseoides*. No structural rearrangement with high sequence similarity occurred in this section. However, the non-coding regions were more divergent than the coding regions, and the LSC and SSC regions were more divergent than the IR regions (Figure S1). Furthermore, the *ycf1*, *ndhF*, *rpl32*, and *psbD* genes in the gene regions and *petN—psbM*, *trnK-UUUU—rps16*, *rps16—trnQ-UUG*, *psbM—trnD-GUC*, *psbZ—trnG-UCC*, *trnT-GGU—psbD*, *rbcL—accD*, and *rpl32—trnL-UAG* in the intergenic spacer regions were quite mutable (Figure S1).

A total of 482 variation polymorphism sites, including 335 single nucleotide polymorphism sites, 147 parsimony-informative sites, and 200 indel events, were detected by nucleotide polymorphism analysis. Nucleotide polymorphism in the IR regions of the cp genome was significantly lower than that in the LSC and SSC regions. The  $P_i$  value of nucleotide diversity in the cp genomes of section *Cyclobalanopsis* ranged from 0 to 0.01808, with an average of 0.00059. Furthermore, the analysis detected 13 highly divergent regions ( $P_i > 0.002$ ), of which eight were located in the gene regions (*trnH-GUG*, *trnC-GCA*, *trnS-UGA*, *ycf1*, *ycf3*, *psaI*, *psbJ*, and *rpl22*) and five in the intergenic spacer regions (*trnK-UUUU—rps16*, *rps16—trnQ-UUG*, *psbM—trnD-GUC*, *rbcL—accD*, and *ndhF—rpl32*). Among the 13 divergent regions, 11 were in the LSC region, and only two were in the SSC region (Figure 3).

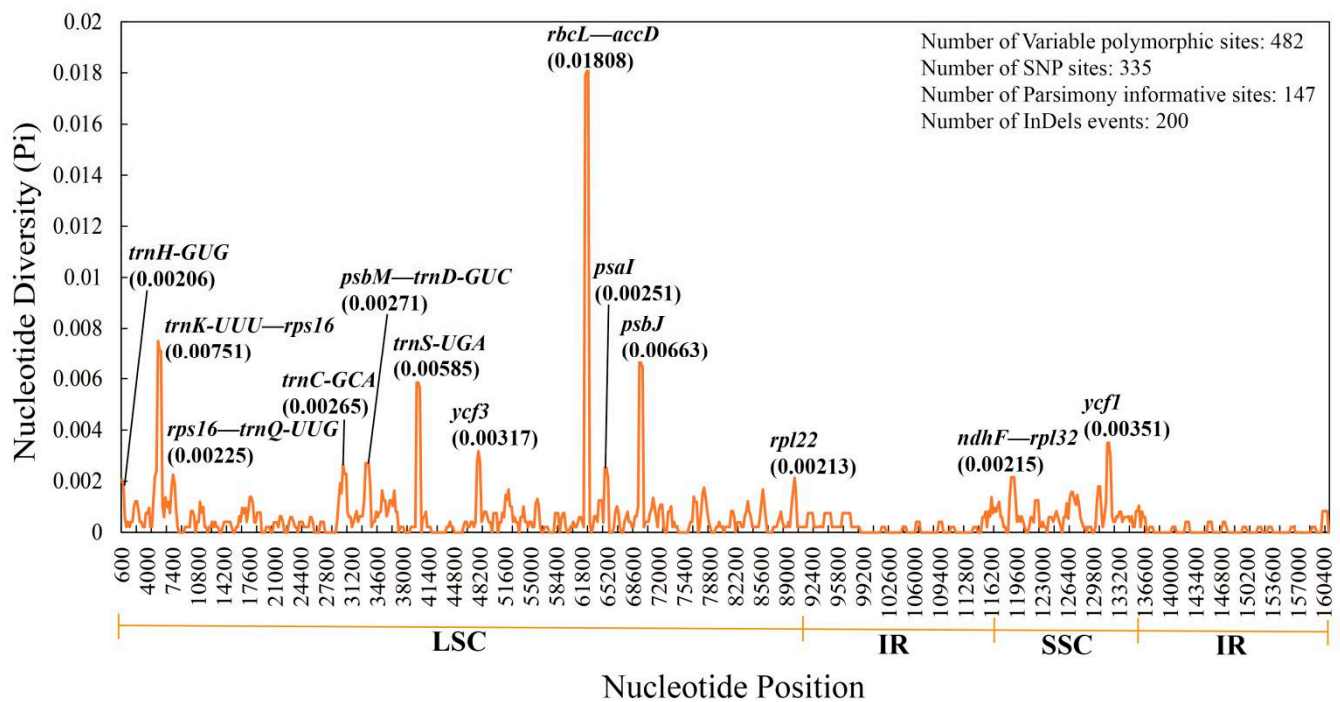
### 3.4. Selective Pressure Analysis

To investigate the evolutionary characteristics of *Quercus* section *Cyclobalanopsis* in the cp genome,  $K_a$ ,  $K_s$ , and the  $K_a/K_s$  ( $\omega$ ) ratio were calculated for the CDS of the 78 shared PCGs in the 16 cp genomes. The results showed that the  $K_a$  values ranged from 0 to 0.06019, the  $K_s$  values ranged from 0 to 0.08333, and the  $K_a/K_s$  ( $\omega$ ) values of 37 CDS of the PCGs were significant ( $K_s > 0$ ,  $p < 0.05$ , Figure 4, Table S4). The  $K_a/K_s$  ( $\omega$ ) values of six CDS of the PCGs (*rps12*, *matK*, *atpF*, *rpoC2*, *rpoC1*, and *ndhK*), which were distributed in the LSC region, were greater than 1, indicating that these genes had undergone positive selection. The  $K_a/K_s$  ( $\omega$ ) values of the other 31 CDS of the PCGs were all less than 1, suggesting that genes were under purifying selection (Figure 4, Table S4).





Figure 2. Comparison of the junction regions (JLA, JLB, JSB, JSA) of the chloroplast genomes of *Quercus* section *Cyclobalanopsis*.

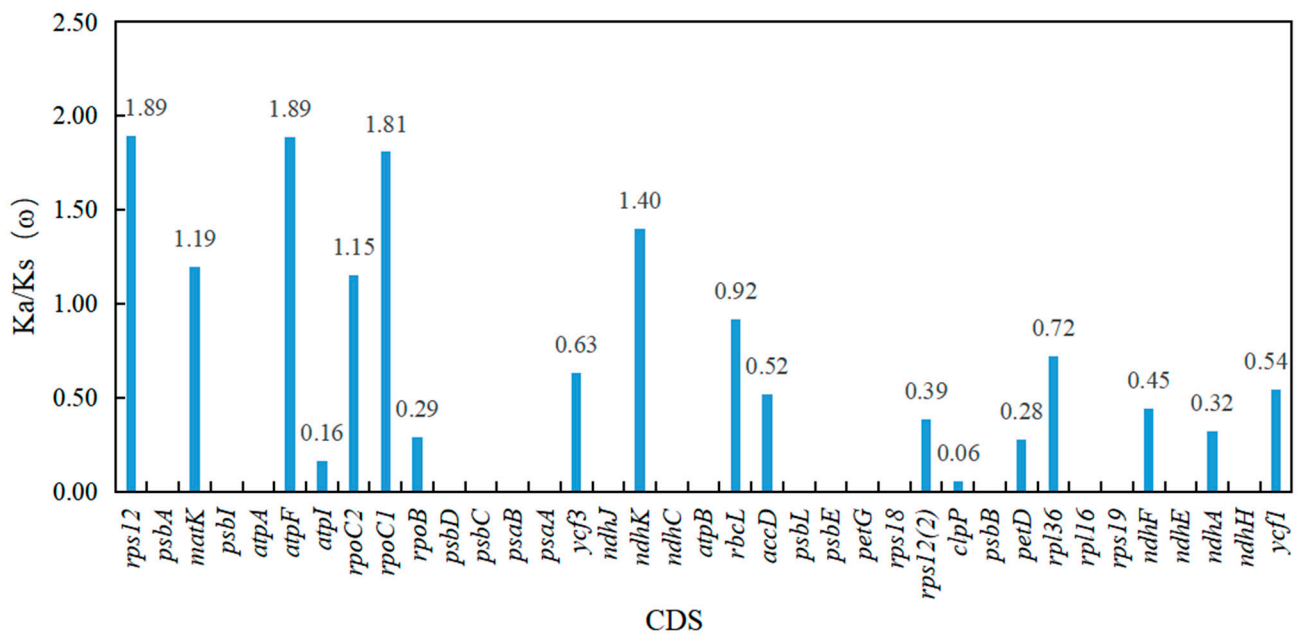


**Figure 3.** Sliding window analysis of 16 chloroplast genomes of *Quercus* section *Cyclobalanopsis*. The *x*-axis represents the site positions of the middle point of the window, and the *y*-axis represents the value of nucleotide diversity (Pi) per window.

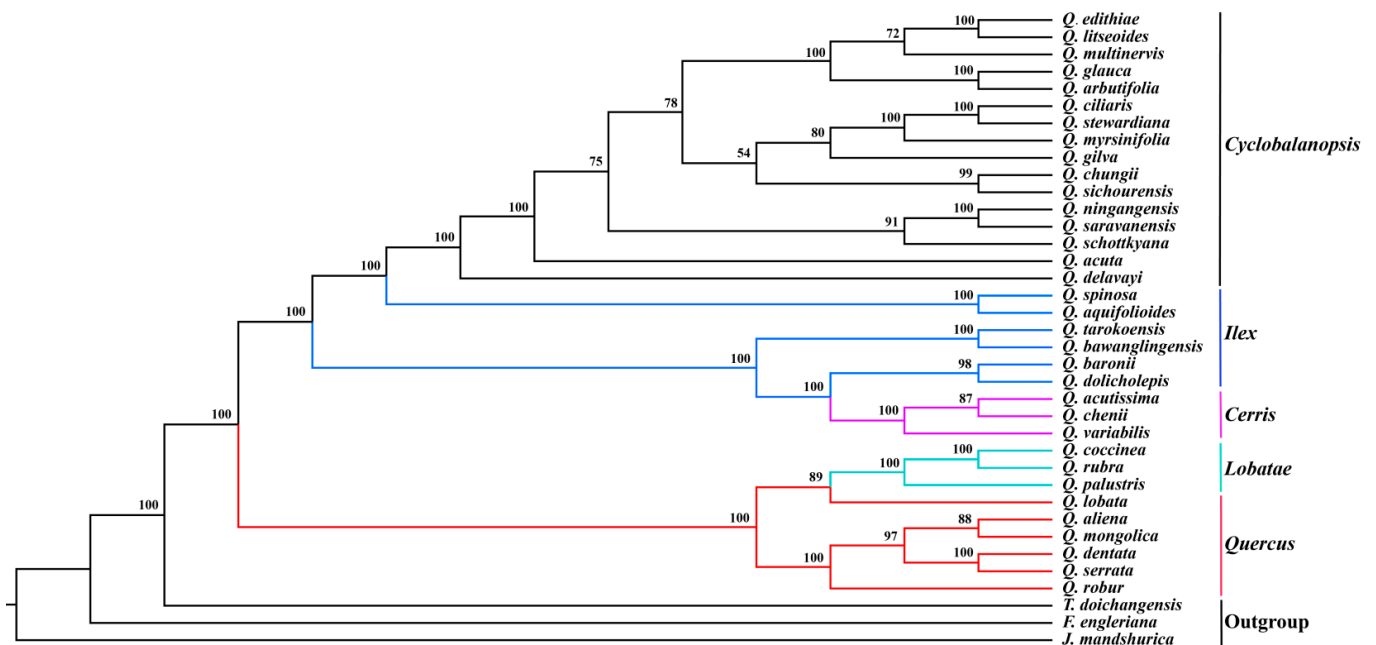
### 3.5. Phylogenetic Analyses

Homologous sequences were screened for the cp genomes of 34 *Quercus* species in five sections and three outgroups (*Trigonobalanus doichangensis*, *Fagus engleriana*, and *Juglans mandshurica*). Thirty-seven homologous sequences with a length of 88,044 bp were generated. The alignment results included almost all genes (Figure S2).

The ML tree and BI tree reconstructed by homologous sequences showed a similar topological structure. All nodes of the phylogenetic trees were supported by 54–100% bootstrap values in ML analysis and 0.67–1.00 Bayesian posterior probabilities in BI analysis (Figures 5 and S3). The results showed that the section *Quercus* and section *Lobatae* formed one clade. *Quercus* section *Ilex* split into two strongly supported clusters; one clade is the species distributed in the Tibetan area (*Quercus spinosa* and *Q. aquifolioides*), whereas the species from East and Central China together with *Quercus* section *Cerris* formed another clade. The bootstrap support in *Quercus* section *Cyclobalanopsis* is not so high in several nodes of the middle part. Up to now, *Quercus* section *Cyclobalanopsis* could be split into five clades. Firstly, *Q. delavayi* and *Q. acuta* diverged from the section *Cyclobalanopsis*, forming two separate clades. *Q. ningangensis* and *Q. saravanensis* were sister to the *Q. schottkyana*, constituting two clades along with other subtropical species. The montane cloud forest species (*Q. arbutifolia* and *Q. litseoides*), tropical species together with the widespread species (*Q. glauca* and *Q. multinervis*) formed the last clade (Figure 5).



**Figure 4.** The Ka/Ks ( $\omega$ ) values of 37 shared functional protein-coding genes of 16 chloroplast genomes of *Quercus* section *Cyclobalanopsis*.



**Figure 5.** The phylogenetic tree among 37 chloroplast genome homologous sequences is based on the ML method. Values besides the branch represented bootstrap support (BS). Abbreviations: *Quercus* (Q.), *Trigonobalanus* (T.), *Fagus* (F.), and *Juglans* (J.).

#### 4. Discussion

##### 4.1. Architecture of cp Genomes in *Quercus* Section *Cyclobalanopsis*

In this study, we present the complete cp genome of *Q. litseoides*. Combined with the 15 related species reported previously, we performed a comparative analysis of the genomic features of *Quercus* section *Cyclobalanopsis*. Angiosperm cp genomes have a highly conserved structure and gene content [16,44]. The cp genomes of *Quercus* are highly conserved in size, quadruple structure, and GC content. There were slight differences in the total numbers of genes and unique genes in *Quercus*. Most *Quercus* species have 86

total PCGs and 79 unique PCGs, except *Quercus rubra*, *Quercus fabri*, *Quercus acutissima*, and *Q. edithiae*, which have one extra gene *ycf15* [7,45–47]. Compared with other *Quercus* species, three tRNA genes (*trnP-GGG*, *trnT-GGU*, and *trnM-CAU*) are missing in *Q. litseoides* [7]. It has also been observed to be missing of *ycf15* and *trnP-GGG* in the cp genomes of some angiosperms [48,49]. The two copy of *trnT-GGU* and *trnM-CAU* genes in most *Quercus* species were caused by annotation and the overlap of these two genes. Most other angiosperms, such as *Musa*, *Oryza*, *Stauntonia*, and *Carya*, only recognized one copy of these two genes in the LSC region [50–53].

Repeat sequences are considered to play an important role in cp genome rearrangement and sequence differentiation [54,55]. A strong A/T bias, LSC concentration (66.7%), and IGS concentration (64.8%) for SSRs, similar to other angiosperm cp genomes, were detected in *Q. litseoides* [56–58]. The number and types of SSRs varied extensively when compared to those of other cp genomes in *Quercus*. The number of SSRs in *Q. litseoides* was higher than that in the other *Quercus* species, whereas fewer SSRs were distributed in the LSC and IGS regions [46,58–63]. These variations support the idea that SSRs can be used as lineage-specific markers for genetic diversity analysis and can be used as markers to understand evolutionary history [64]. The large variation in long repeats in closely related species may reflect a certain degree of evolutionary flexibility [65]. Forty-eight long repeats were detected, and several repeats occurred in the same genes (four in *ycf1*, four in *ycf3*, six in *ycf2*, and four in *rpl12*).

The variable boundary regions are believed to be the arriving force for the variation in the angiosperm cp genomes [66–69]. The IR regions are very important in stabilizing the structure of the cp genome and are mainly responsible for variations in the length of the plastome [70]. The genes distributed in the IR regions of *Q. litseoides* in this study were similar to those of most other species, with little difference in the distribution of boundary genes. There was no obvious expansion or contraction in the IR regions of *Q. litseoides*. For the *Quercus* section *Cyclobalanopsis*, shifts of less than 16 bp were detected in JLB and JLA. The *ycf1* gene crossed the boundary of the IR and SSC regions, which had more variation in length. As expected, the relatively trivial cp genome length and variations in the boundary demonstrated the conservation of *Quercus* section *Cyclobalanopsis* plastomes. Previous studies on land plants also found that the expansion and contraction of the IR regions caused the mutual transfer of genes between the SC and IR regions or the increase and decrease of genes [71].

As the IR regions contain conserved rRNA genes with lower variability, the LSC and SSC regions had a higher level of sequence diversity than the IR regions in the *Quercus* section *Cyclobalanopsis* similar to other cp genomes [20,58,72]. The existence of a copy-dependent repair mechanism in many types of plants leads to a low replacement rate in the IR regions [72]. These mutated regions are more prone to nucleotide substitution during evolution, providing efficient molecular markers for subsequent species identification and useful data and phylogenetic information for genetic evolution analysis. In total, seven regions with higher Pi values (>0.015) were detected in *Quercus* [7], whereas only one was detected in the section *Cyclobalanopsis* (Figure 3). Compared with that of the other sections of *Quercus*, section *Cyclobalanopsis* had the lowest nucleotide diversity.

Nucleotide substitution is the driving force behind genomic evolution. Non-synonymous substitutions may alter protein function such that natural selection tends to remove these deleterious mutations, causing most species to be under negative selection pressure [73,74]. Analysis of adaptive evolution contributes to a profound understanding of gene variation, changes in protein structure and function, and the evolutionary history of species [75]. In this study, six PCGs, *rps12*, *matK*, *atpF*, *rpoC2*, *rpoC1*, and *ndhK*, were positively selected, providing evidence of the adaptive evolution of proteins. Genes with different functions have evolved at different rates [76]. The adaptive evolution of *matK*, which is involved in biosynthesis, has been detected by positive selection in multiple angiosperms [77,78]. The *atpF* and *ndhK* genes are associated with photosynthesis, whereas *rps12*, *rpoC2*, and *rpoC1* genes are associated with transcription and translation. Genes of the genetic and photo-



synthetic systems play an important role in the adaptation of angiosperms to terrestrial ecological environments [50,79,80]. *Quercus litseoides* live at high altitudes and must adapt to high UV radiation intensity, hypoxia, low temperature, and drought stress. These genes may be an important genetic basis for evolutionary adaptation at the chloroplast level.

#### 4.2. Phylogeny of Chloroplast Genome of *Quercus*

The phylogenetic study of *Quercus* is a great challenge because of its species richness, wide distribution, and serious hybridization introgression [6,7,10]. Our cp genome phylogenetic tree supports that sections *Quercus* and *Lobatae* formed one lineage (belonging to subgenus *Quercus*), and sections *Cyclobalanopsis*, *Ilex*, and *Cerris* formed another lineage (subgenus *Cerris*) [6]. Based on the RAD-seq data, section *Ilex* is a monophyletic lineage, whereas the phylogenetic reconstruction of the cp genome was paraphyletic. The section *Cerris* formed one lineage and nested into section *Ilex* in this study, which is also different from the phylogeny based on RAD-seq data [6].

*Quercus litseoides* has the closest relationship with *Q. edithiae*, which has an overlapping geographic area. Based on the current samples, the cp genome phylogeny of section *Cyclobalanopsis* is very different from that based on RAD-seq data. When constructing phylogenetic trees based on RAD-seq data, the *Quercus* section *Cyclobalanopsis* was divided into two clades: one defined by compound trichome bases (CTB lineage) and one by single-celled trichome bases (STB lineage). However, the two lineages and relationships among species were not supported by the cp genome phylogeny in this study [5]. With the development of next-generation sequencing technologies, we could add more taxa and samples to explore and compare the phylogenomics of the cp and nuclear genomes of *Quercus* section *Cyclobalanopsis*.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13071184/s1>. Figure S1: Visualization of the aligned sequence of the 16 chloroplast genomes of *Quercus* section *Cyclobalanopsis* with *Q. litseoides* as a reference using mVISTA. The gray arrows above show the locations of the reference sequence genes, and the direction is forward or reverse. The position of the genome is shown on the horizontal axis at the bottom of each block. The alignment similarity percentages are shown on the right side of the graph (vertical axis). Different colors represent different regions: blue for exons, cyan for introns, and red for intergenic spacer regions. Figure S2: Visualization of homologous sequence alignment of chloroplast genomes. The final alignment sequence is on the right and the corresponding genes are on the left. The visualized map shows the reception and relative locations of these genes. Different colors in the figure represent the similarity between the final alignment sequence and the original sequence, that is, blue  $\leq 50\%$ , green  $\leq 75\%$ , orange  $\leq 99.999\%$ , and red  $> 99.999\%$ . Figure S3. The phylogenetic tree among 37 chloroplast genome homologous sequences is based on the BI method. Values besides the branch represented Bayesian posterior probabilities (PP). Abbreviations: *Quercus* (Q.), *Trigonobalanus* (T.), *Fagus* (F.), and *Juglans* (J.). Table S1: Information on the chloroplast genomes used in this study. Table S2: Minisatellite sequences in the chloroplast genome of *Q. litseoides*. Table S3: Forward repeat sequences (F), reverse repeat sequences (R), complementary repeat sequences (C), and palindromic repeat sequences (P) in the chloroplast genome of *Q. litseoides*. Table S4: Ka, Ks, and Ka/Ks ( $\omega$ ) values of 78 shared functional protein-coding genes in 16 chloroplast genomes of *Quercus* section *Cyclobalanopsis*.

**Author Contributions:** Conceptualization, Y.-G.S. and L.-T.Y.; methodology, Y.L. and T.-R.W.; software, Y.L. and T.-R.W.; validation, T.-R.W., L.-T.Y. and Y.-G.S.; formal analysis, Y.L.; investigation, T.-R.W.; resources, Y.-G.S.; data curation, Y.L. and T.-R.W.; writing—original draft preparation, Y.L. and T.-R.W.; writing—review and editing, Y.-G.S., G.K., M.-H.L., L.-T.Y. and T.-R.W.; visualization, Y.L., T.-R.W. and Y.-G.S.; supervision, Y.-G.S. and L.-T.Y.; project administration, Y.-G.S.; funding acquisition, Y.-G.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China, grant number 31901217, and the Special Fund for Scientific Research of Shanghai Landscaping & City Appearance Administrative Bureau, grant number G192422.



**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the finding of this study are openly available in the GenBank of NCBI at <https://www.ncbi.nlm.nih.gov> (accessed on 25 May 2022), reference number (ON598394).

**Acknowledgments:** We want to thank Jiang-Ping Shu and Yu-Feng Gu from The National Orchid Conservation Centre of China and The Orchid Conservation and Research Centre of Shenzhen, who helped us with the collection of material. We also want to thank Wuhan Benagen Technology Co., Ltd. for their help on sequencing.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. BGCI. *State of the World's Trees*; BGCI: Richmond, UK, 2021; pp. 3–6.
2. Fazan, L.; Song, Y.G.; Kozłowski, G. The Woody Planet: From Past Triumph to Manmade Decline. *Plants* **2020**, *9*, 1593. [CrossRef]
3. Watson, J.E.M.; Evans, T.; Venter, O.; Williams, B.; Tulloch, A.; Stewart, C.; Thompson, I.; Ray, J.C.; Murray, K.; Salazar, A.; et al. The exceptional value of intact forest ecosystems. *Nat. Ecol. Evol.* **2018**, *2*, 599–610. [CrossRef]
4. Denk, T.; Grimm, G.W.; Manos, P.S.; Deng, M.; Hipp, A.L. 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.*; Gil-Pelegrin, E., Peguero-Pina, J., Eds.; Tree Physiology: Springer, Cham, 2017; Volume 7, pp. 13–38. [CrossRef]
5. Deng, M.; Jiang, X.L.; Hipp, A.L.; Manos, P.S.; Hahn, M. 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.* **2018**, *119*, 170–181. [CrossRef] [PubMed]
6. Hipp, A.L.; Manos, P.S.; Hahn, M.; Avishai, M.; Bodenes, C.; Cavender-Bares, J.; Crawl, A.A.; Deng, M.; Denk, T.; Fitz-Gibbon, S.; et al. Genomic landscape of the global oak phylogeny. *New Phytol.* **2020**, *226*, 1198–1212. [CrossRef] [PubMed]
7. Yang, Y.; Zhou, T.; Qian, Z.; Zhao, G. Phylogenetic relationships in Chinese oaks (Fagaceae, *Quercus*): Evidence from plastid genome using low-coverage whole genome sequencing. *Genomics* **2021**, *113*, 1438–1447. [CrossRef] [PubMed]
8. Huang, C.C.; Chang, Y.T.; Bartholomew, B. Fagaceae. In *Flora of China, English Version*; Science Press and Missouri Botanical Garden Press: Beijing, China; St. Louis, MO, USA, 1999; Volume 4, pp. 380–400.
9. Carrero, C.; Jerome, D.; Beckman, E.; Byrne, A.; Coombes, A.J.; Deng, M.; González-Rodríguez, A.; Hoang, V.S.; Khoo, E.; Nguyen, N.; et al. *The Red List of Oaks 2020*; The Morton Arboretum: Lisle, IL, USA, 2020; p. 5.
10. Manos, P.S.; Doyle, J.J.; Nixon, K.C. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Mol. Phylogenet. Evol.* **1999**, *12*, 333–349. [CrossRef] [PubMed]
11. Nixon, K.C. *Quercus*. In *Flora of North America North of Mexico*; Editorial Committee, Ed.; Oxford University Press: New York, NY, USA, 1997; pp. 445–447.
12. CFH. 2022. Available online: <http://cfh.ac.cn/> (accessed on 29 May 2022).
13. Song, Y.G.; Petitpierre, B.; Deng, M.; Wu, J.P.; Kozłowski, G. Predicting climate change impacts on the threatened *Quercus arbutifolia* in montane cloud forests in southern China and Vietnam: Conservation implications. *For. Ecol. Manag.* **2019**, *444*, 269–279. [CrossRef]
14. Deng, M.; Hipp, A.; Song, Y.G.; Li, Q.S.; Coombes, A.; Cotton, A. Leaf epidermal features of *Quercus* subgenus *Cyclobalanopsis* (Fagaceae) and their systematic significance. *Bot. J. Linn. Soc.* **2014**, *176*, 224–259. [CrossRef]
15. Bobik, K.; Burch-Smith, T.M. Chloroplast signaling within, between and beyond cells. *Front. Plant Sci.* **2015**, *6*, 781. [CrossRef]
16. Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [CrossRef]
17. Howe, C.J.; Barbrook, A.C.; Koumandou, V.L.; Nisbet, R.E.; Symington, H.A.; Wightman, T.F. Evolution of the chloroplast genome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2003**, *358*, 99–107. [CrossRef] [PubMed]
18. Asaf, S.; Khan, A.L.; Aaqil Khan, M.; Muhammad Imran, Q.; Kang, S.M.; Al-Hosni, K.; Jeong, E.J.; Lee, K.E.; Lee, I.J. Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species. *PLoS ONE* **2017**, *12*, e0182281. [CrossRef] [PubMed]
19. Birky, C.W.; Maruyama, T.; Fuerst, P. An Approach to Population and Evolutionary Genetic Theory for Genes in Mitochondria and Chloroplasts, and Some Results. *Genetics* **1983**, *103*, 513–527. [CrossRef] [PubMed]
20. He, L.; Qian, J.; Li, X.; Sun, Z.; Xu, X.; Chen, S. Complete Chloroplast Genome of Medicinal Plant *Lonicera japonica*: Genome Rearrangement, Intron Gain and Loss, and Implications for Phylogenetic Studies. *Molecules* **2017**, *22*, 249. [CrossRef] [PubMed]
21. Doyle, J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
22. Batzoglou, S.; Berger, B.; Mesirov, J.; Lander, E.S. Sequencing a genome by walking with clone-end sequences: A mathematical analysis. *Genome Res.* **1999**, *9*, 1163–1174. [CrossRef]
23. Ewing, B.; Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **1998**, *8*, 186–194. [CrossRef]

24. Chen, Y.; Chen, Y.; Shi, C.; Huang, Z.; Zhang, Y.; Li, S.; Li, Y.; Ye, J.; Yu, C.; Li, Z.; et al. SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **2018**, *7*, gix120. [[CrossRef](#)]
25. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)]
26. Shi, L.; Chen, H.; Jiang, M.; Wang, L.; Wu, X.; Huang, L.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **2019**, *47*, W65–W73. [[CrossRef](#)]
27. Lohse, M.; Drechsel, O.; Kahlau, S.; Bock, R. OrganellarGenomeDRAW—A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **2013**, *41*, W575–W581. [[CrossRef](#)] [[PubMed](#)]
28. Beier, S.; Thiel, T.; Munch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]
29. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)]
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. [[CrossRef](#)]
31. Liang, C.; Wang, L.; Lei, J.; Duan, B.; Ma, W.; Xiao, S.; Qi, H.; Wang, Z.; Liu, Y.; Shen, X.; et al. A Comparative Analysis of the Chloroplast Genomes of Four *Salvia* Medicinal Plants. *Engineering* **2019**, *5*, 907–915. [[CrossRef](#)]
32. Amiryousefi, A.; Hyvonen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [[CrossRef](#)] [[PubMed](#)]
33. Brudno, M.; Malde, S.; Poliakov, A.; Do, C.B.; Couronne, O.; Dubchak, I.; Batzoglou, S. Glocal alignment: Finding rearrangements during alignment. *Bioinformatics* **2003**, *19* (Suppl. S1), i54–i62. [[CrossRef](#)]
34. 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. [[CrossRef](#)]
35. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
36. Tippmann, H.F. Analysis for free: Comparing programs for sequence analysis. *Brief Bioinform.* **2004**, *5*, 82–87. [[CrossRef](#)]
37. Rozas, J.; Ferrer-Mata, A.; Sanchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sanchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [[CrossRef](#)] [[PubMed](#)]
38. Gou, W.; Jia, S.B.; Price, M.; Guo, X.L.; Zhou, S.D.; He, X.J. Complete Plastid Genome Sequencing of Eight Species from *Hansenia*, *Haplospheera* and *Sinodielsia* (Apiaceae): Comparative Analyses and Phylogenetic Implications. *Plants* **2020**, *9*, 1523–1539. [[CrossRef](#)] [[PubMed](#)]
39. Hurst, L.D. The Ka/Ks ratio: Diagnosing the form of sequence evolution. *Trends Genet.* **2002**, *18*, 486. [[CrossRef](#)]
40. Bi, G.; Mao, Y.; Xing, Q.; Cao, M. HomBlocks: A multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block searching. *Genomics* **2018**, *110*, 18–22. [[CrossRef](#)]
41. Darzentas, N. Circoletto: Visualizing sequence similarity with Circos. *Bioinformatics* **2010**, *26*, 2620–2621. [[CrossRef](#)] [[PubMed](#)]
42. 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. [[CrossRef](#)]
43. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Hohna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)]
44. Palmer, J.D. Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* **1985**, *19*, 325–354. [[CrossRef](#)]
45. Alexander, L.W.; Woeste, K.E. Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. *Tree Genet. Genomes* **2014**, *10*, 803–812. [[CrossRef](#)]
46. Li, X.; Li, Y.; Zang, M.; Li, M.; Fang, Y. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Quercus acutissima*. *Int. J. Mol. Sci.* **2018**, *19*, 2443. [[CrossRef](#)]
47. Xu, Y.; Chen, H.; Qi, M.; Su, W.; Zhang, Y.; Du, F.K. The complete chloroplast genome of *Quercus fabri* (Fagaceae) from China. *Mitochondrial DNA Part B* **2019**, *4*, 2857–2858. [[CrossRef](#)] [[PubMed](#)]
48. Jansen, R.K.; Cai, Z.; Raubeson, L.A.; Daniell, H.; Depamphilis, C.W.; Leebens-Mack, J.; Muller, K.F.; Guisinger-Bellian, M.; Haberle, R.C.; Hansen, A.K.; et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19369–19374. [[CrossRef](#)] [[PubMed](#)]
49. Liaud, M.F.; Zhang, D.X.; Cerff, R. Differential intron loss and endosymbiotic transfer of chloroplast glyceraldehyde-3-phosphate dehydrogenase genes to the nucleus. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8918–8922. [[CrossRef](#)] [[PubMed](#)]
50. Gao, L.Z.; Liu, Y.L.; Zhang, D.; Li, W.; Gao, J.; Liu, Y.; Li, K.; Shi, C.; Zhao, Y.; Zhao, Y.J.; et al. Evolution of *Oryza* chloroplast genomes promoted adaptation to diverse ecological habitats. *Commun. Biol.* **2019**, *2*, 278. [[CrossRef](#)] [[PubMed](#)]
51. Shen, J.; Li, X.; Chen, X.; Huang, X.; Jin, S. The Complete Chloroplast Genome of *Carya cathayensis* and Phylogenetic Analysis. *Genes* **2022**, *13*, 369. [[CrossRef](#)] [[PubMed](#)]
52. Song, W.; Ji, C.; Chen, Z.; Cai, H.; Wu, X.; Shi, C.; Wang, S. Comparative Analysis the Complete Chloroplast Genomes of Nine *Musa* Species: Genomic Features, Comparative Analysis, and Phylogenetic Implications. *Front. Plant Sci.* **2022**, *13*, 832884. [[CrossRef](#)] [[PubMed](#)]

53. Wen, F.; Wu, X.; Li, T.; Jia, M.; Liu, X.; Liao, L. The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China. *BMC Genom.* **2021**, *22*, 161. [[CrossRef](#)]
54. Timme, R.E.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* **2007**, *94*, 302–312. [[CrossRef](#)]
55. Weng, M.L.; Blazier, J.C.; Govindu, M.; Jansen, R.K. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* **2014**, *31*, 645–659. [[CrossRef](#)]
56. Morton, B.R. The influence of neighboring base composition on substitutions in plant chloroplast coding sequences. *Mol. Biol. Evol.* **1997**, *14*, 189–194. [[CrossRef](#)]
57. Morton, B.R.; Clegg, M.T. Neighboring Base Composition Is Strongly Correlated with Base Substitution Bias in a Region of the Chloroplast Genome. *J. Mol. Evol.* **1995**, *41*, 597–603. [[CrossRef](#)] [[PubMed](#)]
58. Zhang, R.S.; Yang, J.; Hu, H.L.; Xia, R.X.; Li, Y.P.; Su, J.F.; Li, Q.; Liu, Y.Q.; Qin, L. A high level of chloroplast genome sequence variability in the Sawtooth Oak *Quercus acutissima*. *Int. J. Biol. Macromol.* **2020**, *152*, 340–348. [[CrossRef](#)] [[PubMed](#)]
59. Liu, X.; Chang, E.; Liu, J.; Jiang, Z. Comparative analysis of the complete chloroplast genomes of six white oaks with high ecological amplitude in China. *J. For. Res.* **2021**, *32*, 2203–2218. [[CrossRef](#)]
60. Liu, X.; Chang, E.M.; Liu, J.F.; Huang, Y.N.; Wang, Y.; Yao, N.; Jiang, Z.P. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Quercus bawanglingensis* Huang, Li et Xing, a Vulnerable Oak Tree in China. *Forests* **2019**, *10*, 587. [[CrossRef](#)]
61. Wang, T.R.; Wang, Z.W.; Song, Y.G.; Kozłowski, G. The complete chloroplast genome sequence of *Quercus ningangensis* and its phylogenetic implication. *Plant Fungal Syst.* **2021**, *66*, 155–165. [[CrossRef](#)]
62. Yang, Y.; Hu, Y.; Ren, T.; Sun, J.; Zhao, G. Remarkably conserved plastid genomes of *Quercus* group *Cerris* in China: Comparative and phylogenetic analyses. *Nord. J. Bot.* **2018**, *36*, e01921. [[CrossRef](#)]
63. 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. [[CrossRef](#)]
64. Powell, W.; Morgante, M.; McDevitt, R.; Vendramin, G.G.; Rafalski, J.A. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7759–7763. [[CrossRef](#)]
65. King, D.G.; Soller, M.; Kashi, S.Y. Evolutionary tuning knobs. *Endeavour* **1997**, *21*, 36–40. [[CrossRef](#)]
66. Hansen, D.R.; Dastidar, S.G.; Cai, Z.; Penafior, C.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol. Phylogenet. Evol.* **2007**, *45*, 547–563. [[CrossRef](#)]
67. Huang, H.; Shi, C.; Liu, Y.; Mao, S.Y.; Gao, L.Z. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evol. Biol.* **2014**, *14*, 151. [[CrossRef](#)] [[PubMed](#)]
68. Kim, K.J.; Lee, H.L. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **2004**, *11*, 247–261. [[CrossRef](#)] [[PubMed](#)]
69. Wang, R.J.; Cheng, C.L.; Chang, C.C.; Wu, C.L.; Su, T.M.; Chaw, S.M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* **2008**, *8*, 36. [[CrossRef](#)] [[PubMed](#)]
70. Marechal, A.; Brisson, N. Recombination and the maintenance of plant organelle genome stability. *New Phytol.* **2010**, *186*, 299–317. [[CrossRef](#)] [[PubMed](#)]
71. Zhu, A.; Guo, W.; Gupta, S.; Fan, W.; Mower, J.P. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **2016**, *209*, 1747–1756. [[CrossRef](#)] [[PubMed](#)]
72. Han, Y.W.; Duan, D.; Ma, X.F.; Jia, Y.; Liu, Z.L.; Zhao, G.F.; Li, Z.H. Efficient Identification of the Forest Tree Species in Aceraceae Using DNA Barcodes. *Front. Plant Sci.* **2016**, *7*, 1707. [[CrossRef](#)]
73. Perry, A.S.; Wolfe, K.H. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J. Mol. Evol.* **2002**, *55*, 501–508. [[CrossRef](#)]
74. Wang, X.; Shi, X.; Chen, S.; Ma, C.; Xu, S. Evolutionary Origin, Gradual Accumulation and Functional Divergence of *Heat Shock Factor* Gene Family with Plant Evolution. *Front. Plant Sci.* **2018**, *9*, 71. [[CrossRef](#)]
75. Wicke, S.; Schaferhoff, B.; dePamphilis, C.W.; Muller, K.F. Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. *Mol. Biol. Evol.* **2014**, *31*, 529–545. [[CrossRef](#)]
76. Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*; Oxford University Press: Oxford, UK, 2000; pp. 385–386.
77. Li, X.; Li, Y.; Sylvester, S.P.; Zang, M.; El-Kassaby, Y.A.; Fang, Y. Evolutionary patterns of nucleotide substitution rates in plastid genomes of *Quercus*. *Ecol. Evol.* **2021**, *11*, 13401–13414. [[CrossRef](#)]
78. Hao, D.C.; Chen, S.L.; Xiao, P.G. Molecular evolution and positive Darwinian selection of the chloroplast maturase matK. *J. Plant Res.* **2010**, *123*, 241–247. [[CrossRef](#)] [[PubMed](#)]
79. Zhao, D.N.; Ren, Y.; Zhang, J.Q. Conservation and innovation: Plastome evolution during rapid radiation of *Rhodiola* on the Qinghai-Tibetan Plateau. *Mol. Phylogenet. Evol.* **2020**, *144*, 106713. [[CrossRef](#)] [[PubMed](#)]
80. Xie, D.F.; Yu, H.X.; Price, M.; Xie, C.; Deng, Y.Q.; Chen, J.P.; Yu, Y.; Zhou, S.D.; He, X.J. Phylogeny of Chinese *Allium* Species in Section *Daghestanica* and Adaptive Evolution of *Allium* (Amaryllidaceae, Alliioideae) Species Revealed by the Chloroplast Complete Genome. *Front. Plant Sci.* **2019**, *10*, 460. [[CrossRef](#)] [[PubMed](#)]