

## Article

# The Complete Chloroplast Genome of An *Ophiorrhiza baviensis* Drake Species Reveals Its Molecular Structure, Comparative, and Phylogenetic Relationships

Mai Huong Pham <sup>1</sup> , Thu Hoai Tran <sup>1</sup>, Thi Dung Le <sup>1</sup>, Tung Lam Le <sup>1</sup> , Ha Hoang <sup>1</sup> and Hoang Ha Chu <sup>1,2,\*</sup><sup>1</sup> Institute of Biotechnology (IBT), Vietnam Academy of Science & Technology (VAST), Hanoi 100000, Vietnam<sup>2</sup> Faculty of Biotechnology, Graduate University of Science and Technology, VAST, Hanoi 100000, Vietnam

\* Correspondence: chuhoangha@ibt.ac.vn

**Abstract:** *Ophiorrhiza baviensis* Drake, a flowering medical plant in the Rubiaceae, exists uncertainly within the *Ophiorrhiza* genus' evolutionary relationships. For the first time, the whole chloroplast (cp) genome of an *O. baviensis* Drake species was sequenced and annotated. Our findings demonstrate that the complete cp genome of *O. baviensis* is 154,770 bp in size, encoding a total of 128 genes, including 87 protein-coding genes, 8 rRNAs, and 33 tRNAs. A total of 59 SSRs were screened in the studied cp genome, along with six highly variable loci, which can be applied to generate significant molecular markers for the *Ophiorrhiza* genus. The comparative analysis of the *O. baviensis* cp genome with two published others of the *Ophiorrhiza* genus revealed a high similarity; however, there were some notable gene rearrangements in the *O. densa* plastome. The maximum likelihood phylogenetic trees were constructed based on the concatenation of the *rps16* gene and the *trnL-trnF* intergenic spacer sequence, indicating a close relationship between the studied *O. baviensis* and other *Ophiorrhiza*. This study will provide a theoretical molecular basis for identifying *O. baviensis* Drake, as well as species of the *Ophiorrhiza* genus, and contribute to shedding light on the chloroplast genome evolution of Rubiaceae.



**Citation:** Pham, M.H.; Tran, T.H.; Le, T.D.; Le, T.L.; Hoang, H.; Chu, H.H. The Complete Chloroplast Genome of An *Ophiorrhiza baviensis* Drake Species Reveals Its Molecular Structure, Comparative, and Phylogenetic Relationships. *Genes* **2023**, *14*, 227. <https://doi.org/10.3390/genes14010227>

Academic Editors: Wajid Zaman and Hakim Manghwar

Received: 30 November 2022

Revised: 19 December 2022

Accepted: 7 January 2023

Published: 15 January 2023



**Copyright:** © 2023 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/>).

**Keywords:** *Ophiorrhiza baviensis*; chloroplast genome; comparative analysis; phylogeny

## 1. Introduction

The chloroplast (cp) is an essential organelle in photosynthetic plant and microbial cells that produces energy to feed the cell through photosynthesis [1]. Each chloroplast contains its own ribosomes and a separate genome from the cell's nuclear genome, ranging in size from 20 to 160 kilobase pairs (kp). The cp genome is uniparentally inherited with a quadripartite structure consisting of one large single-copy (LSC) region, one small single-copy (SSC) region, and two inverted repeat regions (IRs) of the same length [2]. As a result of the small size of the cp genome, which contains only around 100 to 120 protein coding genes, chloroplasts are often the first target for sequencing in evolutionary analysis, barcoding, and meta-barcoding [2]. In the NCBI Genbank database at present, there are more than 1000 cp genomes of plant species. However, this number is very small compared to the existing plant diversity on the planet, which raises the need to collect and store sequences of uncharacterized species.

For medicinal plants such as *Ophiorrhiza baviensis*, the potential for exploitation and the need for systematic classification are even more essential. *O. baviensis* is a species of flowering plant in the Rubiaceae family, first described scientifically by Drake in 1895, and re-identified by Wu et al. [3]. Information on the ecology and genomic characteristics of this species is extremely limited, with only four sequences of *O. baviensis*—the gene junctions *trnL-trnF* (#MH626989.1), *rps16* (#MH626923.1), the external transcribed spacer (ETS) (#MH626743.1), and *ITS* (#MH626804.1)—available on the Genbank database of the National Center for Biotechnology Information (USA) (NCBI). Each sequence is less than

1000 base pairs (bp) in size, only two of which belong to the chloroplast genome. Thus, there is a need to study the entire chloroplast genome of *O. baviensis* species for taxonomy and diversity assessment, as well as chloroplast genome characterization, conservation, and future research. With an estimated chloroplast genome size of 154 kb, the potential for exploiting genomic information on the *O. baviensis* chloroplast genome is very large.

Recently, PacBio sequencing technology has been applied to sequence cp genomes, and there have been studies demonstrating the superior ability of PacBio in de novo assembly with 99% accuracy; moreover, as the repeatability increases, this can exceed 99.9% [4]. PacBio sequencing is also a great technology in resolving gaps in rRNA, i.e., internal transcribed spacer (ITS) regions and the surrounding regions to obtain accurate molecular biology information for species identification. For the first time, we report a new complete chloroplast genome of *O. Baviensis* Drake from Vietnam and compare it with previously published *Ophiorrhiza* complete chloroplast genome data to evaluate the genome organization, phylogenetic relationships, and conserved genetic resources.

## 2. Materials and Methods

### 2.1. Sample Collection and Chloroplast Genome Sequencing

*O. baviensis* samples were collected in Ba Vi National Park, Hanoi, Vietnam in August 2022 (code number: Xacan 01), 1217.6 m, 21°3'32" N; 105°4'58" E (Figure 1). The voucher specimens were placed in the herbarium of the Institute of Ecology and Biological Resources (HN), Hanoi, Vietnam. Fresh leaves with the same code number were used to extract genomic DNA.



**Figure 1.** *O. baviensis* Drake. (A) Habitat; (B) Morphological characteristic of infructescence in side view; Photos by Thu Hoai Tran.

### 2.2. DNA Extraction and Chloroplast Genome Sequencing

We treated samples prior to extraction with the Chloroplast Isolation Kit (ab234623-Abcam, Cambridge, UK) for cp enrichment to increase the cpDNA concentration. The total DNA was extracted by the GeneAll®Exgene™ Plant SV mini kit using the enriched samples (including both genomic DNA and cp DNA). The extracted DNA integrity was evaluated by electrophoresis on a 0.8% agarose gel for 45 minutes at 120 V, and the DNA concentration was measured by Nanodrop 2000 (Thermo, Waltham, MA, USA) and Qubit 2.0 devices to ensure quality for library preparation and sequencing on the Pacbio system according to the manufacturer's instructions.

The total DNA was fragmented and the DNA damage from fragmentation, as well as the 5'/3' ends, underwent repair using the SMRTbell Damage Repair Kit SPv3 (#100-992-200, Pacific Biosciences, Menlo Park, CA, USA) before being attached to PacBio adapters. Products without adapters are rejected by the Exo III and Exo VII enzymes. The SMRTbell library was cleaned with Ampure PB beads (Beckman Coulter, Brea, CA, USA) and checked for length and concentration using the Bioanalyzer 2100. Subsequently, it was cleaned and sized using BluePippin (SageScience, Beverly, MA, USA) with a gel concentration of 0.75% to filter out library DNA fragments above 20 kb in length. The library was lastly checked for size and fragmentation with the Bioanalyzer 2100 before loading to the SMRT Cell (#101-008-000, PacBio).

The prepared library was loaded on one chip and sequenced on a PacBio SEQUEL system at the National Key Laboratory for Gene Technology, Institution of Biotechnology (Hanoi, Vietnam). SMRTbell library was attached with polymerase and purified using the Sequel Binding Internal Ctrl Kit 2.0 (#101-400-900, PacBio) and the SMRTbell Clean Up Column v2 Kit-Dif (101-184-100, PacBio) according to the procedure generated by the Sample Setup software included in the SMRTLink portal version 5.1.

### 2.3. Genome Assembly and Annotation

Total DNA was sequenced using the PacBio platform. Sequences derived from the cp genome were identified through the pbmm2 program using the cp genome of the reference *Ophiorrhiza* species (accession number: NC\_057496.1) obtained from the Genbank database [5]. Then, the Hierarchical Genome Assembly Process version 4 (HGAP4) software was used to assemble the cp genome [6]. Protein-coding genes and RNA were annotated by the GeSeq webtool [7], while tRNAscan-SE software version 2.0 was applied to verify the tRNA genes [8]. The OrganellarGenomeDRAW (OGDRAW) web-tool was selected to generate the circular gene map [9]. Repeat elements were identified using two approaches. The web-based MISA finder was used for detecting microsatellites in nucleotide sequences, with the following parameters: 10 repeats for mono-, 5 for di-, 4 for tri-, and 3 for tetra-, penta-, and hexa-nucleotide SSRs [10]. Size comparison of the SSRs among the SSRs of each type was used to count polymorphic SSRs. The size and pattern of repeats in the cp genome were identified using the REPuter with the following set of parameters: minimum repeat size 20 bp, hamming distance 3 kb, and 90% or more sequence similarity [11].

### 2.4. Genome Comparison and Phylogenetic Identification

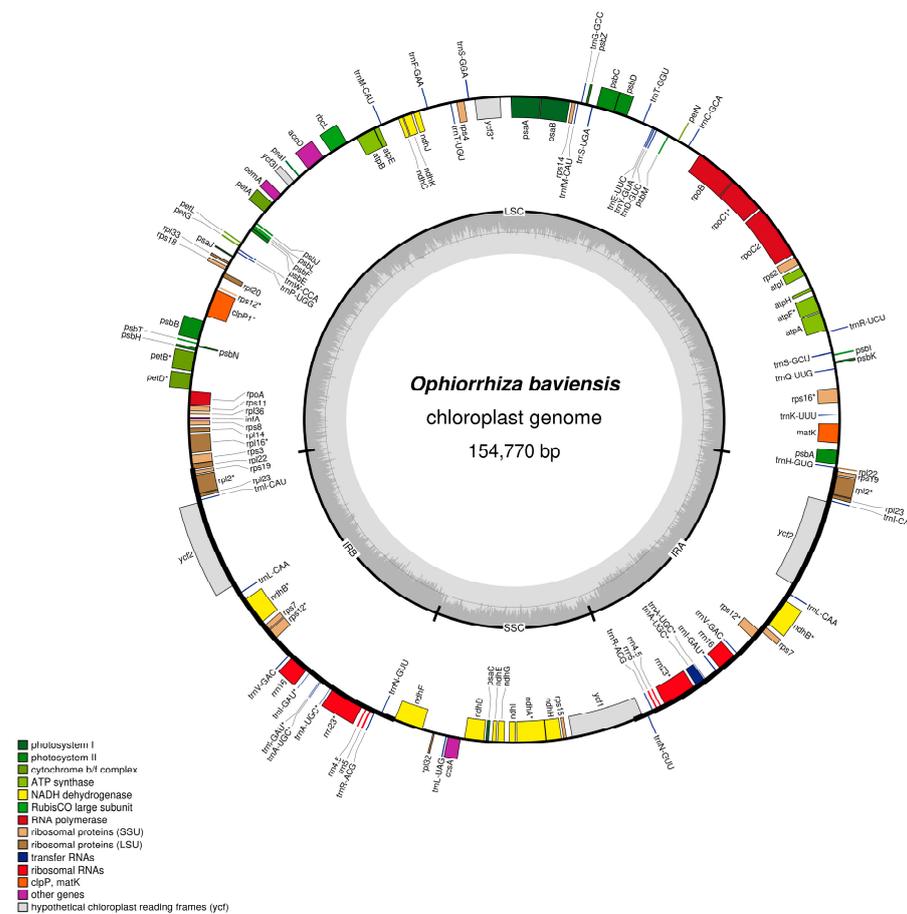
For cp genome comparison, we collected available cp genomes of *Ophiorrhiza* species (*O. pumila*—NC\_057496.1 and *O. densa*—NC\_058252.1) from the GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>, accessed on 15 November 2022). The overall genome structure, gene content, genome size, and number of repeats across the genomes were compared. The entire cp genome sequences of the *Ophiorrhiza* species were aligned through MAFFT software with default parameters and visualized in the mVISTA webtool with the LAGAN mode [12]. We used the annotated cp genome of the project as the reference genome in the mVISTA diagram. Subsequently, Irscope was used to visualize and compare the contiguous region between the large and small single-copy, along with the inverted repeat regions of the genomes. We also examined codon usage bias and sequence divergence via computational nucleotide diversity ( $P_i$ ) analysis among cp genomes in DnaSP software version 6.12.03 [13]. For the sequence divergence analysis, we applied a window size of 600 bp with a step size of 200 bp.

A concatenation of the *rps16* gene and *trnL-trnF* intergenic spacer sequences from the *Ophiorrhiza* species and two *Xanthophyllum* species of the Rubiaceae family from the Genbank database was used to identify the phylogenetic relationships of the studied *O. Baviensis* Drake. The nucleotide sequences were aligned with MAFFT software with default parameters [14] before the maximum likelihood (GTR+CAT model) phylogenetic tree was constructed using FastTree [15] with a 1000 bootstrap and visualized by FigTree software version 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>, accessed on 1 July 2021).

### 3. Results

#### 3.1. Chloroplast Genome Assembly and Annotation

Using the PacBio SEQUEL I system, 28,402,467,862 bp of raw sequence data were generated with a mean read length of 1938 bp, an N50 contig size of 2412 bp, and approximately 9% of the raw reads belonging to the *O. baviensis* cp genome with 158 × coverage. The resequencing assembly resulted in a circular cp genome size of 154,770 bp (Figure 2), and the percentage of GC content was 37.6%. As reported in most angiosperm cp genomes, the assembled *O. baviensis* Drake plastome demonstrated the typical quadripartite structure consisting of four regions, LSC (84,626 bp), SSC (18,574 bp), and a pair of inverted repeats (IRs 25,685 bp).



**Figure 2.** Chloroplast genome map of *O. baviensis* Drake in Vietnam. Genes shown inside the circle are transcribed clockwise, whereas genes outside are transcribed counterclockwise. The light gray inner circle shows the AT content, the dark gray corresponds to the GC content.

In addition, the annotation results from GeSeq and tRNAscan-SE revealed that the *O. baviensis* Drake cp genome possessed a total of 128 genes, of which there were 87 protein-coding genes, 33 tRNA genes, and 8 rRNA genes (16S, 23S, 5S, and 4.5S) (Table 1). The annotated gene models were assigned into three major groups based on their functions. Regarding the photosynthesis-related gene category, there were 44 genes encoding the subunits of ATP synthase, cytochrome complex, photosystem I and II, and putative NADPH dehydrogenase, along with the large subunit of Rubisco related to the photosynthetic electron transport chain. The other 76 genes were functionally characterized in the transcription and translation processes. The majority were tRNA genes, and the others were rRNA genes and genes encoding DNA-dependent RNA polymerase, the subunits of the ribosome, and ribosome proteins. The remaining nine genes were classified in the category of other genes, consisting of five genes with reported functions in RNA processing (*matK*),

c-type cytochrome synthesis (*ccsA*), fatty acid synthesis (*accD*), carbon metabolism (*cemA*), and proteolysis (*clpP*). In addition, four genes encoding the conserved reading frames (*ycf1*, *ycf2*, and *ycf3*) were also annotated in the cp genome.

**Table 1.** Summary of the chloroplast genome of *O. baviensis* Drake species.

Genome Features	<i>O. baviensis</i> Drake
Genome size (bp)	154,770 bp
LSC size (bp)	84,826
SSC size (bp)	18,574
IR size (bp)	25,685
GC content (%)	37.6
No. of genes	128
No. of PCGs	87
No. of tRNA	33
No. of rRNA	8

Otherwise, each IR region of the *O. baviensis* cp genome was annotated to comprise 18 genes (all 4 rRNA genes, 7 tRNA genes, 1 NADH-dehydrogenase protein-coding gene, 4 ribosomal protein-coding genes, and 2 other genes). There were 17 cp genes that harbored introns, among which 15 genes (*atpF*, *rpl2* ( $\times 2$ ), *rpl16*, *ndhA*, *ndhB* ( $\times 2$ ), *rpoC1*, *rps12*, *rps16*, *trnA-UGC* ( $\times 2$ ), *trnG-GCC*, and *trnI-GAU* ( $\times 2$ )) contained a single intron, while two genes (*ycf3*, *clpP*) had double introns (Table 2).

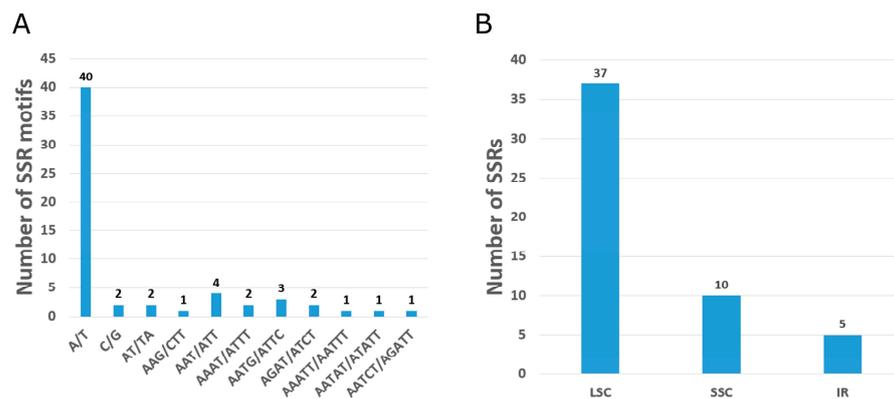
**Table 2.** Gene composition of *O. baviensis* Drake chloroplast genome.

Category of Genes	Group of Genes	Name of Genes
Photosynthesis	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpFa</i> , <i>atpH</i> , <i>atpI</i>
	Subunits of NADH-dehydrogenase	<i>ndhAa</i> , <i>ndhB</i> ( $\times 2$ ) <i>a</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Subunits of cytochrome b/f complex	<i>petL</i> , <i>petB</i> , <i>petG</i> , <i>petA</i> , <i>petD</i> , <i>petN</i>
	Subunits of photosystem I	<i>psaJ</i> , <i>psaC</i> , <i>psaA</i> , <i>psaI</i> , <i>psaB</i>
	Subunits of photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
	Subunit of rubisco	<i>rbcL</i>
	Large subunit of ribosome	<i>rpl14</i> , <i>rpl16a</i> , <i>rpl2</i> ( $\times 2$ ) <i>a</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> ( $\times 2$ ), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
Transcription and translation	DNA-dependent RNA polymerase	<i>rpoB</i> , <i>rpoA</i> , <i>rpoC1a</i> , <i>rpoC2</i>
	Small subunit of ribosomal proteins	<i>rps11</i> , <i>rps12</i> ( $\times 2$ ) <i>a</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16a</i> , <i>rps18</i> , <i>rps19</i> ( $\times 2$ ), <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> ( $\times 2$ ), <i>rps8</i>
	rRNA genes	<i>rrn23S</i> ( $\times 2$ ), <i>rrn16S</i> ( $\times 2$ ), <i>rrn5S</i> ( $\times 2$ ), <i>rrn4.5S</i> ( $\times 2$ ) <i>trnA-UGC</i> ( $\times 2$ ) <i>a</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnG-GCCa</i> , <i>trnH-GUG</i> , <i>trnI-GAU</i> ( $\times 2$ ) <i>a</i> , <i>trnL-CAA</i> ( $\times 2$ ), <i>trnL-UAG</i> , <i>trnN-GUU</i> ( $\times 2$ ), <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> ( $\times 2$ ), <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> ( $\times 2$ ), <i>trnW-CCA</i> , <i>trnY-GUA</i>
	tRNA genes	
Other genes	Translational initiation factor	<i>infA</i>
	Subunit of acetyl-CoA-carboxylase (fatty acid synthesis)	<i>accD</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
	Envelope membrane protein (carbon metabolism)	<i>cemA</i>
	Protease	<i>clpPb</i>
	Maturase (RNA processing)	<i>matK</i>
Conserved open reading frames	<i>ycf1</i> , <i>ycf2</i> ( $\times 2$ ), <i>ycf3b</i>	

Genes marked with the sign are the gene with a single (a) or double (b) introns and duplicated genes ( $\times 2$ ).

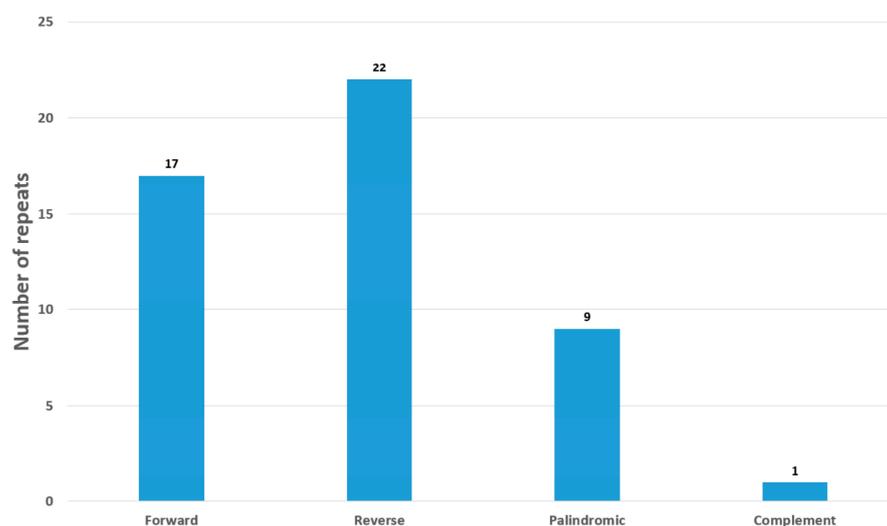
### 3.2. Repeat Sequences and Codon Analysis

A total of 59 simple sequence repeats (SSRs) were investigated in the *O. baviensis* Drake chloroplast genome via the MISA web-tool. Almost all of the screened repeats were mono repeats (composed of A/T and C) with the size ranging from 10 to 16 bp (Figure 3A). Two di-, five tri-, seven tetra-, and three penta-nucleotide SSRs were found in the *O. baviensis* Drake plastid. A total of 53 SSRs were classified as simple based SSRs and the six remaining SSRs presented in a compound formation. The majority of SSR types were discovered in the LSC, while the IR regions included the smallest number of SSRs (Figure 3B).



**Figure 3.** Analysis of single sequence repeats (SSRs) of the *O. baviensis* Drake chloroplast genome. (A) Number of identified SSR sequence motifs; (B) Frequency of repeat types in LSC, SSC, and IR regions.

The cp genome of *O. baviensis* Drake was annotated to possess 49 long repeats including 9 palindromic repeats, along with 12 forward and 22 reverse repeats. There was only one complement repeat (Figure 4). The unit size of the detected repeats ranged from 20 to 58 bp, while a majority of the repeat size (67%) was shorter than 30 bp.



**Figure 4.** Repeat analysis of *O. baviensis* Drake chloroplast genome.

The codon usage frequency of 64 protein-coding genes was evaluated for three cp genomes: *O. baviensis* Drake and two other available *Ophiorrhiza* species. The total number of codons found in the coding regions was 51,517, while the A- and U-ending were found more frequently than the G/C-ending (Table 3). Leucine was the most prevalent among the 20 amino acids with a percentage of 10.46% (5068 codons), followed by serine with 9.95% (4817 codons). Meanwhile, the rarest was tryptophan with a total of 681 codons accounting

for approximately 1.4%. A total of 30 codons exhibited the codon usage bias ( $RSCU < 1$ ), while 32 codons were observed to be more frequent than the expected usage at equilibrium ( $RSCU > 1$ ) (Table 3). The usage frequency for the start codons AUG and UGG (methionine and tryptophan) exhibited no bias ( $RSCU = 1$ ).

**Table 3.** Relative synonymous codon usage (RSCU) for protein-coding genes in *O. baviensis*.

Codon	AA	Frequency	RCSU	Codon	AA	Frequency	RCSU	Codon	AA	Frequency	RCSU
UAA	*	1259	1.22	AUC	I	1205	0.80	CGG	R	420	0.75
UAG	*	825	0.80	AUA	I	1471	0.98	AGA	R	1093	1.94
UGA	*	1004	0.98	AAA	K	2050	1.35	AGG	R	627	1.12
GCU	A	446	1.23	AAG	K	982	0.65	UCU	S	1113	1.39
GCC	A	351	0.97	UUA	L	1040	1.23	UCC	S	982	1.22
GCA	A	401	1.11	UUG	L	1095	1.30	UCA	S	824	1.03
GCG	A	250	0.69	CUU	L	1063	1.26	UCG	S	622	0.77
UGU	C	679	1.20	CUC	L	653	0.77	AGU	S	747	0.93
UGC	C	449	0.80	CUA	L	737	0.87	AGC	S	529	0.66
GAU	D	1012	1.42	CUG	L	480	0.57	ACU	T	668	1.13
GAC	D	413	0.58	AUG	M	856	1.00	ACC	T	651	1.10
GAA	E	1337	1.43	AAU	N	1779	1.38	ACA	T	647	1.09
GAG	E	537	0.57	AAC	N	800	0.62	ACG	T	406	0.68
UUU	F	2212	1.20	CCU	P	611	1.03	GUU	V	784	1.38
UUC	F	1481	0.80	CCC	P	618	1.04	GUC	V	411	0.72
GGU	G	540	0.96	CCA	P	726	1.23	GUA	V	682	1.20
GGC	G	383	0.68	CCG	P	414	0.70	GUG	V	402	0.71
GGA	G	747	1.33	CAA	Q	987	1.40	UGG	W	681	1.00
GGG	G	577	1.03	CAG	Q	420	0.60	UAU	Y	1345	1.36
CAU	H	880	1.36	CGU	R	376	0.67	UAC	Y	637	0.64
CAC	H	414	0.64	CGC	R	280	0.50				
AUU	I	1830	1.22	CGA	R	576	1.02				

\* Stop codon.

### 3.3. Chloroplast Genome Comparison

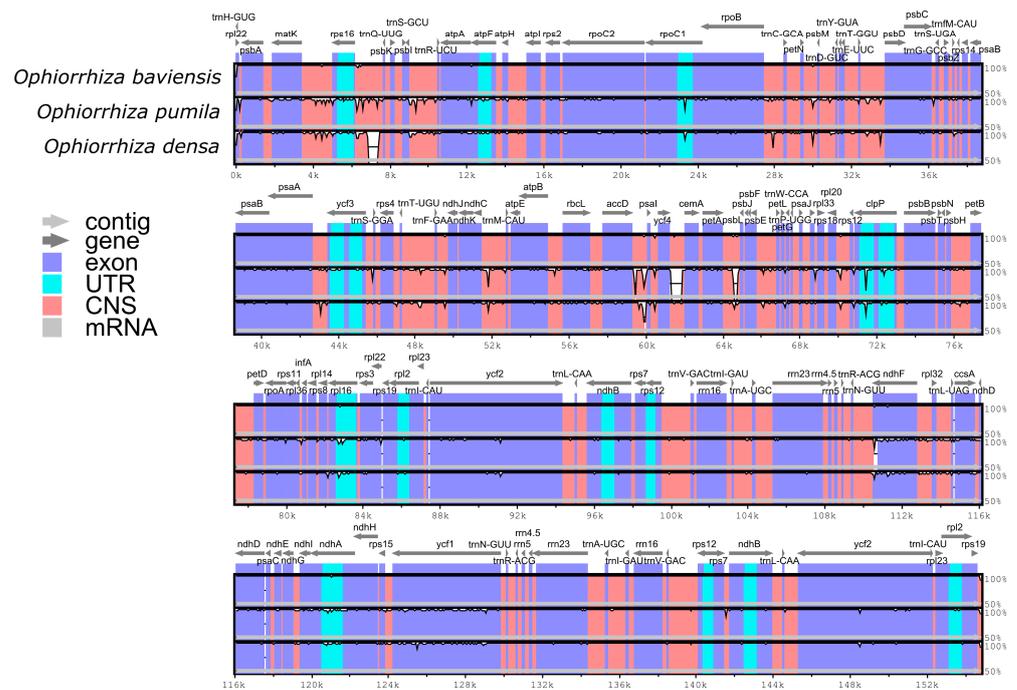
To characterize genomic divergence, the percentage of sequence identity was evaluated for three *Ophiorrhiza* species with the functional annotation of *O. baviensis* Drake as a reference. The comparison using the mVISTA program revealed that the gene organization among the three species was highly similar and there were several regions of sequence variation (Figure 5). The results exhibited a higher frequency of divergence in the LSC and SSC regions than in the IR regions. Moreover, the coding regions of the three cp genomes were observed to be more conserved, whereas a majority of the detected variations were screened in the conserved non-coding sequences (CNS). Among the protein-coding gene sequences, the highly disparate genes consisted of *matK*, *rpoC2*, *rpoB*, *clpP*, *rpl16*, *ndhF*, *ndhA*, and *ycf1*.

The sliding window analysis indicated that the average polymorphism information (Pi) values of the LSC (Pi = 0.005635) and SSC (Pi = 0.007472) regions were greater than that of the IR (Pi = 0.001285) regions, which showed that most of the variations were located in the LSC and SSC regions (Figure 6). Of the three *Ophiorrhiza* species, the average value of nucleotide diversity (Pi) was 0.00441.

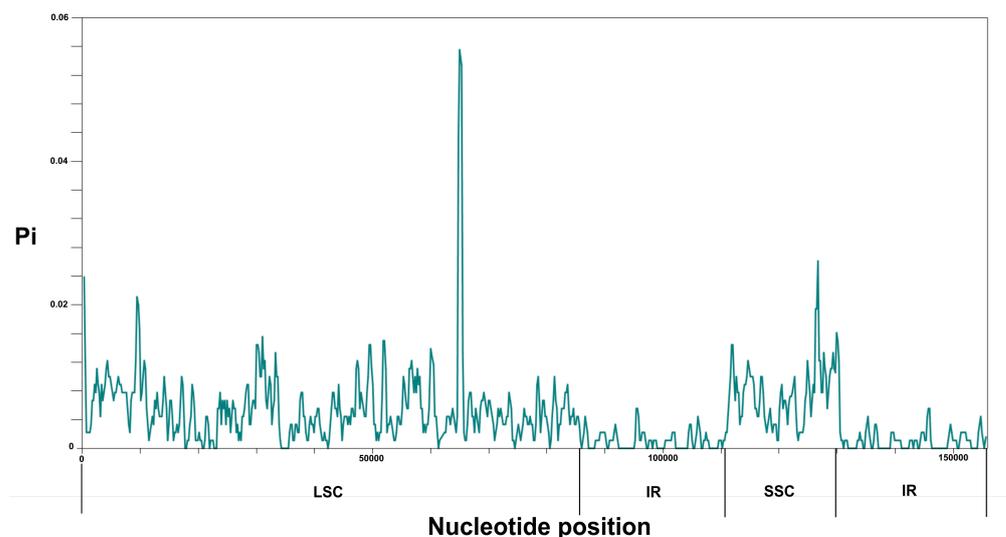
### 3.4. IR Contraction and Expansion in the Chloroplast Genome

The IR/LSC and IR/SSC boundaries of three *Ophiorrhiza* cp genomes were compared using the IRscope program. Overall, the results indicated that the region size, gene organization, and gene content showed a high similarity among the cp genome of *O. baviensis* and *O. pumila* (Figure 7). On the other hand, the *O. densa* cp genome showed several variants with the two abovementioned *Ophiorrhiza* species. The size of IR regions ranged from 25,684 bp (*O. baviensis* Drake) to 26,066 bp (*O. pumila*), and the size of IR of *O. densa* was 25,701 bp. The *rpl22* gene was located within the LSC region with a 102 bp overlap with

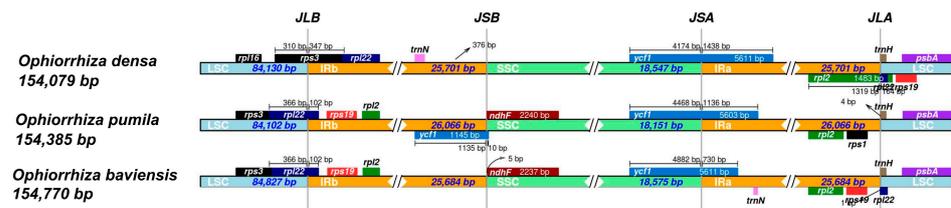
the IRb for *O. baviensis* and *O. pumila*, while *O. densa* showed a 347 bp overlap of the *rps3* gene in this boundary. Apart from *O. densa*, the *ndhF* gene was detected on the boundary of the SSC and IRb region. The border across IRa and SSC was found in the *ycf1* gene with 1438, 1316, and 730 bp tail sections of the gene placed in the IRa of *O. densa*, *O. pumila*, and *O. baviensis*, respectively (Figure 7). The IRa and LSC boundary showed the presence of the *trnH* gene in the forward strand of all three species and the *rpl22* gene in the reverse strand of *O. baviensis* and *O. densa*. The results of the IR analysis indicated extensive contraction and expansion of the IR regions in the three species.



**Figure 5.** Complete chloroplast genome alignments of the three *Ophiorrhiza* species. The horizontal axis indicates the coordinates within the chloroplast genome. The vertical scale indicates the percent identity within 50–100%. Annotated genes are displayed along the top.



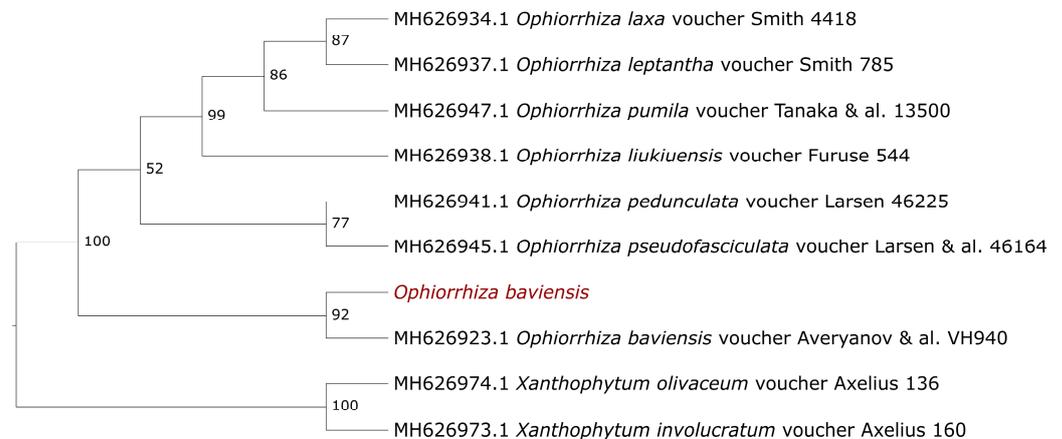
**Figure 6.** Nucleotide diversity (Pi) values among the three *Ophiorrhiza* species. X-axis: the position in the genome; Y-axis: Pi value. Pi, polymorphism information.



**Figure 7.** Comparison of LSC, IR, and SSC junction positions among the three *Ophiorrhiza* chloroplast genomes. JLB (junction IRb/LSC), JSB (junction IRb/SSC), JSA (junction IRa/SSC), JLA (junction IRa/LSC).

### 3.5. Phylogenetic Inference

The number of available sequences of *O. baviensis* on the Genbank databases, especially belonging to the cp genome, is limited (only the *rps16* gene and the *trnL-trnF* intergenic spacer). Therefore, we extracted these sequences from the assembled cp genome and used them to access the phylogenetic relationship of the studied *O. baviensis* at the species level. Figure 8 shows the phylogenetic resolution based on the concatenated sequence between the *rps16* gene and the *trnL-trnF* intergenic spacer with a high bootstrap value of 92% between the studied *O. baviensis* Drake and the reference *O. baviensis* voucher Averyanov & al. VH940 (AAU) (Accession number: MH626923.1). With a bootstrap value of 100%, all eight *Ophiorrhiza* species were grouped separately from the two *Xanthophytum* species as an outgroup. In the case of barcoding among the Rubiaceae family, the combined *rps16-trnL-F* intergenic spacer sequences provided a high capacity for phylogenetic resolution.



**Figure 8.** The maximum likelihood phylogenetic trees based on the concatenated sequences between the *rps16* genes and the *trnL-trnF* intergenic spacer. Numbers on the branches indicate bootstrap percentage after 1000 replications in constructing the tree. The species investigated in this study are colored in red.

## 4. Discussion

Rubiaceae is a family of flowering plants containing 620 genera with approximately 13,500 species over the world, which makes it the fourth-largest angiosperm family. Over 300 cp genomes in the Rubiaceae family have been published in the Genbank database until now, only three of which belong to *Ophiorrhiza*. The genus *Ophiorrhiza* consists of about 200–300 species mainly distributed in humid tropical forests from East India to the Western Pacific, and from South China to Northern Australia [16,17]. Bioactive compounds from this family, such as quinine, emetine, caffeine, and camptothecin are of major pharmaceutical importance; thus, many species in the genus *Ophiorrhiza* are of interest [18]. In the present study, we sequenced and annotated the entire cp genome of a Vietnamese medicinal plant.

Angiosperm cp genomes have a highly conserved gene order and gene content with 127–134 genes found across the chloroplast genomes. The analyzed *O. baviensis* cp genomes

demonstrated the typical quadripartite structure and showed the expected size range (~154 kb) for *Ophiorrhiza* species and the conserved gene contents. Our gene annotation results were similar to the genetic properties of angiosperm chloroplast genomes. The number of genes present in the cp genome from *O. baviensis* was 128, of which, 17 genes included one or two introns. In addition, the deletions of the *petB* and *petD* introns were observed in the studied *O. baviensis* cp genome, which also occurred in *O. pumila* species. Introns play an important role in gene expression regulation. Recent research has revealed gene or intron loss in chloroplast genomes [19–21], among which *petB* and *petD* intron loss was reported in many angiosperms [22].

In addition to two copies of IR regions, 49 small repeats were found to be located within coding and non-coding regions of the *O. baviensis* plastome. The cp genome includes numerous dispersed repeats, which are supposed to be biomarkers of mutational hotspots [23,24]. The repeat number is similar to the data of other species belonging to the Rubiaceae family [25,26]. Repeats are closely related to angiosperm plastome reconstruction and can be assumed as recognition signals of recombination because of their potential to generate secondary structures. In this study, the similar number of repeats in comparison with previous estimates might not demonstrate inter- and intra-specific plastome recombination. In higher plants, SSRs are identified as crucial molecular markers for the investigations of population variation due to their distinct uniparental inheritance, and they are commonly used to evaluate genetic diversity and population structure in evolutionary studies [27–29]. In total, 59 SSRs were screened in the *O. baviensis* cp genomes with strong A/T bias. These repeats play a significant role for generating genetic markers in *O. baviensis* species, which may be applied to assess the variation at the intraspecific level in phylogenetic and ecological studies.

Comparative analyses on *O. baviensis* and two available *Ophiorrhiza* cp genomes were implemented to explore the plastome structure in the taxa. The cp genome size of the three *Ophiorrhiza* ranged from 154,079 bp (*O. densa*) to 154,770 bp (*O. baviensis*), the figure for *O. pumila* was 154,385 bp. Gene organization and codon usage patterns exhibited high conservation, which could be applicable for further population genetics and phylogenetic studies. Moreover, the three *Ophiorrhiza* cp genomes were less variable in their coding regions than in their noncoding regions, which is consistent with the common pattern in most angiosperms [30] (Figure 5). Codon usage preference is closely related to gene expression and can affect the level of mRNA and proteins in the genome [31–33]. The most prevalent amino acid in the *Ophiorrhiza* was leucine (Leu), which has also been commonly detected in the other angiosperms. The high similarity in codon usage may indicate that these *Ophiorrhiza* species underwent similar environmental pressure through their evolutionary processes. The *Ophiorrhiza* cp genomes indicated that the RSCU values of most codons ending in A/U were greater than 1, which may be caused by a bias toward a high A/T ratio in composition. Additionally, we investigated that the partial sequences of the *ycf1* gene along with five intergenic spacers (IGSs), including *petA-psbJ*, *trnH-GUG-psbA*, *trnS-GCU-trnR-UCU*, *psbM-trnD-GUC*, and *ndhC-trnM-CAU*, had relatively high nucleotide diversity values ( $P_i > 0.015$ ). These divergence regions could be studied to provide molecular markers for DNA barcoding and phylogenetic research in *Ophiorrhiza*.

While the three plastomes showed an approximate similarity in genome size, the size of the structural regions exhibited significant differences in a detailed comparison of junction sites (Figure 7). The regions of the cp genome frequently undergo length variations during the evolution of terrestrial plants, which leads to the emergence of many boundary features [34]. The expansion and contraction of the boundaries between IRs and the single-copy (SC) regions are the primary causes of the size change in cp genomes and influence the evolution rate of cp genomes [35,36]. Our finding revealed that the boundary-gene set of the *Ophiorrhiza* species included *rpl22*, *rps19*, *ndhF*, *ycf1*, and *trnH*. Several notable gene rearrangements were observed in the *O. densa* plastome; these were the presence of the *rps3* gene at the JLB instead of the *rpl22* gene, the expansion of the *rpl2* gene to the JLA, and the absence of the *rps19* gene in the IR regions. Expansion and contraction, as well as variation,

at the junction of the SC–IR regions were characterized, suggesting that gene organization in the IR regions can report the distance between species to some extent.

The majority of taxonomic levels of plant phylogenetic connections have been demonstrated using complete chloroplast genomes and protein-coding genes [37,38]. The current study provides the phylogeny of the *Ophiorrhiza* genus based on the combined *rps16-trnL-F* intergenic spacer sequences. The previous study of Razafimandimbison and Rydin demonstrated that *O. baviensis* had been resolved as a sister relationship with *O. japonica* and *O. hayatana* [39]. In terms of species classification, the phylogenetic tree based on the concatenation of the *rps16* gene and the *trnL-F* intergenic spacer indicated the close relationship between the studied plant and the *O. baviensis* voucher Averyanov & al. VH940 (AAU) with a high bootstrap value of 92%. This approach showed effectiveness in the classification of the lower taxonomic levels among the Rubiaceae family. Further, the combination of these barcodes can lead to better species classification compared to the results from a single gene [39]. This study will help to clarify the evolutionary position of *O. baviensis* in the *Ophiorrhiza* genus, as well as offering applicable cp genome data for further research into the genesis and diversification of the Rubiaceae family. Overall, our phylogenetic investigation of the *O. baviensis* cp genome was successful in discovering the intragenetic connections within the *Ophiorrhiza* genus.

## 5. Conclusions

In this study, the first complete chloroplast genome of an *O. baviensis* Drake species from Vietnam was characterized and compared with two other published *Ophiorrhiza* plastomes. The assembly resulted in a whole cp genome of 154,770 bp in size. According to the comparative result, the structure and gene content of three *Ophiorrhiza* cp genomes exhibited a high similarity, and the SC-IR junction analysis revealed the expansion and contraction of IR regions. Additionally, the phylogenetic tree indicated close relationships between our novel cp genome sequence and other *Ophiorrhiza* species. This study provides the potential to employ cp genomes for enhancing species classification and genetic source conservation during further study of the Rubiaceae family.

**Author Contributions:** Conceptualization, H.H.C.; sampling, T.H.T.; methodology, T.H.T., H.H. and T.D.L.; software, M.H.P.; validation, T.L.L.; formal analysis, M.H.P.; data curation, T.L.L.; writing—original draft preparation, M.H.P.; writing—review and editing, H.H.C.; visualization, M.H.P.; supervision, H.H.C. and T.H.T.; project administration, H.H.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the project of the Vietnam Academy of Science and Technology (VAST): “Sequencing and characterizing the chloroplast genome of an *Ophiorrhiza baviensis* species by PacBio SMRT next-generation sequencing technology for genetic classification and conservation” (project no. CSCL08.02/22-22).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** This complete chloroplast genome of *O. baviensis* Drake has been deposited at DDBJ/ENA/GenBank under the accession number OP902221.

**Acknowledgments:** We thank Khang Sinh Nguyen, researcher at the Institute of Ecology and Biological Resources, Vietnam Academy of Science and Technology for authenticating the taxonomic identification of the plant samples.

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

## References

1. Neuhaus, H.E.; Emes, M.J. Nonphotosynthetic Metabolism in Plastids. *Annu. Rev. Plant Biol.* **2000**, *51*, 111. [[CrossRef](#)]
2. Bendich, A.J. Circular Chloroplast Chromosomes: The Grand Illusion. *Plant Cell* **2004**, *16*, 1661–1666. [[CrossRef](#)]
3. Lei, W.; Liu, W.-J.; Nguyen, K.S. Revision of Three Taxa of *Ophiorrhiza* (Rubiaceae) from China. *Phytotaxa* **2019**, *387*, 129–139. [[CrossRef](#)]

4. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [[CrossRef](#)] [[PubMed](#)]
5. PacificBiosciences. Pbbmm2: A Minimapp2 Frontend for PacBio Native Data Formats. Available online: <https://github.com/PacificBiosciences/pbbmm2> (accessed on 10 January 2021).
6. Chin, C.-S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E.; et al. Nonhybrid, Finished Microbial Genome Assemblies from Long-Read SMRT Sequencing Data. *Nat. Methods* **2013**, *10*, 563–569. [[CrossRef](#)]
7. 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. [[CrossRef](#)] [[PubMed](#)]
8. Chan, P.P.; Lin, B.Y.; Mak, A.J.; Lowe, T.M. TRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. *Nucleic Acids Res.* **2021**, *49*, 9077–9096. [[CrossRef](#)] [[PubMed](#)]
9. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): A Tool for the Easy Generation of High-Quality Custom Graphical Maps of Plastid and Mitochondrial Genomes. *Curr. Genet.* **2007**, *52*, 267–274. [[CrossRef](#)] [[PubMed](#)]
10. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-Web: A Web Server for Microsatellite Prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]
11. Kurtz, S.; Schleiermacher, C. REPuter: Fast Computation of Maximal Repeats in Complete Genomes. *Bioinformatics* **1999**, *15*, 426–427. [[CrossRef](#)]
12. 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](#)]
13. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [[CrossRef](#)] [[PubMed](#)]
14. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief. Bioinform.* **2019**, *20*, 1160–1166. [[CrossRef](#)] [[PubMed](#)]
15. Price, M.; Dehal, P.; Arkin, A. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)]
16. Chen, T.; Taylor, C. Ophiorrhiza. In *Flora of China*; Press, S., Ed.; Beijing & Missouri Botanical Garden Press: St. Louis, MO, USA, 2011; Volume 19, pp. 258–282.
17. Lei, W.; Tan, Y.; Hareesh, V.S.; Liu, Q. Ophiorrhiza Macrocarpa (Rubiaceae), a New Viviparous Species from Yunnan, Southwestern China. *Nord. J. Bot.* **2018**, *36*, njb-01637. [[CrossRef](#)]
18. Hamzah, A.S. Isolation, Characterization and Biological Activities of Chemical Constituents of Ophiorrhiza and Hedyotis Species. Ph.D. Dissertation, Universiti Pertanian Malaysia, Serdang, Malaysia, 1994.
19. GAO, L.; SU, Y.-J.; WANG, T. Plastid Genome Sequencing, Comparative Genomics, and Phylogenomics: Current Status and Prospects. *J. Syst. Evol.* **2010**, *48*, 77–93. [[CrossRef](#)]
20. Frailey, D.C.; Chaluvadi, S.R.; Vaughn, J.N.; Coatney, C.G.; Bennetzen, J.L. Gene Loss and Genome Rearrangement in the Plastids of Five Hemiparasites in the Family Orobanchaceae. *BMC Plant Biol.* **2018**, *18*, 30. [[CrossRef](#)]
21. Oyebanji, O.; Zhang, R.; Chen, S.-Y.; Yi, T.-S. New Insights Into the Plastome Evolution of the Millettoid/Phaseoloid Clade (Papilionoideae, Leguminosae). *Front. Plant Sci.* **2020**, *11*, 151. [[CrossRef](#)]
22. 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](#)]
23. Abdullah; Mehmood, F.; Shahzadi, I.; Ali, Z.; Islam, M.; Naeem, M.; Mirza, B.; Lockhart, P.J.; Ahmed, I.; Waheed, M.T. Correlations among Oligonucleotide Repeats, Nucleotide Substitutions, and Insertion–Deletion Mutations in Chloroplast Genomes of Plant Family Malvaceae. *J. Syst. Evol.* **2021**, *59*, 388–402. [[CrossRef](#)]
24. Liu, Q.; Li, X.; Li, M.; Xu, W.; Schwarzacher, T.; Heslop-Harrison, J.S. Comparative Chloroplast Genome Analyses of Avena: Insights into Evolutionary Dynamics and Phylogeny. *BMC Plant Biol.* **2020**, *20*, 406. [[CrossRef](#)] [[PubMed](#)]
25. Ly, S.N.; Garavito, A.; De Block, P.; Asselman, P.; Guyeux, C.; Charr, J.-C.; Janssens, S.; Mouly, A.; Hamon, P.; Guyot, R. Chloroplast Genomes of Rubiaceae: Comparative Genomics and Molecular Phylogeny in Subfamily Ixoroideae. *PLoS ONE* **2020**, *15*, e0232295. [[CrossRef](#)] [[PubMed](#)]
26. Amenu, S.G.; Wei, N.; Wu, L.; Oyebanji, O.; Hu, G.; Zhou, Y.; Wang, Q. Phylogenomic and Comparative Analyses of Coffeaeae Alliance (Rubiaceae): Deep Insights into Phylogenetic Relationships and Plastome Evolution. *BMC Plant Biol.* **2022**, *22*, 88. [[CrossRef](#)] [[PubMed](#)]
27. Varshney, R.K.; Sigmund, R.; Börner, A.; Korzun, V.; Stein, N.; Sorrells, M.E.; Langridge, P.; Graner, A. Interspecific Transferability and Comparative Mapping of Barley EST-SSR Markers in Wheat, Rye and Rice. *Plant Sci.* **2005**, *168*, 195–202. [[CrossRef](#)]
28. Dong, W.; Liu, H.; Xu, C.; Zuo, Y.; Chen, Z.; Zhou, S. A Chloroplast Genomic Strategy for Designing Taxon Specific DNA Mini-Barcodes: A Case Study on Ginsengs. *BMC Genet.* **2014**, *15*, 138. [[CrossRef](#)]
29. Provan, J.; Powell, W.; Hollingsworth, P.M. Chloroplast Microsatellites: New Tools for Studies in Plant Ecology and Evolution. *Trends Ecol. Evol.* **2001**, *16*, 142–147. [[CrossRef](#)] [[PubMed](#)]
30. Yang, C.-H.; Liu, X.; Cui, Y.-X.; Nie, L.-P.; Lin, Y.-L.; Wei, X.-P.; Wang, Y.; Yao, H. Molecular Structure and Phylogenetic Analyses of the Complete Chloroplast Genomes of Three Original Species of Pyrrosiae Folium. *Chin. J. Nat. Med.* **2020**, *18*, 573–581. [[CrossRef](#)]
31. Zhou, M.; Guo, J.; Cha, J.; Chae, M.; Chen, S.; Barral, J.M.; Sachs, M.S.; Liu, Y. Non-Optimal Codon Usage Affects Expression, Structure and Function of Clock Protein FRQ. *Nature* **2013**, *495*, 111–115. [[CrossRef](#)]

32. Somaratne, Y.; Guan, D.-L.; Wang, W.-Q.; Zhao, L.; Xu, S.-Q. The Complete Chloroplast Genomes of Two *Lespedeza* Species: Insights into Codon Usage Bias, RNA Editing Sites, and Phylogenetic Relationships in Desmodieae (Fabaceae: Papilionoideae). *Plants* **2020**, *9*, 51. [[CrossRef](#)]
33. Lyu, X.; Liu, Y. Nonoptimal Codon Usage Is Critical for Protein Structure and Function of the Master General Amino Acid Control Regulator CPC-1. *Mbio* **2020**, *11*, e02605-20. [[CrossRef](#)]
34. Ding, S.; Dong, X.; Yang, J.; Guo, C.; Cao, B.; Guo, Y.; Hu, G. Complete Chloroplast Genome of *Clethra Fargesii* Franch., an Original Sympetalous Plant from Central China: Comparative Analysis, Adaptive Evolution, and Phylogenetic Relationships. *Forests* **2021**, *12*, 441. [[CrossRef](#)]
35. 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. Int. J. Rapid Publ. Rep. Genes Genomes* **2004**, *11*, 247–261. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, H.; Li, C.; Miao, H.; Xiong, S. Insights from the Complete Chloroplast Genome into the Evolution of *Sesamum indicum* L. *PLoS ONE* **2013**, *8*, e80508. [[CrossRef](#)]
37. De Las Rivas, J.; Lozano, J.J.; Ortiz, A.R. Comparative Analysis of Chloroplast Genomes: Functional Annotation, Genome-Based Phylogeny, and Deduced Evolutionary Patterns. *Genome Res.* **2002**, *12*, 567–583. [[CrossRef](#)]
38. Moore, M.J.; Bell, C.D.; Soltis, P.S.; Soltis, D.E. Using Plastid Genome-Scale Data to Resolve Enigmatic Relationships among Basal Angiosperms. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19363–19368. [[CrossRef](#)] [[PubMed](#)]
39. Razafimandimbison, S.G.; Rydin, C. Molecular-Based Assessments of Tribal and Generic Limits and Relationships in Rubiaceae (Gentianales): Polyphyly of Pomazoteae and Paraphyly of Ophiorrhizeae and Ophiorrhiza. *Taxon* **2019**, *68*, 72–91. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.