

Article

Structural and Comparative Analysis of the Complete Chloroplast Genome of a Mangrove Plant: *Scyphiphora hydrophyllacea* Gaertn. f. and Related Rubiaceae Species

Ying Zhang ^{1,*}, Jing-Wen Zhang ², Yong Yang ² and Xin-Nian Li ¹

¹ Life Science and Technology School, Lingnan Normal University, Zhanjiang 524048, China; lixinn@foxmail.com

² Ministry of Education Key Laboratory for Ecology of Tropical Islands, Hainan Normal University, Haikou 570100, China; Zhangjingwen19114@126.com (J.-W.Z.); yybiology@163.com (Y.Y.)

* Correspondence: Zhangyingred@yahoo.com; Tel.: +86-07593182545

Received: 31 August 2019; Accepted: 1 November 2019; Published: 8 November 2019



Abstract: *Scyphiphora hydrophyllacea* Gaertn. f. (Rubiaceae) is an endangered mangrove species found in China, and its only known location is in Hainan Island. Previous studies conducted on *S. hydrophyllacea* have mainly focused on its location, biological characteristics, and medical effects. However, to date, there has been no published report regarding the genetics or genome of this endangered mangrove species. In this study, we developed valuable chloroplast genome-related molecular resources of *S. hydrophyllacea* by comparing with it related Rubiaceae species. The chloroplast genome of *S. hydrophyllacea* was found to be a circular molecule with a total size of 155,132 bp, and it is observed to have a quadripartite structure. The whole chloroplast genome contains 132 genes, of which 88 and 36 are protein-coding and transfer RNA genes, respectively; it also contains four ribosomal RNA genes with an overall GC content of 37.60%. A total of 52 microsatellites were detected in the *S. hydrophyllacea* chloroplast genome, and microsatellite marker detection identified A/T mononucleotides as majority simple sequence repeats in all nine Rubiaceae chloroplast genomes. Comparative analyses of these nine chloroplast genomes revealed variable regions, including *matK*, *rps16*, and *atpF*. All nine species shared 13 RNA-editing sites distributed across eight coding genes. Phylogenetic analyses based on the complete sequences of the chloroplast genomes revealed that the position of *S. hydrophyllacea* is closer to the *Coffeae* genus than to *Cinchoneae*, *Naucleaeae*, *Morindeae*, and *Rubieae* in the Rubiaceae family. The genome information reported in this study could find further application in the evolution and population genetic studies, and it helps improve our understanding of the endangered mechanism and the development of conservation strategies of this endangered mangrove plant.

Keywords: endangered mangrove; *Scyphiphora hydrophyllacea*; Rubiaceae; Gentianales; chloroplast genome; phylogenetic analysis

1. Introduction

Scyphiphora hydrophyllacea is a shrub mangrove that belongs to the *Scyphiphora* genus (family: Rubiaceae), a monotypic genus whose distribution range extends south India and as well as Ceylon, Indochina, and Hainan in China through the Malay Archipelago and Philippines to Australia and New Caledonia and northward to the Solomon Islands and Palau [1]. Its distinguishing characteristics include rounded glossy leaves, fringed stipules, small white flowers, and eight-ribbed drupe-like fruits. Its terminal nodes and shoots are also distinctively covered by a resinous substance [2]. This species is

often located along the high intertidal zones of the midestuarine reaches where it grows in pockets of scattered isolated shrubs and is often regarded as a minor constituent of the mangrove habitat. Based on the categories and criteria of the International Union for the Conservation of Nature Red List of Threatened Species, *S. hydrophyllacea* has been classified into the Least Concern category, which has a global loss of 20% [3]. In China, this species is found only in Hainan Island, and it has been included in a list of key wild plants under provincial protection in Hainan (2006). Although the incidences of fruit set in this mangrove is high, only a low percentage of seed germination has been reported [1]. Today, *S. hydrophyllacea* continues to be regarded as an important medical plant because of its medical properties such as antihepatocarcinogenic and antioxidant effects [4]. Several phytochemicals including flavonoids, terpinoids, and iriddoids have been reported in this species [5]. However, so far, no studies have been conducted to investigate its genetic background. The chloroplast (cp) genome of *S. hydrophyllacea* reported in this paper provides valuable information for further studies of the cp molecular biology of the species. These data will also promote work on genetic breeding and germplasm protective research, and are projected to help clarify the molecular evolution status of *S. hydrophyllacea* in Rubiaceae.

Cp are well known as the main site of photosynthesis, which is a process that provides the energy required for the synthesis of glucose, important fatty acids, starch, and pigments [6]. The size of 120–210 kb is typical for a higher plant cp genome, which usually encodes 120–140 genes. The typical quadripartite structure of the cp genome consist of a small single-copy region (SSC), a large single-copy region (LSC), and two inverted repeat (IR) regions [7]. Chloroplasts are independent genetic systems with a highly conserved genomic structure. Unlike the nuclear genome, cp DNA has the characteristics of multiple copies, low molecular weight, and a simple structure, which is considered to be beneficial and is rather conservative [8]. With ongoing developments in DNA sequence technologies, and a booming increase in the number of researchers focused on cp genome research, approximately 3621 plant cp genomes are now publicly available in the National Center for Biotechnology Information (NCBI) database. Rubiaceae is one of the largest families of angiosperms and consists of approximately 600 genera and more than 10,000 species [9], and yet only a few cp genome sequences are registered in the NCBI. In Rubiaceae, only two species—*S. hydrophyllacea* and *Rustia occidentalis*—belong to mangrove plants. In keeping with the important role that cp plays in the salt tolerance of higher plants [10], the whole cp genome information will provide the molecular data to explain this mangrove adaptation for tidal habitat.

2. Materials and Methods

2.1. Plant Material, DNA Extraction, and Sequencing

Mature and healthy leaves of *S. hydrophyllacea* from Sanya (18°13′21.09″N, 109°36′59.73″E), Hainan, China, were collected and then preserved in ice for further study. The corresponding voucher specimens of *S. hydrophyllacea* were deposited at the Hainan Normal University herbarium (BHM-001). The total DNA of leaves was extracted by using a plant DNA extraction Kit (Tiangen, Beijing, China) and following the manufacturer's instructions. The Illumina HiSeq platform was used to sequence the total DNA, which was carried out by a genome sequencing company (TGS, Shenzhen, China). The average 350-bp paired-end library was manufactured and sequenced using the Illumina Genome Analyzer (Hiseq PE150, Shenzhen, China).

2.2. Genomic Assembly, Annotation and Validation

To evaluate the quality of sequenced raw reads, the software FastQC (0.11.7) was used. Then, the cp genome related reads were filtered by mapping all the raw reads to the published cp genome sequences in Rubiaceae. The SPAdes (3.9.0) software was used to assemble the contig sequence [11]. All the transfer RNA sequences were verified using the software tRNAscan-SE version (2.0) [12]. Then, the Ribosome RNA sequences were analyzed with RNAmmer 1.2 Server. For the annotation of the

S. hydrophyllacea cp genome, the DOGMA program was used [13]. Furthermore, the annotation results were checked manually, and then the codon positions were also adjusted via comparison to homologs from other cp genomes in Rubiaceae. The structural features of the *S. hydrophyllacea* cp genome were illustrated using the software OGDRAW (1.3.1) [14].

2.3. Simple Sequence Repeat Analysis

Cp simple sequence repeats (SSRs) in nine cp genomes of Rubiaceae (including *S. hydrophyllacea*) were detected using MISA (<http://pgrc.ipk-gatersleben.de/misa/>, accessed on 7 April 2017). The parameters were set as follows: The minimum numbers of repeats for mononucleotide, dinucleotides, trinucleotides, tetranucleotides, pentanucleotide, and hexanucleotides were 10, 5, 4, 3, 3, and 3, respectively [15].

2.4. Codon Usage Analysis

The codon usage of the *S. hydrophyllacea* cp genome was analyzed using codonW software. The following conditions were used to minimize deviation in the results: (1) the length of every sequence coding for amino acids in protein must be more than 300 nucleotides (nt); and (2) repeat sequences were removed [15]. Possible RNA-editing sites in the *S. hydrophyllacea* protein-coding genes were predicted using the program predictive RNA editor for plants (PREPSuite) with the cutoff value set to 0.8 [16].

2.5. Genome Comparison

Nine cp genome sequences of Rubiaceae including *S. hydrophyllacea* (MN390972), *Coffea arabica* (NC_008535.1), *Coffea canephora* (NC_030053.1), *Emmenopterys henryi* (NC_036300.1), *Galium aparine* (NC_036969.1), *Galium mollugo* (NC_036970.1), *Gynochthodes nanlingensis* (NC_028614.1), *Mitragyna officinalis* (NC_028009.1), and *Mitragyna speciosa* (NC_034698.1) were used. We selected the shuffle-LAGAN mode in the mVISTA software to compare the variation in nine Rubiaceae cp genomes [17]. The borders between single-copy regions (LSC and SSC) and IR regions were compared among nine Rubiaceae cp genomes using the IRscope software [18].

2.6. Phylogenetic Analysis

To gain an insight into the position of *Scyphiphora* in Gentianales and in an attempt to hypothesize when changes have taken place between/among the species and major clades, 40 cp genomes of Gentianales (data present in NCBI on 20 July 2019), including *S. hydrophyllacea*, were compared with each other. Furthermore, three mangroves species in Combretaceae—*Lumnitzera littorea* (Jack) Voigt, *Lumnitzera racemosa* Willd., and *Laguncularia racemosa* Gaern. f.—were chosen as out-groups. To minimize the overrepresentation of duplicated sequences, one of the IR regions in each plasmid was removed before analyses. Using the software MAFFT v7.427 [19], multiple sequence alignment was performed using default values. The software IQ-TREE v1.6.10 (<http://www.iqtree.org>) was used to select the best model, TVM+F+R3, and build the maximum likelihood tree with default parameters. For the insertion and deletion events analysis in Rubiaceae, multiple sequence alignment was analyzed among 10 cp genomes. According to the annotation results of the *S. hydrophyllacea* cp genome, all compared exons including indel regions were manually extracted using MEGA v6.0. The main indel of the indel length >10 bp was kept [20].

3. Results and Discussion

3.1. Basic Characteristics of cp GENOME of *S. hydrophyllacea*

The typical tetrad structure of the cp genome found in most plants [21] was also found in the *S. hydrophyllacea* cp genome with paired IR sequences encoded in opposite directions and LSC and SSC regions, as shown in Figure 1. The cp genome sequence of *S. hydrophyllacea* was deposited in GenBank

under accession number MN390972. The total cp genome of *S. hydrophyllacea* was 153,132 bp in length, similar to other Rubiaceae cp genomes [22,23]. The LSC region was 85,239 bp, the SSC region was 18,165 bp, and the IR regions were 25,864 bp in the cp genome of *S. hydrophyllacea*.

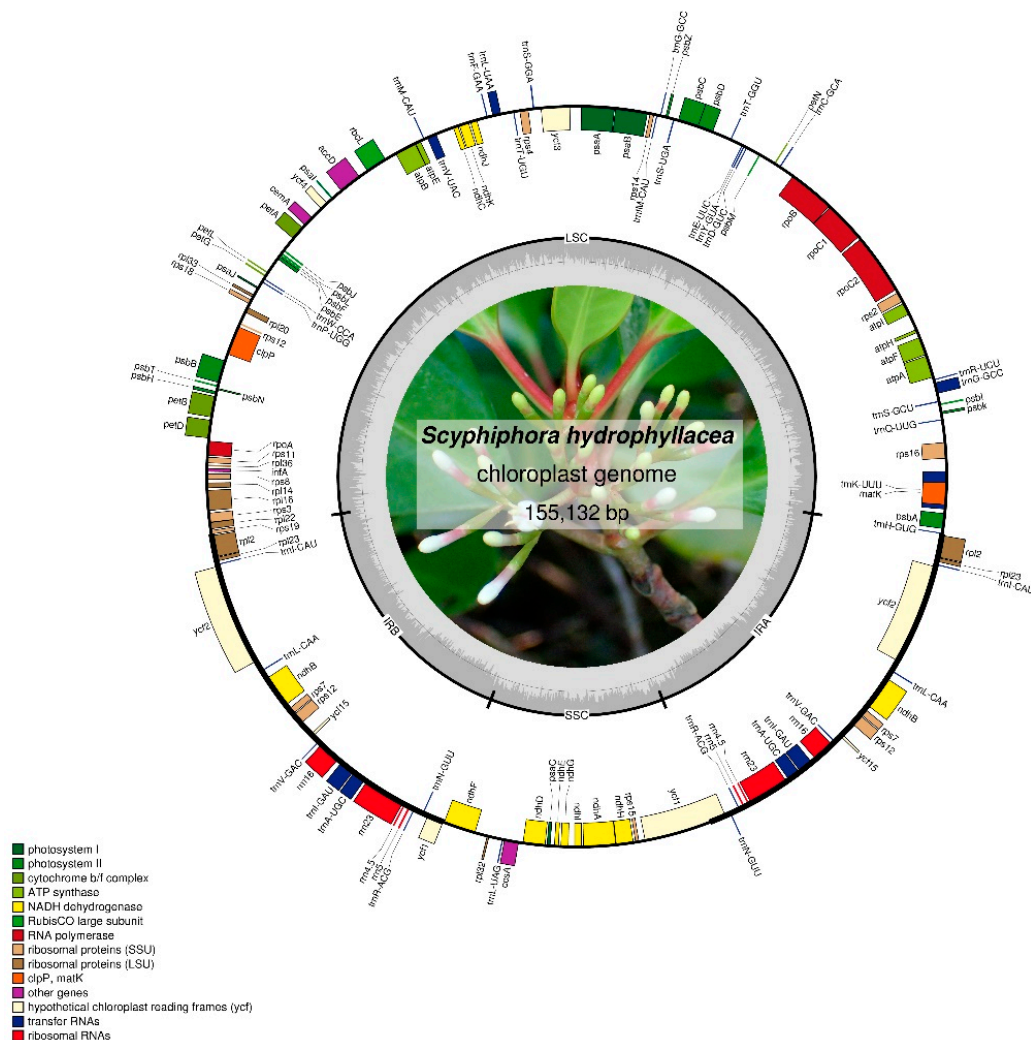


Figure 1. Gene map of the *S. hydrophyllacea* chloroplast genome. The genes drawn outside and inside the outer circle transcribed clockwise and counterclockwise, respectively. Genes of different functional groups are color coded. Guanine and cytosine (GC) content and Adenine and thymine (AT) content are represented on the inner circle by darker gray and lighter gray, respectively.

In the *S. hydrophyllacea* cp genome, a total of 132 genes were found, of which 113 are unique consisting of 80 protein-coding genes, 29 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes (Figure 1, Table 1). Of these, eight protein-coding genes, four rRNAs, and seven tRNAs are found to be duplicated in the IR regions. The protein-coding genes present in the *S. hydrophyllacea* cp genome include nine genes encoding large ribosomal proteins, in which *rpl2* and *rpl23* have two gene copies in IRs and furthermore, *rpl2* has one intron; 12 small ribosomal protein genes; five genes encoding photosystem I components, 15 genes related to photosystem II, and six genes encoding adenosine triphosphate (ATP) synthase and electron transport chain component (Table 1). The gene *rps12* is a trans-spliced gene with its 5' terminal located at the LSC region and the 3' end with a copy located in each of the two IR regions, which is a common phenomenon in higher plants [24]. Similar patterns of protein-coding genes are also present in other Rubiaceae plants [22,23].

Table 1. Genes of the cp genome of *S. hydrophyllacea*. rRNA: Ribosomal RNA.

Functions Category	Group of Genes	Gene Name
Self-replication	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7^a, rps8, rps11, rps12^{acde}, rps14, rps15, rps16, rps18, rps19</i>
	large subunit of ribosome	<i>rpl2^{ab}, rpl14, rpl16, rpl20, rpl22, rpl23^a, rpl32, rpl33, rpl36</i>
	rRNA genes	<i>rrn4.5^a, rrn5^a, rrn16^a, rrn23^a</i>
	DNA-dependent RNA polymerase	<i>rpoA, rpoB, rpoC1^b, rpoC2</i>
	rRNA Genes	<i>trnY-GUA, trnW-CCA, trnV-UAC, trnV-GAC^a, trnT-UGU, trnT-GGU, trnS-UGA, trnS-GGA, trnS-GCU, trnR-UCU, trnR-ACG^a, trnQ-UUG, trnP-UGG, trnN-GUU^a, trnM-CAU, trnL-UAG, trnL-UAA, trnL-CAA^a, trnK-UUU, trnI-GAU^a, trnI-CAU^a, trnH-GUG, trnG-GCC, trnG-UCC, trnfm-CAU, trnF-GAA, trnE-UUC, trnD-GUC, trnC-GCA, trnA-UGC^a</i>
Genes for Photosynthesis	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF^b, atpH, atpI</i>
	Subunits of NADH-dehydrogenase	<i>ndhA^b, ndhB^{ab}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH^b, ndhI, ndhJ, ndhK</i>
	Subunits of cytochrome b/f complexes	<i>petA, petB, petD, petG, petL, petN</i>
	Subunits of photosystem I	<i>psaA, psaB, psaC, psal, psaj</i>
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunits of rubisco	<i>rbcL</i>
Other Genes	Subunits of Acetyl-CoA-carboxylase	<i>accD</i>
	Envelop membrane protein	<i>cemA</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
	Protease	<i>clpP^c</i>
	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>
	Elongation factor	
Genes of Unknown Function	Conserved open reading frames	<i>ycf1^a, ycf2^a, ycf3^c, ycf4, ycf15^a</i>

A—Two gene copies in inverted repeat (IRs); b—Gene containing a single intron; c—Gene containing two introns; d—Pseudogene; e—Gene divided into two independent transcription units.

3.2. SSR Analysis

SSRs are accepted as important molecular markers for population variation studies in higher plants and are usually composed of 1–6 nt [15]. SSRs in the cp genome, similar to those in the nuclear genomes, are highly variable and are often used as genetic markers [25]. In this study, nine cp genome sequences of Rubiaceae plants (including *S. hydrophyllacea*) were used to determine SSR loci using MISA software (Figure 2). A total of 52 microsatellites were identified in the *S. hydrophyllacea* cp genome (Figure 2A, Figure S1, Table S1). Moreover, 43, 38, 46, 66, 67, 64, 54, and 45 SSRs were detected in *C. arabica*, *C. canephora*, *E. henryi*, *G. aparine*, *G. mollugo*, *G. nanlingensis*, *M. speciosa*, and *M. officinalis*, respectively (Figure 2A, Table S1). *G. aparine* (66 SSRs), *G. mollugo* (67 SSRs), and *C. canephora* (38 SSRs) have the highest and lowest number of SSRs, respectively. All SSRs were classified into five types of microsatellites: Mononucleotide, dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide (Figure 2A,B). Consistent with previous reports, most of the SSRs are mononucleotide repeats [26]. In agreement with previous research, the number of mononucleotide repeats is more than the sum of

other types (Figure 2B), and all mononucleotide repeats consist of A or T bases, which is analogous to other land plants [15]. As for the SSR loci, the repeats located in the LSC region are more frequent compared with those in the SSC region and IR regions in all the analyzed Rubiaceae plants (Figure 2C). The frequency of identified SSR motifs in different repeat class types of these nine species are listed in Figure 2D. Mononucleotide A/T showed the highest frequency in all repeats.

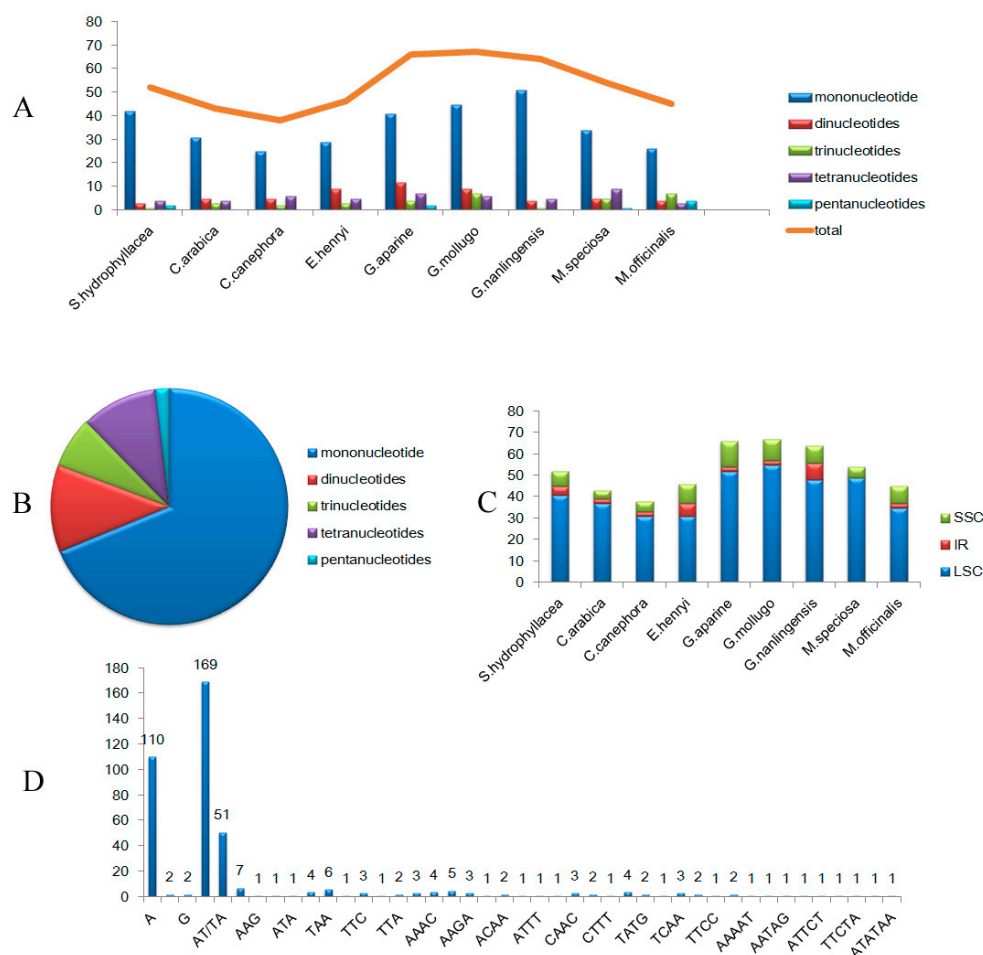


Figure 2. Analysis of simple sequence repeats (SSRs) in nine Rubiaceae (including *S. hydrophyllacea*) chloroplast genome sequences. (A) Number of different SSRs types detected in nine Rubiaceae (including *S. hydrophyllacea*) chloroplast genome sequences. (B) Presence of different SSR types in all SSRs of nine Rubiaceae (including *S. hydrophyllacea*) chloroplast genome sequences. (C) Number of SSRs in the large single-copy (LSC), IR, and small single-copy (SSC) regions in nine Rubiaceae (including *S. hydrophyllacea*) chloroplast genome sequences. (D) Number of identified SSR motifs in different repeat class types.

3.3. Codon Usage and Putative RNA Editing Sites in cp Genes of *S. hydrophyllacea*

In this study, the codon usage frequency and the relative synonymous codon usage (RSCU) in the *S. hydrophyllacea* plastome were analyzed. All protein-coding genes presented a total of 68,907 bp and 22,969 codons in the *S. hydrophyllacea* cp genome. Among all the codons, leucine (Leu) was the most abundant amino acid with a frequency of 10.58%, followed by isoleucine (Ile) with a frequency of 8.61%, whereas cysteine (Cys) was less abundant with a frequency of 1.06% (Figure 3, Table S2 and S3). Leucine and isoleucine are among the more common codons in comparison with other previously reported land plant cp genomes [27,28]. All 19 A/U-ending codons had an RSCU value of >1, whereas two amino acids, methionine (Met) and tryptophan (Trp), with C/G-ending codons had RSCU values of <1 and showed no codon bias. The results for the number of codons (Nc) of each protein-coding gene

ranged from 28.65% (*petN* gene) to 61.00% (*PetG*) (Table S3). The codon usage bias of the cp genome may be caused by selection and mutation [29]; meanwhile, a better understanding of exogenous gene expression and molecular evolution mechanisms of *S. hydrophyllacea* can be gained from further research on codons.

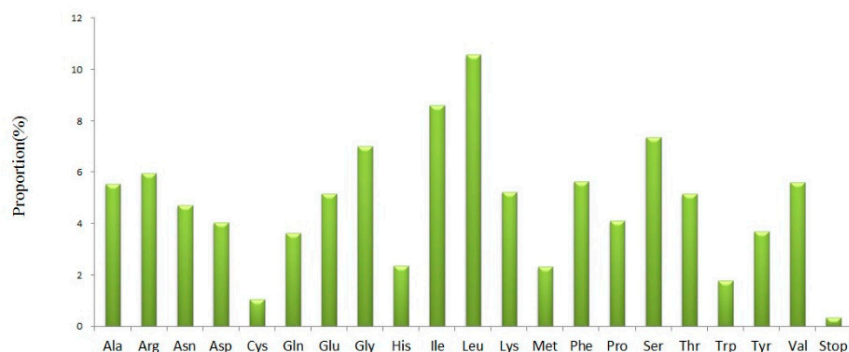


Figure 3. Amino acid frequencies in *S. hydrophyllacea* chloroplast genome protein-coding sequences.

Potential RNA-editing sites in *S. hydrophyllacea* plastome were analyzed using the PREP program, and the results showed that the most frequent conversions at the codon positions consist of serine (Ser) changing to leucine (Leu) (Table 2). A total of 46 editing sites in 18 protein-coding genes were identified, with the *ndhB* and *ndhD* genes having the highest number of predicted RNA-editing sites, which is analogous to other land plants [27]. Furthermore, *rpoB* has four predicted RNA-editing sites, whereas *accD*, *atpA*, *matK*, and *ndhA* have three editing sites. All the RNA-editing conversions in the *S. hydrophyllacea* cp genome resulted in hydrophobic products comprising isoleucine, leucine, tryptophan, tyrosine, valine, methionine, and phenylalanine. These results are also congruent with previous reports, which found that the most RNA-editing sites in higher plants led to amino acid change from polar to apolar and resulted in an increase in protein hydrophobicity [15,29,30].

Table 2. Predicted RNA-editing site in the *S. hydrophyllacea* chloroplast genome.

Gene	Nucleotide Position	Amino Acid Position	Codon Conversion	Effect	Score
<i>accD</i>	280	94	CCC => TCC	P => S	1
	845	282	TCG => TTG	S => L	0.8
	887	296	GCC => GTC	A => V	0.8
<i>atpA</i>	773	258	TCA => TTA	S => L	1
	791	264	CCC => CTC	P => L	1
	914	305	TCA => TTA	S => L	1
<i>matK</i>	292	98	CTT => TTT	L => F	0.86
	454	152	CAT => TAT	H => Y	1
	643	215	CAT => TAT	H => Y	1
<i>ndhA</i>	341	114	TCA => TTA	S => L	1
	566	189	TCA => TTA	S => L	1
	1073	358	TCC => TTC	S => F	1
<i>ndhB</i>	149	50	TCA => TTA	S => L	1
	467	156	CCA => CTA	P => L	1
	586	196	CAT => TAT	H => Y	1
	611	204	TCA => TTA	S => L	0.8
	737	246	CCA => CTA	P => L	1
	746	249	TCT => TTT	S => F	1
	830	277	TCA => TTA	S => L	1
	836	279	TCA => TTA	S => L	1
	1481	494	CCA => CTA	P => L	1
<i>ndhD</i>	29	10	ACG => ATG	T => M	1
	626	209	TCA => TTA	S => L	1
	701	234	TCG => TTG	S => L	1
	905	302	TCA => TTA	S => L	1
	1103	368	GCT => GTT	A => V	1
	1325	442	TCA => TTA	S => L	0.8

Table 2. Cont.

Gene	Nucleotide Position	Amino Acid Position	Codon Conversion	Effect	Score
	1337	446	TCA => TTA	S => L	0.8
<i>ndhF</i>	290	97	TCA => TTA	S => L	1
<i>ndhG</i>	314	105	ACA => ATA	T => I	0.8
<i>petB</i>	418	140	CGG => TGG	R => W	1
	611	204	CCA => CTA	P => L	1
<i>psaI</i>	80	27	TCT => TTT	S => F	0.86
<i>psbE</i>	214	72	CCT => TCT	P => S	1
<i>rpl20</i>	320	107	TCA => TTA	S => L	0.86
<i>rpoB</i>	473	158	TCA => TTA	S => L	0.86
	551	184	TCA => TTA	S => L	1
	566	189	TCG => TTG	S => L	1
	2414	805	TCA => TTA	S => L	0.86
<i>rpoC1</i>	41	14	TCA => TTA	S => L	1
<i>rpoC2</i>	2296	766	CGG => TGG	R => W	1
	3746	1249	TCA => TTA	S => L	0.86
<i>rps14</i>	80	27	TCA => TTA	S => L	1
	149	50	CCA => CTA	P => L	1
<i>rps2</i>	248	83	TCA => TTA	S => L	1
<i>rps8</i>	143	48	GCG => GTG	A => V	1

3.4. Comparison of Basic Characteristics of the cp Genome of Nine Rubiaceae Species

The cp genome of Rubiaceae has the typical circular structure with lengths ranging from 152,712 to 155,600 bp (Table 3), and the cp genome of *G. aparine* has the shortest one. The LSC length of Rubiaceae is 83,594–86,298 bp, with the longest found in *S. hydrophyllacea* and the shortest found in *G. aparine*. The SSC length ranges from 17,054 bp to 18,208 bp, and the IR length varies from 25,594 bp to 26,076 bp. In most cases, the differences in the length of the IRs determine the length differences of the cp genome [31]. However, the largest difference in length was found in the LSC region rather than in the SSC and IR regions among the Rubiaceae cp genomes. The GC content of the Rubiaceae cp genomes was similar, and ranged from 37.18% to 38.52%, in which *G. nanlingensis* has the highest GC content (Table 3). The obtained cp genome of *S. hydrophyllacea* exhibits the typical angiosperm quadripartite structure. Moreover, gene content, order, and GC content were consistent with those of the other members of the Rubiaceae family [22,23,32].

The first discovery of cp RNA-editing in cp came from the maize *rpl2* transcript in 1991, in which an ACG codon changed to a start codon AUG, which was defined as the post-transcriptional modification of pre-RNAs [33]. Comparisons of RNA editing sites among all nine studied Rubiaceae species revealed that *M. officinalis* has highest number of RNA-editing sites (58 in 23 genes), followed by *E. henryi* (58 in 21 genes). Meanwhile, the lowest number of RNA-editing sites is found in *G. aparine* (44 in 19 genes, Table S4). All nine Rubiaceae species shared 13 editing sites distributed in eight genes (Table 4), and the highly conserved RNA-editing sites occurred between genera (Table S4). Even though the most frequent editing events in higher plants are C-to-U/T changes, U/T-to-C editing has also been observed in this research [33]. In *S. hydrophyllacea*, on the other hand, 46 RNA-editing sites were found in 25 genes, all with C-to-T editing. Furthermore, not one U/T-to-C editing in all RNA-editing sites has been found in the other seven Rubiaceae species (Table S4). In all species except for *G. nanlingensis*, the *ndhB* gene was observed to have the highest number of editing sites, followed by the *ndhD* gene. In the *G. nanlingensis* cp genome, there are two editing sites in the *ndhB* gene. At the same time, a notable RNA-editing event was also detected in all nine Rubiaceae species at the initiator codon (ACG), resulting in an ATG translational start codon in the *ndhD* gene, which is analogous to several other plants [27,33]. For the *ycf3* gene, one editing site is found in both *Galium* species, and no editing site was found in the other seven tested cp genomes.

Table 3. Comparison of the basic characteristics of the chloroplast genome in nine Rubiaceae species. tRNA: Transfer RNA.

	<i>Scyphiphora hydrophyllacea</i>	<i>Coffea arabica</i>	<i>Coffea canephora</i>	<i>Emmenopterys henryi</i>	<i>Galium aparine</i>	<i>Galium mollugo</i>	<i>Gynochthodes nanlingensis</i>	<i>Mitragyna officinalis</i>	<i>Mitragyna speciosa</i>
Length (bp)	153,132	155,189	154,751	155,379	152,712	153,677	154,086	153,398	155,600
GC content (%)	37.60	37.43	37.47	37.64	37.28	37.18	38.52	38.05	37.52
AT content (%)	62.40	62.57	62.53	62.36	62.72	62.82	61.48	61.95	62.48
LSC length (bp)	85,239	85,166	84,850	85,554	83,594	84,471	84,329	84,302	86,298
SSC length (bp)	18,165	18,137	18,133	18,245	17,054	17,054	18,113	17,562	18,114
IR length (bp)	25,864	25,943	25,884	25,790	26,032	26,076	25,822	25,767	25,594
Gene number	132	133	133	132	132	131	133	133	131
Pseudogene number	1	2	2	1	3	1	2	1	1
Gene number in IR regions	19	16	17	17	16	16	18	18	16
Protein-coding gene number	88	85	86	86	84	85	90	91	85
Protein-coding gene (%)	66.67	63.91	64.66	65.15	63.64	64.89	67.67	68.42	64.89
rRNA gene number	8	8	8	8	8	8	8	8	8
rRNA (%)	6.06	6.02	6.02	6.06	6.06	6.11	6.02	6.02	6.11
tRNA gene number	36	38	37	37	37	37	33	33	37
tRNA (%)	27.27	28.57	27.82	28.03	28.03	28.24	24.81	24.81	28.24

Table 4. List of RNA-editing sites shared by the nine plastomes predicted by the PREP program.

[illegible]

To compare the sequence variation between species, an alignment of nine Rubiaceae species plastid genome sequences was carried out using the mVISTA program (Figure 4). Overall, the comparative genomic analysis showed that nine Rubiaceae cp genomes were relatively conserved. In agreement with similar studies in other plants, the IR region appeared to be more conserved than the LSC and SSC regions [15,34]. The noncoding regions appeared to be more variable globally than the coding regions in the cp genomes of Rubiaceae species. In all nine cp genome sequences, some highly divergent regions, including *matK*, *rps16*, *atpF*, *psaB*, *ycf3*, *psbH*, *petD*, *rpl16*, *rpl22*, *ndhF*, and *ccsA* were identified, which might be used as a source of potential molecular markers for Rubiaceae plants. However, further work is necessary to verify the suitability of these potential molecular markers for the phylogenetic studies of Rubiaceae.

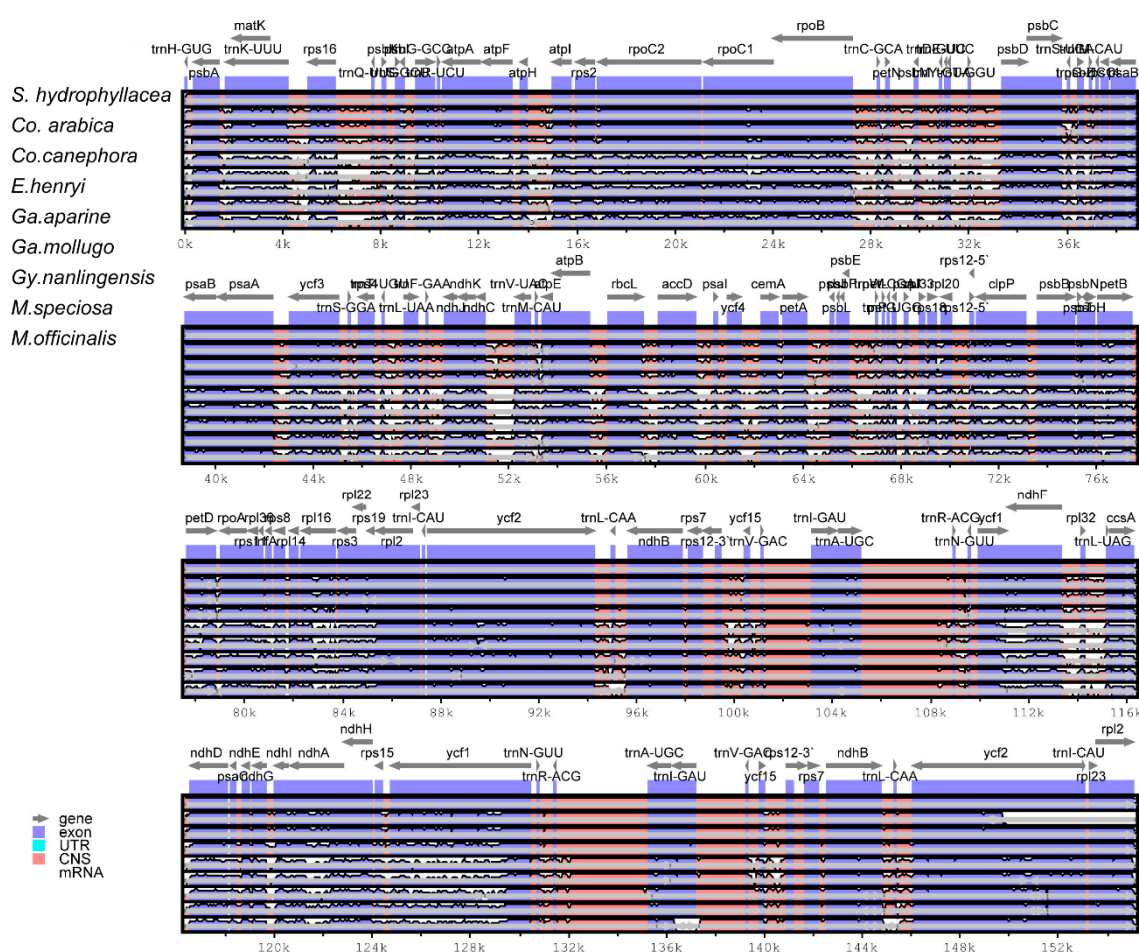


Figure 4. Sequence alignment of nine Rubiaceae species chloroplast genomes, with *S. hydrophyllacea* as the reference. The y-axis indicates the percent of identity between 50% and 100%. Genome regions are color-coded as protein-coding regions, rRNA coding regions, tRNA coding regions, and conserved noncoding sequences.

The IR region is always considered to be consistent and stable in the cp genome, and is also common in plant evolution with the events of border region contraction or expansion. In this study, the IR boundaries of the *S. hydrophyllacea* cp genome were analyzed and compared with those of the other eight Rubiaceae species (Figure 5). The events of expansion or contraction within the border regions between the two IR regions and the single-copy regions are considered to contribute to the genome size variations among plant lineages [35]. According to our research, IR regions are more conservative than LSC and SSC regions in the cp genomes of Rubiaceae. Although there are still expansion or contraction events in IR regions observed among the studied representatives of Rubiaceae, they contributed little to the observed differences in the overall size of the cp genomes. Interestingly, *C. canephora* showed

obvious differences compared with the other eight Rubiaceae species with the *rpl2* gene in LSC/IR, which was found in the IR region in other eight species. The location of *ycf2* in the SSC/IR region was replaced by *ycf1* in the other eight cp genomes (Figure 5).

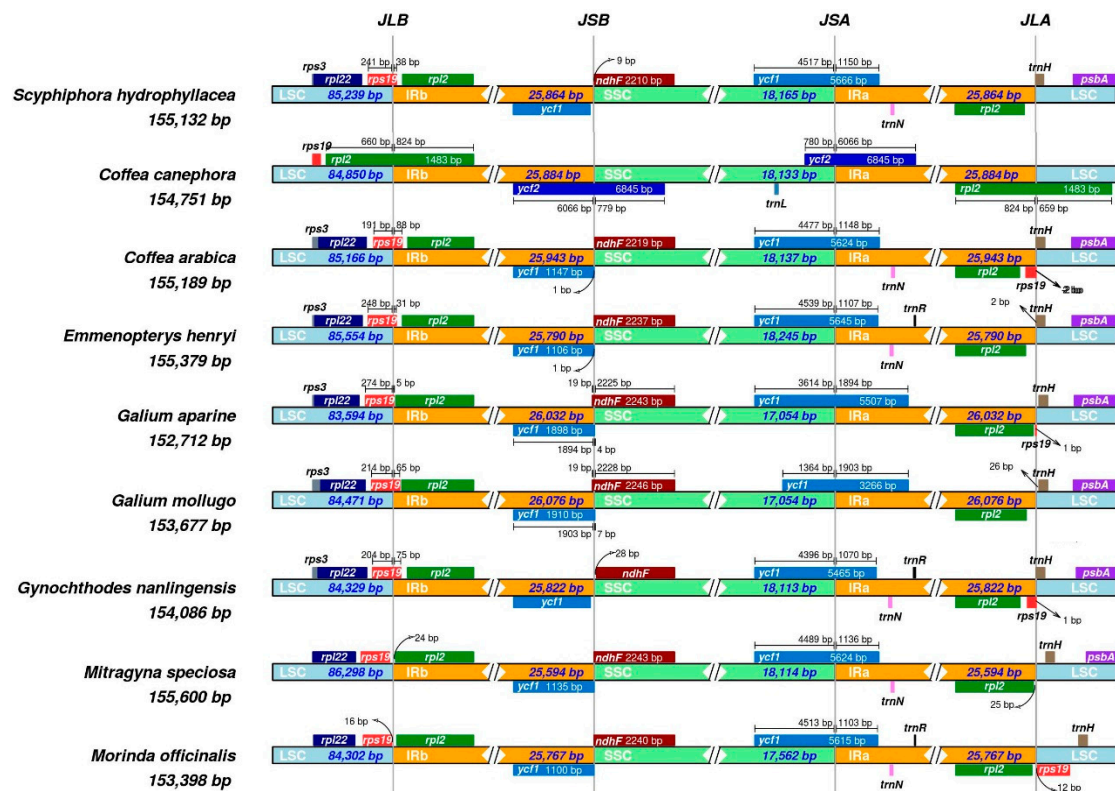


Figure 5. IR contraction/expansion analysis of nine Rubiaceae species. JLB (LSC/IRb), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC) denote the junction sites between each of the corresponding two regions on the genome.

3.5. Phylogenetic Relationships in Gentianales

The alignment of complete plastid genome sequences resulted in a well-resolved phylogenetic topology of 40 Gentianales taxa (Figure 6). In general, species representing Rubiaceae, Gentianaceae, Apocynaceae, and Asclepiadaceae were clustered into three groups. Furthermore, Apocynaceae and Asclepiadaceae had the nearest distance and were clustered into one group. Six subfamilies were clustered from those 10 Rubiaceae species (*C. arabica*, *C. canephora*, *S. hydrophyllacea*, *E. henryi*, *D. sinensis*, *M. speciosa*, *M. officinalis*, *N. cadamba*, *G. aparine*, and *G. mollugo*), and the position of *S. hydrophyllacea* appeared to be closer to the genus *Coffea* than to species representing Cinchoneae, Naucleaeae, Morindeae, and Rubieae. Previous research discovered the tribal and generic relationships in Rubiaceae via analyses of morphology, nuclear, ribosomal internal transcribed spacer (ITS), the restriction sites of cpDNA, and single chloroplast gene (*rbcL*) [9,36]. Four species (*M. officinalis*, *E. henryi*, *C. arabica*, and *C. canephora*) in Rubiaceae based on whole protein-coding genes of the cp genome were used to evaluate the phylogenetic relationships in Gentianales plants [32]. Furthermore, the closely related phylogenetic relationships of Rubiaceae plants (two *Coffea* species) with an out-group plant such as the Solanaceae family was also analyzed using Conserved Ortholog Set II makers [37] or chloroplast genes [22]. Phylogenetic analyses based on the complete plastid genome sequence instead of a few genes have been conducted in several high land plant species [20]. Our phylogenetic analyses resolved similar topologies, which confirm the results of previous phylogenetic analyses in Rubiaceae based on fewer genes [9].

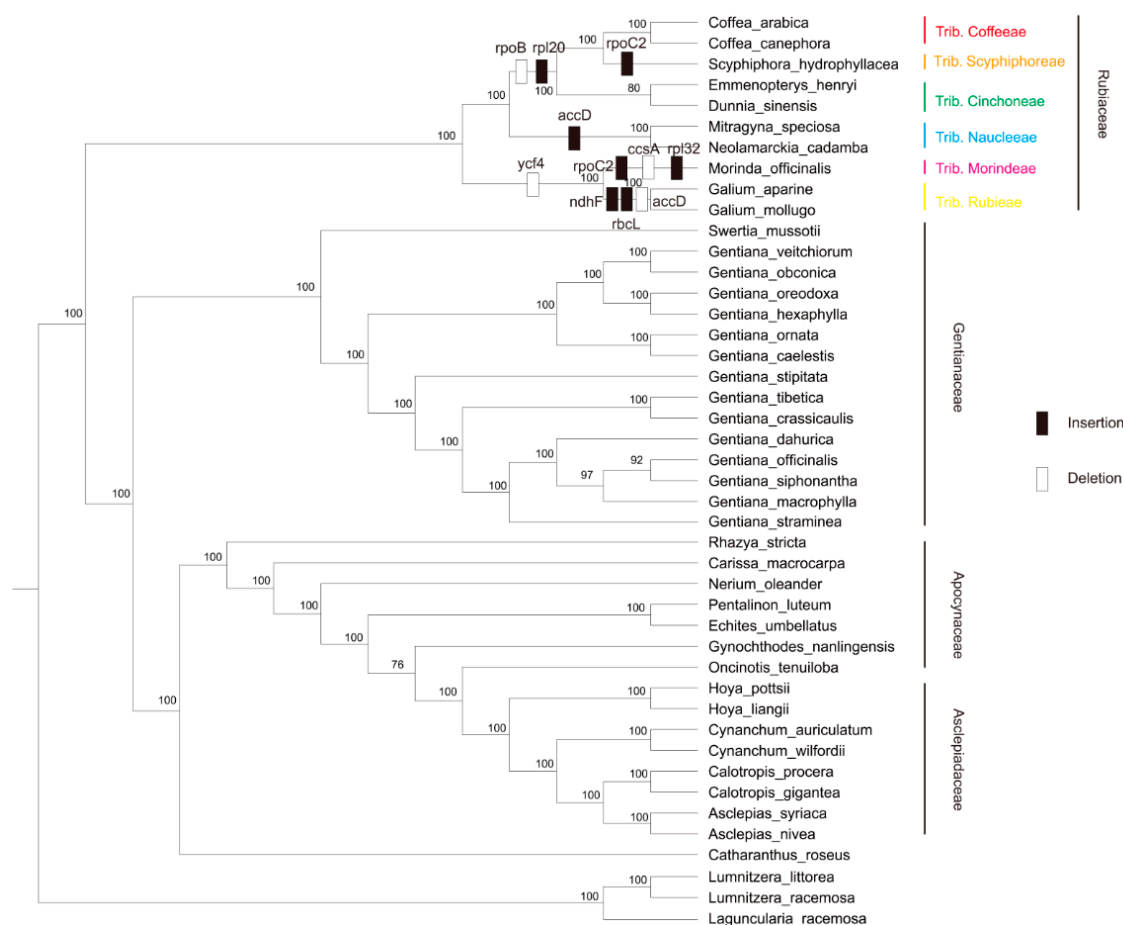


Figure 6. Cladogram illustrating the phylogenetic relationships of Gentianales based on complete chloroplast genome sequences. Currently recognized suprageneric groups are listed on the right.

According to the gene annotation of *S. hydrophyllacea*, the exons including the indel region were analyzed using MEGA, and insertion events were found in the following genes: *rpoC2*, *rpl20*, *rpl32*, *ndhF*, and *rbcL*. Deletion events were also found in *rpoB*, *accD*, *ccsA*, and *ycf4* (Figure 6). The insertion and deletion sequence lengths in each species are listed in Table S5. The *accD* gene encoding one of the four subunits of the acetyl-CoA carboxylase enzyme in most cps showed insertion events in Nicotianoideae and Solanoideae plants, which is regarded as a possible ancestral trait of these species [33]. In this research, the deletion events in the *accD* gene were found in *M. speciosa*, *N. cadamba*, and two *Galium* species. Furthermore, no insertion events in the *accD* gene were found in all the Rubiaceae species investigated in this study.

4. Conclusions

The complete cp genome sequence obtained from one endangered mangrove, *S. hydrophyllacea*, was compared with that of eight other Rubiaceae cp genomes. The cp genomes of those Rubiaceae species have undergone evolution at the gene level rather than the genome level, because no significant structural changes were found. The IR/SSC and IR/LSC junctions are relatively conservative in Rubiaceae except for *C. canephora*. Eleven cp DNA markers were developed from the relatively highly variable regions, which may be used for further studies that focus on the identification of markers. All the chosen Rubiaceae taxa were completely distinguished with high bootstrap support based on the whole cp genome sequences. Gene insertion events in five genes and deletion events in four genes were found in Rubiaceae cp genomes. The data presented in this study will help improve our understanding of the evolutionary history of Gentianales. The availability of this cp genome sequence

will serve as a tool to advance the study of protection in *S. hydrophyllacea* and help researchers explore the endangered mechanism of and genetic questions about this species.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1999-4907/10/11/1000/s1>, Table S1: SSRs in chloroplast genomes of nine Rubiaceae plants, Table S2: Codon usage in protein-coding genes from *S. hydrophyllacea*, Table S3: Codon usage for individual protein genes, Table S4: List of RNA-editing sites predicted by the PREP program in the selected chloroplast genome, Table S5: Insertion and deletion events of chloroplast genes in Rubiaceae species. Figure S1. The distribution, type, and presence of microsatellites (SSRs) in the chloroplast genome of *S. hydrophyllacea*. (A) Number of different SSR types; (B) Proportion of SSRs in LSC, SSC, and IR regions; (C) Number of identified SSR motifs in different repeat class types.

Author Contributions: All the authors listed have made substantial, direct, and intellectual contributions to the work, and approved it for publication. Y.Z. analyzed the data and wrote the manuscript; Y.Z., Y.Y. and X.-N.L. conceived the experiments; J.-W.Z. and Y.Y. performed the experiments.

Funding: This work was supported by grants from the Hainan Natural Science Foundation (Grant No.318MS176) and the National Natural Scientific Foundation of China (Grant No.41776148 and 31760119).

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

References

- Tomlinson, P. *The Botany of Mangroves*, 1st ed.; Cambridge University Press: New York, NY, USA, 1986; pp. 261–264.
- Duke, N.C. *Australia's Mangroves: The Authoritative Guide to Australia's Mangrove Plants*, 1st ed.; University of Queensland Press: Brisbane, QLD, Australia, 2006; pp. 172–173.
- Polidoro, B.A.; Carpenter, K.E.; Collins, L.; Duke, N.C.; Ellison, A.M.; Ellison, J.C.; Farnsworth, E.J.; Fernando, E.S.; Kathiresan, K.; Koedam, N.E.; et al. The loss of species: Mangrove extinction risk and geographic areas of global concern. *PLoS ONE* **2010**, *5*, e10095. [[CrossRef](#)] [[PubMed](#)]
- Samarakoon, S.R.; Shanmuganathan, C.; Ediriweera, M.K.; Piyathilaka, P.; Tennekoon, K.H.; Thabrew, I.; Galhena, P.; De Silva, E.D. Anti-hepatocarcinogenic and Anti-oxidant Effects of Mangrove Plant *Scyphiphora hydrophyllacea*. *Pharm. Mag.* **2017**, *13*, S76–S83. [[CrossRef](#)] [[PubMed](#)]
- Feng, C.L.; Gong, M.F.; Zeng, Y.B.; Dai, H.F.; Mei, W.L. Scyphiphin C, a new iridoid from *Scyphiphora Hydrophyllacea*. *Molecules* **2010**, *15*, 2473–2477. [[CrossRef](#)] [[PubMed](#)]
- Neuhaus, H.E.; Emes, M.J. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Biol.* **2000**, *51*, 111–140. [[CrossRef](#)]
- 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](#)]
- Li, Y.; Zhang, J.; Li, L.; Gao, L.; Xu, J. Structural and Comparative Analysis of the Complete Chloroplast Genome of *Pyrus hopeiensis*—“Wild Plants with a Tiny Population”—and Three Other *Pyrus* Species. *Int. J. Mol. Sci.* **2018**, *19*, 3262. [[CrossRef](#)]
- Andreasen, K.; Bremer, B. Combined phylogenetic analysis in the Rubiaceae-Ixoroideae: Morphology, nuclear and chloroplast DNA data. *Am. J. Bot.* **2000**, *87*, 1731–1748. [[CrossRef](#)]
- Bejaoui, F.; Salas, J.J.; Nouairi, I.; Smaoui, A.; Abdelly, C.; Martinez-Force, E.; Youssef, N.B. Changes in chloroplast lipid contents and chloroplast ultrastructure in *Sulla carnosa* and *Sulla coronaria* leaves under salt stress. *J. Plant Physiol.* **2016**, *198*, 32–38. [[CrossRef](#)]
- Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: Single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)]
- Lowe, T.M.; Chan, P.P. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57. [[CrossRef](#)]
- Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [[CrossRef](#)] [[PubMed](#)]
- Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [[CrossRef](#)] [[PubMed](#)]
- Liu, H.Y.; Yu, Y.; Deng, Y.Q.; Li, J.; Huang, Z.X.; Zhou, S.D. The Chloroplast Genome of *Lilium henrici*: Genome Structure and Comparative Analysis. *Molecules* **2018**, *23*, 1276. [[CrossRef](#)] [[PubMed](#)]

16. Mower, J.P. The PREP suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* **2009**, *37*, W253–W259. [[CrossRef](#)] [[PubMed](#)]
17. 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](#)] [[PubMed](#)]
18. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [[CrossRef](#)]
19. Nakamura, T.; Yamada, K.D.; Tomii, K.; Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **2018**, *34*, 2490–2492. [[CrossRef](#)]
20. Amiryousefi, A.; Hyvonen, J.; Poczai, P. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae. *PLoS ONE* **2018**, *13*, e0196069. [[CrossRef](#)]
21. Xu, C.; Dong, W.; Li, W.; Lu, Y.; Xie, X.; Jin, X.; Shi, J.; He, K.; Suo, Z. Comparative Analysis of Six Lagerstroemia Complete Chloroplast Genomes. *Front. Plant Sci.* **2017**, *8*, 15. [[CrossRef](#)]
22. Samson, N.; Bausher, M.G.; Lee, S.B.; Jansen, R.K.; Daniell, H. The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: Organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. *Plant Biotechnol. J.* **2007**, *5*, 339–353. [[CrossRef](#)]
23. Zhang, R.; Li, Q.; Gao, J.; Qu, M.; Ding, P. The complete chloroplast genome sequence of the medicinal plant *Morinda officinalis* (Rubiaceae), an endemic to China. *Mitochondrial DNA Part A* **2016**, *27*, 4324–4325. [[CrossRef](#)] [[PubMed](#)]
24. Hildebrand, M.B.; Hallick, R.; Passavant, C.; Bourque, D. Trans-Splicing in Chloroplasts: The rps12 Loci of *Nicotiana tabacum*. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 372–376. [[CrossRef](#)] [[PubMed](#)]
25. Fu, P.C.; Zhang, Y.Z.; Geng, H.M.; Chen, S.L. The complete chloroplast genome sequence of *Gentiana lawrencei* var. *farreri* (Gentianaceae) and comparative analysis with its congeneric species. *PeerJ* **2016**, *4*, e2540. [[CrossRef](#)] [[PubMed](#)]
26. George, B.; Bhatt, B.S.; Awasthi, M.; George, B.; Singh, A.K. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr. Genet.* **2015**, *61*, 665–677. [[CrossRef](#)]
27. Saina, J.K.; Li, Z.Z.; Gichira, A.W.; Liao, Y.Y. The Complete Chloroplast Genome Sequence of Tree of Heaven (*Ailanthus altissima* (Mill.) (Sapindales: Simaroubaceae), an Important Pantropical Tree. *Int. J. Mol. Sci.* **2018**, *19*, 929. [[CrossRef](#)]
28. Yang, Y.; Zhu, J.; Feng, L.; Zhou, T.; Bai, G.; Yang, J.; Zhao, G. Plastid Genome Comparative and Phylogenetic Analyses of the Key Genera in Fagaceae: Highlighting the Effect of Codon Composition Bias in Phylogenetic Inference. *Front. Plant Sci.* **2018**, *9*, 82. [[CrossRef](#)]
29. Morton, B.R. The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. *J. Mol. Evol.* **2003**, *56*, 616–629. [[CrossRef](#)]
30. Lopes, A.D.S.; Pacheco, T.G.; Nimz, T.; Vieira, L.D.N.; Guerra, M.P.; Nodari, R.O.; De Souza, E.M.; Pedrosa, F.D.O.; Rogalski, M. The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta* **2018**, *247*, 1011–1030.
31. Guisinger, M.M.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* **2011**, *28*, 583–600. [[CrossRef](#)]
32. Duan, R.Y.; Huang, M.Y.; Yang, L.M.; Liu, Z.W. Characterization of the complete chloroplast genome of *Emmenopterys henryi* (Gentianales: Rubiaceae), an endangered relict tree species endemic to China. *Conserv. Genet. Resour.* **2017**, *9*, 1–3. [[CrossRef](#)]
33. Tsudzuki, T.; Wakasugi, T.; Sugiura, M. Comparative analysis of RNA editing sites in higher plant chloroplasts. *J. Mol. Evol.* **2001**, *53*, 327–332. [[CrossRef](#)]
34. Li, P.; Lu, R.S.; Xu, W.Q.; Ohi-Toma, T.; Cai, M.Q.; Qiu, Y.X.; Cameron, K.M.; Fu, C.X. Comparative Genomics and Phylogenomics of East Asian Tulips (*Amana*, Liliaceae). *Front. Plant Sci.* **2017**, *8*, 451. [[CrossRef](#)] [[PubMed](#)]
35. Dong, W.; Xu, C.; Cheng, T.; Zhou, S. Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. *PLoS ONE* **2013**, *8*, e77965. [[CrossRef](#)] [[PubMed](#)]

36. Tosh, J.; Dessein, S.; Buerki, S.; Groeninckx, I.; Mouly, A.; Bremer, B.; Smets, E.F.; De, B.P. Evolutionary history of the Afro-Madagascan *Ixora* species (Rubiaceae): Species diversification and distribution of key morphological traits inferred from dated molecular phylogenetic trees. *Ann. Bot.* **2013**, *112*, 1723–1742. [[CrossRef](#)]
37. Guyot, R.; Lefebvre-Pautigny, F.; Tranchant-Dubreuil, C.; Rigoreau, M.; Hamon, P.; Leroy, T.; Hamon, S.; Poncet, V.; Crouzillat, D.; Kochko, A.D. Ancestral synteny shared between distantly-related plant species from the asterid (*Coffea canephora* and *Solanum* Sp.) and rosid (*Vitis vinifera*) clades. *BMC Genom.* **2012**, *13*, 103. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).