Article

Comparative Analysis of Six Complete Plastomes of *Tripterospermum* spp.

Xiong-De Tu [1,2,†], Wen-Jun Lin [1,†], Hou-Hua Fu [1], Yi-Zhe Lin [1], Jun Shen [1], Shuai Chen [1], Zhong-Jian Liu [1,2], Ming-He Li [1,2,*] and Shi-Pin Chen [1,2,*]

[1] College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2210455001@fafu.edu.cn (X.-D.T.); linwenjun@fafu.edu.cn (W.-J.L.); fhh202216@163.com (H.-H.F.); linyizhe_0220@163.com (Y.-Z.L.); 17859565356@163.com (J.S.); csssd19@163.com (S.C.); zjliu@fafu.edu.cn (Z.-J.L.)

[2] Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at Landscape Architecture and Arts, Fujian Agriculture and Forestry University, Fuzhou 350002, China

[*] Correspondence: fjalmh@fafu.edu.cn (M.-H.L.); chenshipin@fafu.edu.cn (S.-P.C.)

[†] These authors contributed equally to this work.

Abstract: The *Tripterospermum*, comprising 34 species, is a genus of Gentianaceae. Members of *Tripterospermum* are mostly perennial, entwined herbs with high medicinal value and rich in iridoids, xanthones, flavonoids, and triterpenes. However, our inadequate understanding of the differences in the plastid genome sequences of *Tripterospermum* species has severely hindered the study of their evolution and phylogeny. Therefore, we first analyzed the 86 Gentianae plastid genomes to explore the phylogenetic relationships within the Gentianae subfamily where *Tripterospermum* is located. Then, we analyzed six plastid genomes of *Tripterospermum*, including two newly sequenced plastid genomes and four previously published plastid genomes, to explore the plastid genomes' evolution and phylogenetic relationships in the genus *Tripterospermum*. The *Tripterospermum* plastomes have a quadripartite structure and are between 150,929 and 151,350 bp in size. The plastomes of *Tripterospermum* encoding 134 genes were detected, including 86 protein-coding genes (CDS), 37 transfer RNA (tRNA) genes, eight ribosomal RNA (rRNA) genes, and three pseudogenes (*infA*, *rps19*, and *ycf1*). The result of the comparison shows that the *Tripterospermum* plastomes are very conserved, with the total plastome GC content ranging from 37.70% to 37.79%. In repeat sequence analysis, the number of single nucleotide repeats (A/T) varies among the six *Tripterospermum* species, and the identified main long repeat types are forward and palindromic repeats. The degree of conservation is higher at the SC/IR boundary. The regions with the highest divergence in the CDS and the intergenic region (IGS) are *psaI* and *rrn4.5-rrn5*, respectively. The average pi of the CDS and the IGS are only 0.071% and 0.232%, respectively, indicating that the *Tripterospermum* plastomes are highly conserved. Phylogenetic analysis indicated that Gentianinae is divided into two clades, with *Tripterospermum* as a sister to *Sinogeniana*. Phylogenetic trees based on CDS and CDS + IGS combined matrices have strong support in *Tripterospermum*. These findings contribute to the elucidation of the plastid genome evolution of *Tripterospermum* and provide a foundation for further exploration and resource utilization within this genus.

Keywords: comparative analysis; Gentianinae; plastid genome; phylogenetic analysis; *Tripterospermum*

1. Introduction

Tripterospermum Blume, established by Blume in 1826 with the type species *T. trinerve* from Java, is a genus of Gentianinae (Gentianaceae). There are 34 accepted species of *Tripterospermum*, distributed in East Asia, Southeast Asia, and the adjacent Himalayan region, with the highest diversity in southwest China, and there are more than 20 species [1,2]. The species of *Tripterospermum* are characterized by perennial twining herbs, opposite leaves, and axillary and terminal inflorescences [3]. *Tripterospermum* plants are rich in

iridoids, xanthones, flavonoids, and triterpenes, revealing diverse biological activities, such as antivirus and antihypertension [4–9].

The subtribe Gentianinae, comprising *Crawfurdia* Wallich, *Gentiana* L., *Kuepferia* Adr. Favre, *Metagentiana* T. N. Ho & S. W. Liu, *Tripterospermum* Blume, and *Sinogentiana* Favre & Yuan, encompasses a total of six genera with approximately 425 species. This subtribe stands as one of the most species-rich clades within Gentianaceae [10,11]. Two genera, *Tripterospermum* and *Crawfurdia*, were concurrently described [12,13]. Due to their climbing habits, these genera are frequently confused [14]. Some authors have even considered them a single genus [15–17], while others classified them as *Gentiana* [18,19]. The monophyly of *Tripterospermum* and *Crawfurdia* has been questioned by previous studies [20,21]. Favre et al. [10] employed ITS and a*tpB-rbcL* sequences for nineteen species of *Tripterospermum*, nine of *Crawfurdia*, and eleven of *Metagentiana*. The results indicated that only *Metagentiana* is polyphyletic, whereas *Crawfurdia* and *Tripterosperum* are monophyletic. However, the phylogenetic support within the genus *Tripterospermum* is low. In a subsequent study, Favre et al. [11] utilized ITS, *atpB-rbcL*, and *trnL-trnF* to reconstruct the phylogenetic relationships of Gentianinae. The findings suggested that *Gentiana* sect. *Otophora* is monophyletic, which is more closely related to *Metagentiana* than *Gentiana*. *Metagentiana* is monophyletic when excluding *M. striata* and *M. souliei*. To account for the monophyly of *Gentiana* and *Metagentiana*, new genera were established, *Kuepferia* and *Sinogentiana*, respectively. More recent studies, incorporating ITS, *atpB-rbcL*, and *trnL-trnF*, conducted extensive sampling of Gentianinae, including 30 *Tripterospermum* species, and constructed a phylogenetic tree. The results indicated that *Crawfurdia*, *Gentiana*, *Kuepferia*, *Metagentiana*, *Sinogentiana*, and *Tripterospermum* are all monophyletic, although the phylogenetic support among *Tripterospermum* species remains low [22].

Traditional molecular markers face challenges in resolving phylogenetic relationships at low taxonomic levels, such as *Tripterospermum*, due to the scarcity of informative sites. One effective approach to enhance molecular phylogenetic datasets is to increase the sampling of loci [23]. The plastid genome, characterized by its small size, dense gene content, moderate mutation rate, and abundant site information [24], has proven to be an effective classification tool for seed plants [25–27]. It has also been proven to be a valuable tool for intergenus and intragenus classification in the Gentianaceae [28–31]. In this study, we present two new plastomes of *Tripterospermum*, *T. filicaule* and *T. nienkui*. These newly obtained sequences were compared with four previously described *Tripterospermum* plastomes. Additionally, to further elucidate the phylogenetic relationships within *Tripterospermum* and its subtribe Gentianinae, we conducted a phylogenetic analysis using the published plastid genomes and constructed a phylogenetic tree that demonstrates robust support within *Tripterospermum* and the broader subtribe Gentianinae.

2. Results

2.1. Characteristics of the Plastome

The plastomes of *T. filicaule* and *T. nienkui* follow the conventional quadripartite arrangement, comprising a sizable single-copy (LSC) region, a compact single-copy (SSC) region, and two inverted repeats (IRs) (Figure 1). The plastome lengths of six *Tripterospermum* species ranged from 150,929 bp (*T. luteoviride*) to 151,350 bp (*T. championii*) (Supplementary Table S1). The length of LSC regions ranged from 82,177 bp (*T. luteoviride*) to 82,506 bp (*T. championii*), and the length of the SSC regions ranged from 17,439 bp (*T. nienkui*) to 17,640 bp (*T. championii*). For the IR regions, the sizes ranged from 25,581 bp (*T. japonicum*, *T. membranaceum*, and *T. nienkui*) to 25,602 bp (*T. championii*). The overall GC content of the plastomes ranged from 37.70% to 37.79%, showcasing a noticeable contrast in GC content among the regions. Specifically, the GC content in the IR regions is higher, ranging from 43.42% to 43.47%, compared to the lower GC content observed in the LSC region (35.52% to 35.61%) and SSC region (31.39% to 31.46%) (Supplementary Table S1).
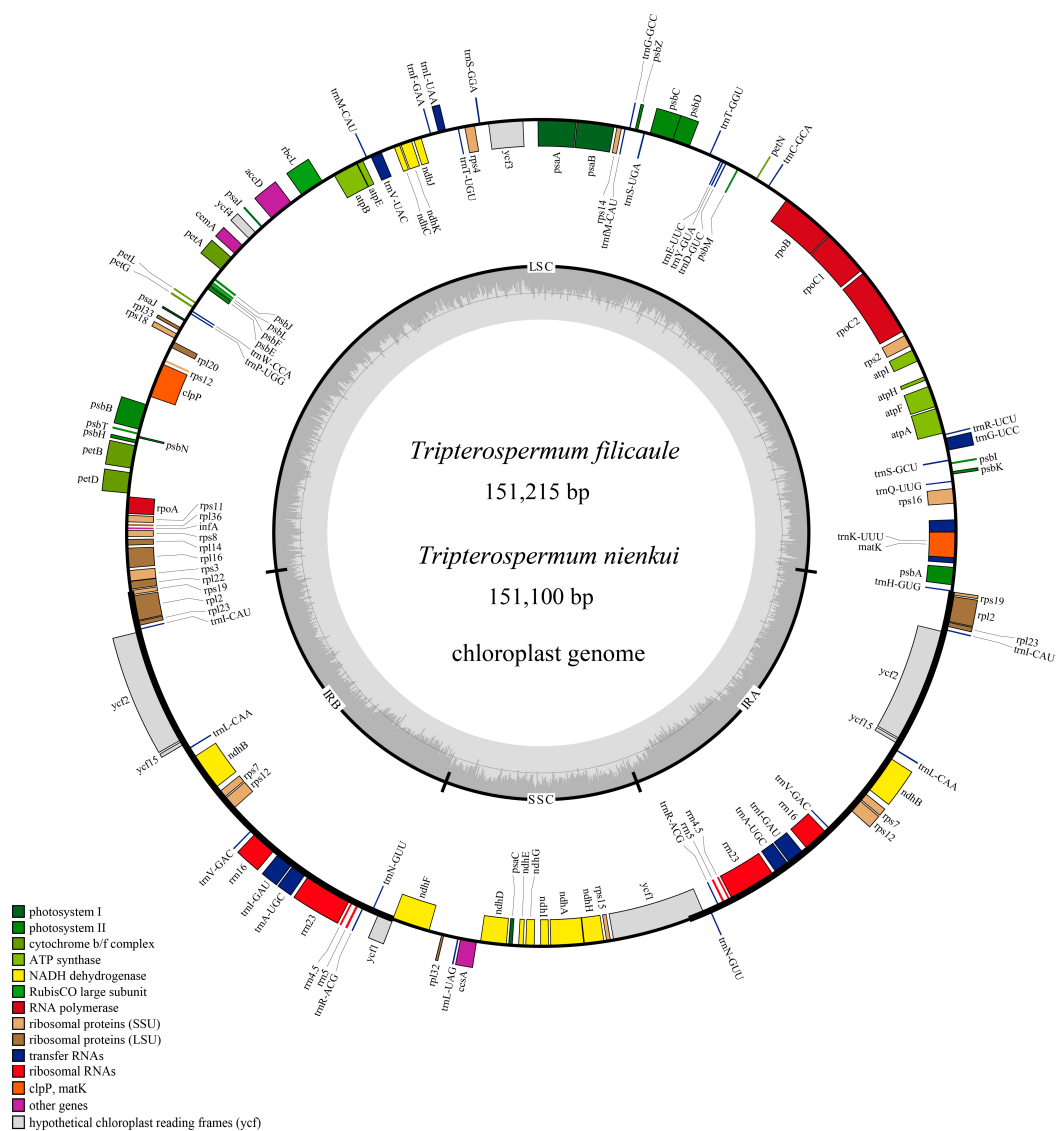
Figure 1. Annotation map of the plastomes for *T. filicaule* and *T. nienkui*. In the inner circle, varying shades of gray depict the distribution of GC content, with darker gray indicating higher GC content. The inner circle further distinguishes between AT content, represented by lighter shades, and GC content, represented by darker shades across the plastome.

The plastomes of six *Tripterospermum* species encoded 134 genes (114 unique), including 86 protein-coding genes (79 unique), 37 transfer RNA (tRNA) genes (30 unique), eight ribosomal RNA (rRNA) genes (four unique), and three pseudogenes (*infA*, *rps19*, and *ycf1*) (Table 1, Supplementary Table S1). The examined plastomes revealed an absence of gene organization rearrangements (Supplementary Figure S1).

Visualization of the SC/IR boundary region in the *Tripterospermum* plastomes revealed a highly conserved pattern (Figure 2). The protein-coding genes *rps19*, *ndhF*, and *ycf1* span the LSC/IRb (JLB), SSC/IRb (JSB), and SSC/IRa (JSA) boundaries. Due to the reverse repetition characteristic of the IR region, *rps19* and *ycf1* generate pseudogenic versions of themselves at IRa and IRb, respectively.

Table 1. Genetic classification of the *Tripterospermum* plastomes.

Category of Genes	Group	Name
Photosynthesis	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	Large subunit of rubisco	*rbcL*
	Cytochrome b/f complex	*petA, petB *, petD *, petG, petL, petN*
	ATP synthase	*atpA, atpB, atpE, atpF *, atpH, atpI*
	NADH dehydrogenase	*ndhA *, ndhB * (*2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Self-replication-related genes	Ribosomal RNA	*rrn4.5 (*2), rrn5 (*2), rrn16 (*2), rrn23 (*2)*
	Transfer RNA	*trnA-UGC * (*2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC *, trnH-GUG, trnI-CAU (*2), trnI-GAU * (*2), trnK-UUU *, trnL-CAA (*2), trnL-UAA *, trnL-UAG, trnM-CAU, trnN-GUU (*2), trnP-UGG, trnQ-UUG, trnR-ACG (*2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC (*2), trnV-UAC *, trnW-CCA, trnY-GUA*
	RNA polymerase	*rpoA, rpoB, rpoC1 *, rpoC2*
	Small subunit of ribosomal proteins	*rps2, rps3, rps4, rps7 (*2), rps8, rps11, rps12 ** (*2), rps14, rps15, rps16 *, rps18, Ψrps19, rps19*
	Large subunit of ribosomal proteins	*rpl2 * (*2), rpl14, rpl16 *, rpl20, rpl22, rpl23 (*2), rpl32, rpl33, rpl36*
Other genes	Protease	*clpP ***
	Maturase	*matK*
	Envelop membrane protein	*cemA*
	Acetyl-CoA-carboxylase	*accD*
	Translation initiation factor	*ΨinfA*
	C-type cytochrome synthesis	*ccsA*
Genes with unknown function	Hypothetical chloroplast reading frames	*Ψycf1, ycf1, ycf2 (*2), ycf3 **, ycf4*

* Gene with one intron, ** gene with two introns, (*2) gene with two copies, and Ψ gene with a pseudogene.

2.2. Codon Usage Analysis

A total of seventy-nine unique protein-coding genes (CDS) were analyzed among the six *Tripterospermum* plastomes, including an assessment of codon usage frequency and the calculation of relative synonymous codon usage (RSCU) (Figure 3, Supplementary Table S2). The total number of codons for these genes varied between 22,457 (*T. japonicum*) and 22,919 (*T. championii*) (Supplementary Table S2) among the six plastomes of *Tripterospermum*. We identified a total of 65 synonymous codons (including stop codons). Among these codons, 32 codons had RSCU $\geq$ 1, and 33 had RSCU < 1 (Supplementary Table S2). Leucine (Leu) was consistently the most frequently encoded amino acid in all *Tripterospermum* plastomes, with a percentage ranging from 10.54% to 10.62%, while cysteine (Cys: 1.15–1.16%) was the least abundant (Figure 3).
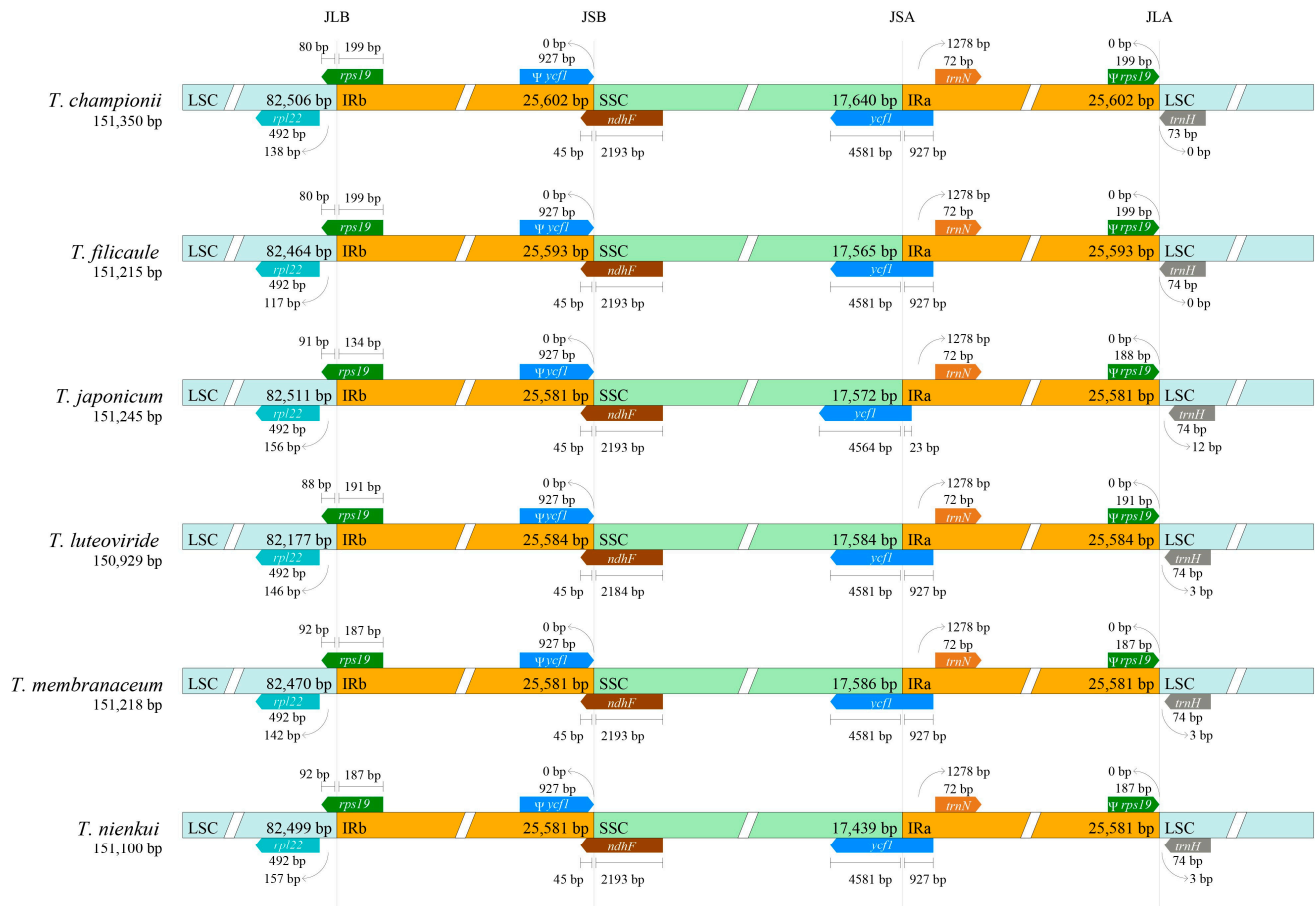
Figure 2. Comparison of junctions among the LSC, SSC, and IRs in the six *Tripterospermum* plastomes. Pseudogenes are marked with (Ψ). The junction sites are labeled as JLB (LSC/IRb), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC).
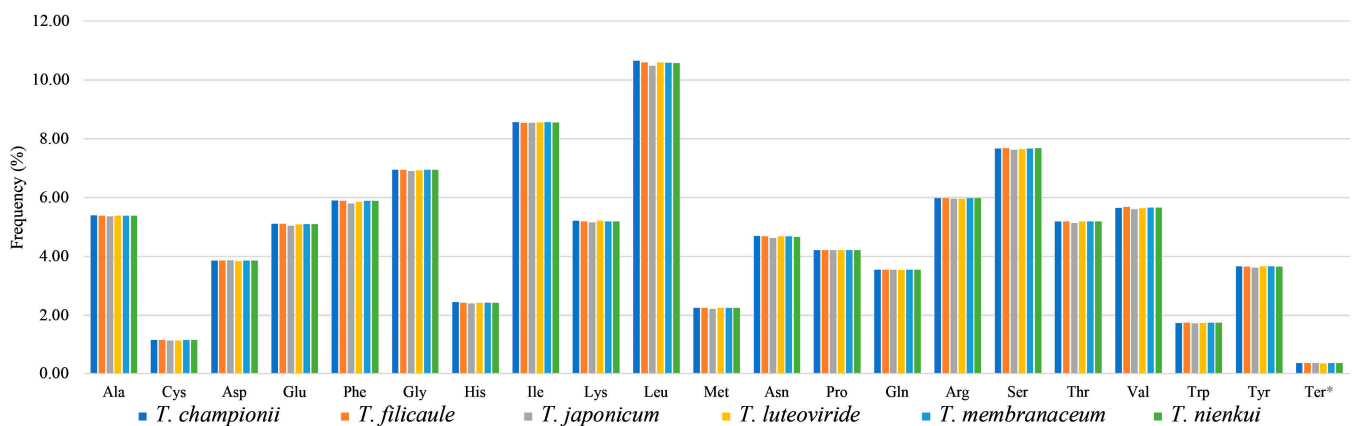

Figure 3. Amino acid frequencies in six *Tripterospermum* plastomes based on protein-coding sequences. * represents the termination codon.

2.3. Repeat Sequence Analysis

Simple sequence repeats (SSRs) were further identified, and a total of 43, 40, 42, 45, 39, and 39 SSRs were identified in the plastomes of *T. championii*, *T. filicaule*, *T. japonicum*, *T. luteoviride*, *T. membranaceum*, and *T. nienkui*, respectively (Figure 4, Supplementary Table S3). Five SSR types, mononucleotides, dinucleotides, trinucleotides, tetranucleotides, and pentanucleotides (mono-, di-, tri-, tetra-, and penta-), all appeared in *Trilterospermum*

species, but we did not find hexanucleotides in *T. championii*. Mononucleotide repeats are the most abundant, comprising 62.50%, followed by tetranucleotide repeats at 14.52%, whereas pentanucleotides and hexanucleotide repeats are very rare among these plastomes, accounting for only 2.42% and 2.01% of the total, respectively. In all analyzed plastomes, SSRs are predominantly situated in the LSC and IGS regions (Supplementary Figure S2).
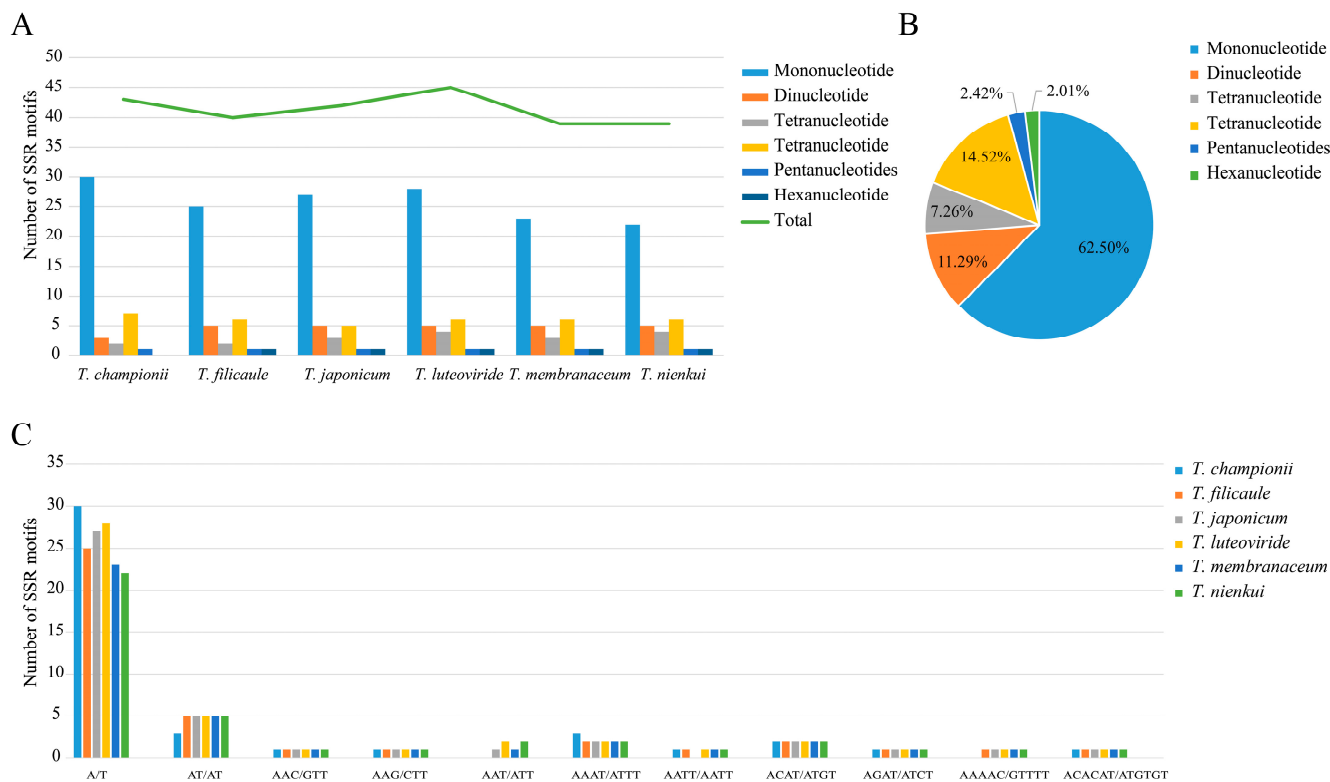

Figure 4. Analyses of SSRs in the six *Tripterospermum* plastomes: (**A**) count and types of SSRs; (**B**) percentage distribution of SSR types; and (**C**) number of SSR motifs.

Four types of long repeats (forward, palindromic, complement, and reverse repeats) were detected in the *Tripterospermum* plastomes. The number of long repeats varied from 14 (*T. filicaule* and *T. luteoviride*) to 29 (*T. championii*) (Supplementary Figure S3, Supplementary Table S4). Forward repeats (56.76% of the total long repeats) and palindromic repeats (34.23% of the total long repeats) are the most abundant in the total repetition and exist in all plastomes. Reverse repeats (8.11% of the total long repeats) are identified in the plastomes of four species, whereas complement repeats (1.00% of the total long repeats) are exclusively found in *T. nienkui* plastomes (Supplementary Figure S3, Supplementary Table S4). We categorized all repeats into three groups based on their length (30–39 bp, 40–49 bp, and 50–59 bp). Among these, ninety-nine (89.19%) have lengths of 30–39 bp, while only six (5.41%) have lengths of 40–49 bp and 50–59 bp.

2.4. Sequence Divergence Analysis

The divergence of the six *Tripterospermum* plastomes was assessed using the mVISTA online platform, with *T. championii* as the reference. (Figure 5). The results demonstrated that the full-length *Tripterospermum* plastomes were highly conserved. The coding regions, including exons and t/rRNA regions, are more conserved than non-coding sequences. To gain a more comprehensive insight into sequence divergence among *Tripterospermum* plastomes, both coding regions and intergenic regions were isolated for the calculation of nucleotide variability (Pi) (Figure 6, Supplementary Table S5). For coding regions, the mean pi is 0.00071, indicating high conservatism. The highest pi value is in the *psaI* region, but it is only 0.01201. There are 34 regions with a pi value of 0 (39.08% of all coding
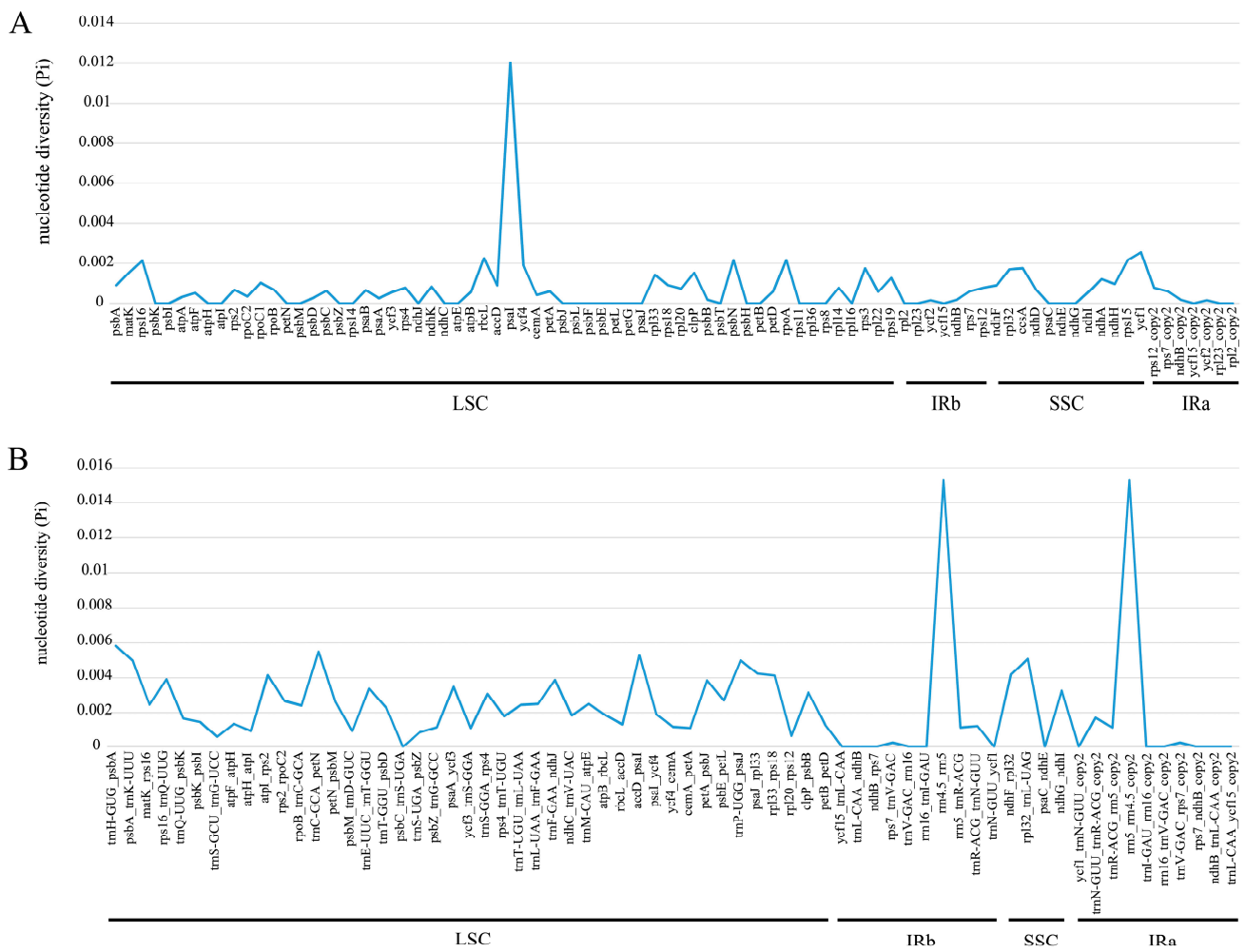
regions) (Figure 6A, Supplementary Table S5). For intergenic regions, the mean pi is 0.00232, indicating high conservatism, and the highest pi value is in the *rrn4.5-rrn5* region, with a pi value of 0.01522. There are 14 regions with a pi value of 0 (20.59% of all intergenic regions) (Figure 6B, Supplementary Table S5).

Figure 5. Sequence alignment of six *Tripterospermum* plastomes using the mVISTA program, with *T. championii* serving as the reference. The X- and Y-scales denote the coordinates within plastomes and the percentage of identity (ranging from 50% to 100%), respectively. Gray arrows indicate the transcription direction of each gene. Genome regions are color coded to represent protein-coding regions, tRNA, rRNA, and conserved non-coding regions.

Figure 6. Nucleotide diversity (Pi) of the aligned six *Tripterospermum* plastomes: (**A**) protein-coding genes; (**B**) intergenic regions.

2.5. Phylogenetic Analyses

To elucidate the phylogenetic relationships of subtribe Gentianae and *Tripterospermum*, we used seventy-nine unique CDSs extracted from the eighty-six plastomes, including eighty-one plastomes from subtribe Gentianinae and five plastomes from Swertiinae, to construct the Gentianae phylogenetic tree (Figure 7). The results showed that each genus of Gentianae is monophyletic and is highly supported, divided into two clades. One clade comprises *Gentiana*, which accounts for the majority and is divided into twelve sections with high support (Figure 7). The other clade includes *Kuepferia*, *Crawfurdia*, *Metagentiana*, *Sinogentiana*, and *Tripterospermum*; *Tripterospermum* is sister to *Sinogentiana*, and towards the base are *Metagentiana*, *Crawfurdia*, and *Kuepferia* (Figure 7). Within the genus *Tripterospermum*, the relationship between the three species, *T. membranaceum*, *T. filicaule*, and *T. nienkui*, has moderate support (Figure 7). To further clarify the impact of using different plastome tree-building strategies on the phylogenetic relationship of *Tripterospermum*, we developed phylogenetic trees based on four datasets: complete plastid genomes, CDS, intergenic regions (IGS), and a combined dataset encompassing both CDS and IGS. The results show that the phylogenetic tree constructed by CDS (Figure 8A) and CDS + IGS (Figure 8C) have similar support, and both are better than IGS (Figure 8B) and complete plastid (Figure 8D) constructed phylogenetic tree, supporting higher and better ability to distinguish *Tripterospermum* species. However, in the genus *Kuepferia*, the phylogenetic tree constructed by CDS has nodes with an unstable relationship (0/45/0.65), while the phylogenetic tree constructed by CDS + IGS had high support (100/100/1.00).
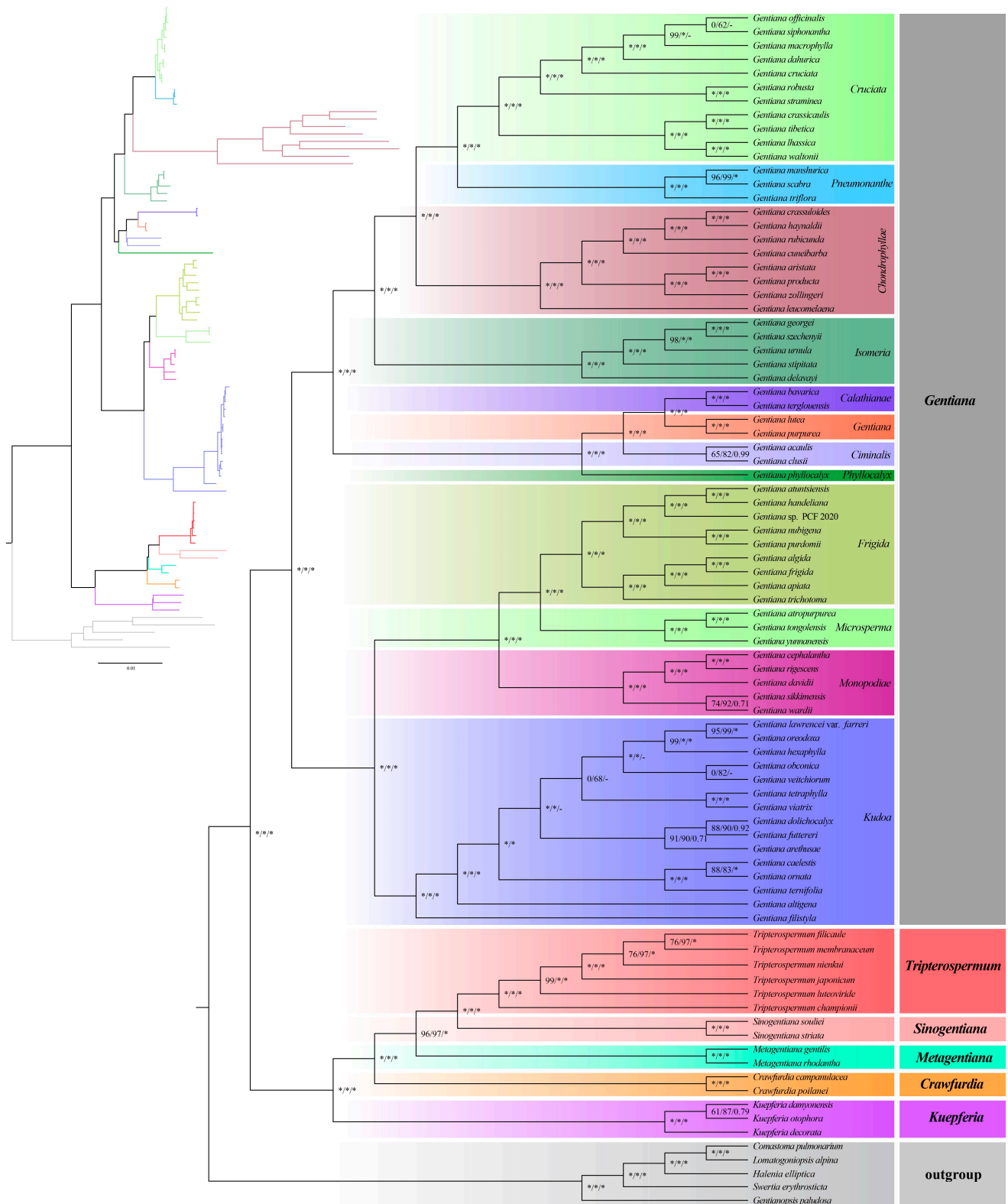
Figure 7. Phylogenetic tree of Gentianinae based on coding genes regions from plastome. Bootstrap percentages and Bayesian posterior probabilities are indicated as values near the nodes (SH-aLRT left, UFBoot middle, and PP right). An asterisk (*) denotes nodes with a confidence level of 100% in bootstrap percentages or 1.00 in Bayesian posterior probabilities.
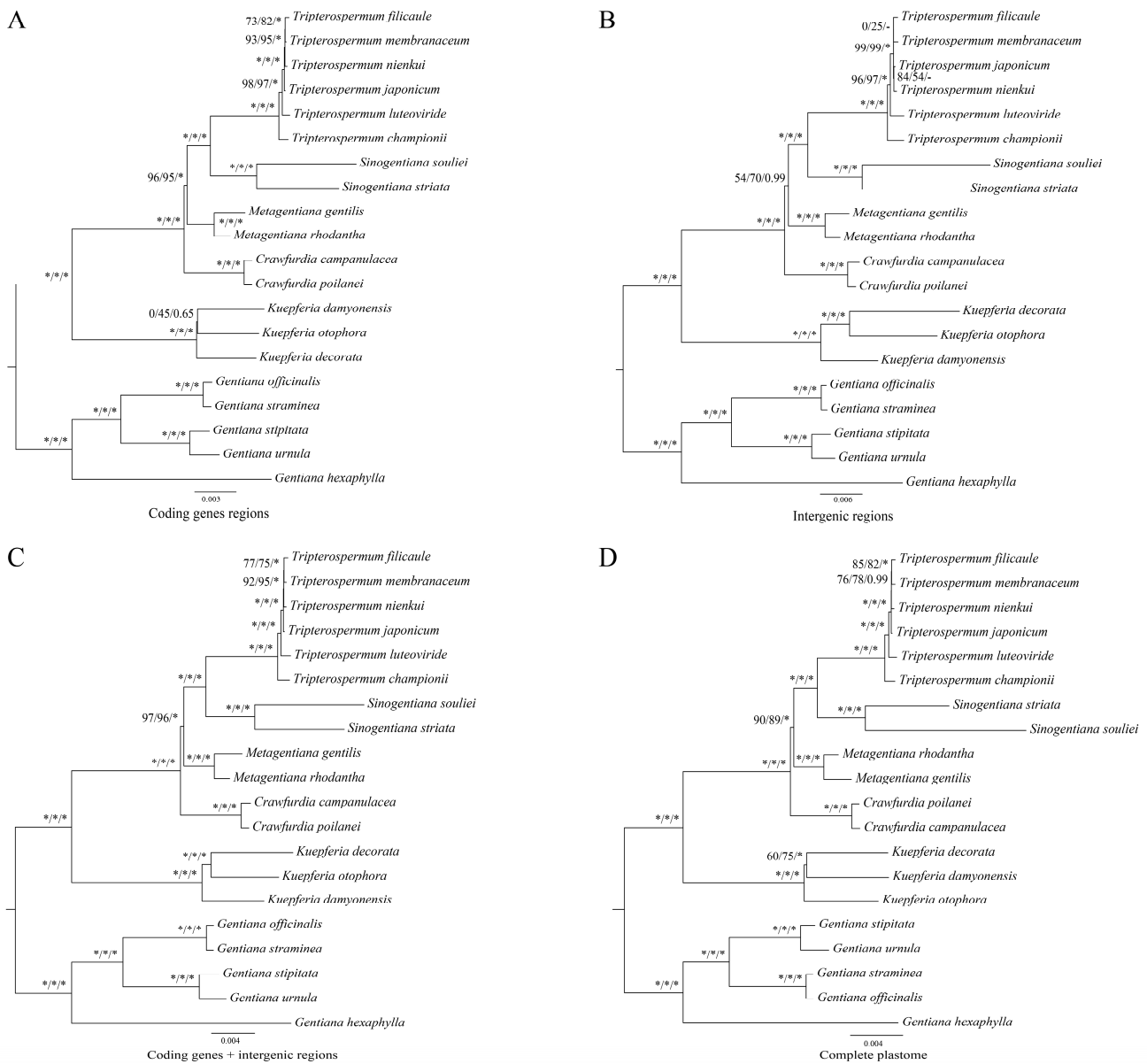
Figure 8. Phylogenetic tree of *Tripterospermum* based on plastome. Bootstrap percentages and Bayesian posterior probabilities are indicated as values near the nodes (SH-aLRT left, UFBoot middle, and PP right). An asterisk (*) denotes nodes with a confidence level of 100% in bootstrap percentages or 1.00 in Bayesian posterior probabilities. (**A**) phylogenetic tree based on coding genes regions (CDS); (**B**) phylogenetic tree based on intergenic regions (IGS); (**C**) phylogenetic tree based on coding genes + intergenic regions (CDS + IGS); (**D**) phylogenetic tree based on complete plastomes.

3. Discussion

3.1. Variation of Plastome Sequences

In this study, we expanded plastome sampling in *Tripterospermum*, providing a valuable opportunity to enhance our understanding of plastome evolution in this complex taxon. We assembled and annotated two *Tripterospermum* plastomes, conforming to the typical quadripartite structure shared by most angiosperms comprising one LSC region, one SSC region, and two IR regions. The size of *Tripterospermum* plastomes varied from 150,929 bp (*T. luteoviride*) to 151,350 bp (*T. championii*). This range aligns with the reported sizes of plastomes in Gentianinae, spanning from 117,780 bp (*Gentiana producta*) to 151,350 bp (*T. championii*) [30]. The six *Tripterospermum* plastomes exhibit conservatism in the number and types of genes, collectively comprising 134 genes, including 86 CDS, 37 tRNAs, and

eight rRNAs. The plastid genomes of Gentianeae species have all undergone pseudogenization of the *rps16* gene. Since then, the loss of *rps16*, *ndh* complexes, the second intron of *clpP*, and the intron of *rpl2* have occurred in many genera or *Gentiana* sections in Gentianinae [30]. However, these phenomena were not found in *Tripterospermum*, and only pseudogenization of the *rps16* gene existed.

This study conducted analyses of Mauve collinearity and plastome boundaries, revealing a highly conserved structure among *Tripterospermum* plastomes. The expansion and contraction of the IR boundary play a pivotal role in the evolution of species [32]. Concerning the gene arrangement at the boundaries within the six *Tripterospermum* plastomes, we observed a high degree of conservation at LSC/IRa (JLA), JLB, JSA, and JSB. Due to the spanning of JLB and JSA by the two genes, *rps19* and *ycf1*, their duplicated pseudogenes appear in another IR region, a phenomenon documented in numerous previous studies [28,30,33].

3.2. Codon Usage and Repeat Sequence Analysis

The bias in codon usage is a significant factor in the evolution of plastomes and influences the expression of gene functions [34]. Organisms that share close genetic relationships often display highly similar codon usage biases [35]. The patterns of codon frequency and RSCU values were notably similar in *Tripterospermum* plastomes. Among all codons, Leu displayed the highest occurrence (10.54–10.62%), while Cys had the lowest frequency (1.15–1.16%). It is worth noting that the total amount of amino acid in the plastid of *T. filicaule*, *T. membranaceum*, and *T. nienkui* was the same, which was 22,857. *T. filicaule* was a sister to *T. membranaceum*, and they were sisters with *T. nienkui*, and the total amount of amino acid in their plastomes was the same, which is related to their close relationship.

The SSRs, also called microsatellites, comprise repeating 1–6 nucleotide motifs (mono-, di-, tri-, tetra-, penta-, and hexa- repeats). SSRs often exhibit high levels of polymorphism and find extensive applications in species authentication, elucidating evolutionary relationships and assessing genetic diversity within plant populations [36–38]. In this study, we identified a total of 39 to 45 SSRs in the plastomes of *Tripterospermum*. We found that the number of SSRs in *Tripterospermum* plastome introns was correlated with phylogenetic relationships. The number of SSRs in the plastome introns of *T. filicaule* was the lowest (2) and located at the top of the phylogenetic tree, followed by *T. membranaceum* (3), *T. nienkui* (3), *T. japonicum* (4), *T. championii* (5), and *T. luteoviride* (5). The plastome introns of *T. championii* and *T. luteoviride* had the highest number of SSRs, located at the base of the phylogenetic tree. Among single-nucleotide repeats, only A/T repeat sequences were observed, with no G/C repeat sequences found. This observation may be attributed to the low occurrence of single-nucleotide repeats and the A/T bias in plastomes [39,40], consistent with previous studies [28]. Furthermore, the numbers of single-nucleotide repeats (A/T) varied across the six *Tripterospermum* plastomes, providing potential utility for future population genetic and phylogenetic studies. Repeat sequences > 30 bp, known as long repeat sequences, were also identified in this study. These sequences may play roles in genome recombination, rearrangement, and phylogeny, and contribute to insertions and substitutions in plastomes [41]. The predominant types of repeats identified were forward repeats and palindromic repeats falling within the 30–39 bp range. These findings are particularly noteworthy for the identification and analysis of genetic diversity in *Tripterospermum* plants.

3.3. Sequence Divergence Analysis and Phylogenetic Analysis

To assess the variations of *Tripterospermum* plastomes, mVISTA was employed to investigate the structural characteristics of the six plastomes. The results indicate that coding regions were more conserved than non-coding regions. To explore highly mutated hotspots in *Tripterospermum* plastomes, both coding and intergenic regions were extracted for pi calculation. The mean pi of the coding and intergenic regions was only 0.071% and 0.232%, respectively. This result shows surprisingly low differences between all aligned sequences, highlighting the conservation of coding and intergenic regions, indicating that

Tripterospermum plastomes are highly conserved. In previous studies, the ITS, *atpB-rbcL*, and *trnL-trnF* sequences could not completely resolve the phylogenetic relationship of *Tripterospermum* at the molecular level [10,11,22]. This requires the introduction of more site sequences.

The phylogenetic relationship within the subtribe Gentianinae, particularly between *Gentiana* and the genera *Crawfurdia*, *Tripterospermum*, and *Metagentiana*, has been a subject of controversy in previous studies [10,15–21]. To address the non-monophyletic issue of *Gentiana* sect. *Otophora* and *Metagentiana*, Favre et al. [10] established new genera, *Kuepferia* and *Sinogentiana*, respectively. A recent study conducted an extensive sampling study on Gentianinae, revealing that *Crawfurdia*, *Gentiana*, *Kuepferia*, *Metagentiana*, *Sinogentiana*, and *Tripterospermum* are all monophyletic. However, the relationships within the *Tripterospermum* genus are not entirely clear, such as the phylogenies between *Tripterospermum* species being low and requiring further improvement [22].

The significance of plastomes in reconstructing phylogenetic relationships and comprehending evolutionary history has been firmly established, with an expanding application in Gentianinae research. However, existing plastid genome systematic studies in this context only involve a limited number of *Tripterospermum* species [29–31]. To explore the potential application of plastid genome systematics in *Tripterospermum*, we utilized 79 unique CDSs extracted from the 86 plastomes to construct a Gentianae phylogenetic tree. The results demonstrated the monophyly of each genus within Gentianae, divided into two well-supported clades. One clade is predominantly represented by *Gentiana*, further divided into twelve sections. The other branch shows *Tripterospermum* as a sister to *Sinogentiana*, followed by *Metagentiana*, *Crawfurdia*, and *Kuepferia*, consistent with Fu et al.'s findings [30]. A short branch-length phylogenetic tree with moderate support is observed in *Tripterospermum*.

To further explore the phylogenetic relationships within *Tripterospermum*, we constructed a phylogenetic tree using four matrices, including complete plastid genomes, CDS, IGS, and CDS + IGS. The trees constructed by CDS and CDS + IGS exhibited similar and higher support compared to those constructed by IGS and the complete plastid matrix and were better at distinguishing *Tripterospermum* species. However, in the genus *Kuepferia*, the phylogenetic tree constructed by CDS showed an unstable relationship node (0/45/0.65), while the tree constructed by CDS + IGS exhibited high support (100/100/1.00). Therefore, CDS and CDS + IGS combined matrices serve as a valuable reference for resolving disputes regarding the phylogenetic relationships of *Tripterospermum*, with the CDS + IGS combined matrix yielding more reliable results overall.

4. Materials and Methods

4.1. Plant Materials and Sequencing

Fresh leaves of *T. filicaule* and *T. nienkui* were collected from Sanming City, China, and the voucher specimens were deposited at the Herbarium of Fujian Agriculture and Forestry University in Fuzhou, China. Total genomic DNA was extracted from leaves using a modified CTAB method [42], the quantity and quality of extracted DNA was assessed by spectrophotometry, and the integrity was evaluated using a 1% (*w/v*) agarose gel electrophoresis. The purified DNA samples were sheared into fragments with an average length of 350 bp for library preparation following the manufacturer's guidelines (Illumina, San Diego, CA, USA). These libraries were sequenced on the Illumina HiSeq-2500 platform at the Beijing Genomics Institute (Shenzhen, China).

4.2. Plastome Assembly and Annotation

The raw reads were verified using Trimmomatic v.0.32 [43] with default settings to obtain high-quality clean reads. The paired-end clean reads were assembled into complete plastid genomes using GetOrganelle v1.7.1 [44] and then annotated by PGA software [45] and were manually adjusted by Geneious 11.1.5 [46]. To avoid confusion of annotation, we re-annotated the plasmids obtained from GenBank. The annotation circle maps of

plastomes were drawn using the online tool OGDRAW v.1.3.1 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html, accessed on 2 October 2023) [47].

4.3. Plastome Structure Comparisons and Sequence Divergence Analysis

The rearrangements between different *Tripterospermum* plastomes were identified by Mauve v.2.4.0 [48]. Whole-genome alignment was performed and plotted with the mVISTA program (http://genome.lbl.gov/vista/mvista/submit.shtml, accessed on 17 October 2023) [49] in the Shuffle-LAGAN model. The *T. championii* (MN199139) plastome was used as a reference. The Perl script CPJSdraw.pl (https://github.com/xul962464/CPJSdraw, accessed on 2 November 2023) was used to compare the genes in the boundary regions of LSC/IRb/SSC/IRa. The nucleotide diversity (Pi) of protein-coding genes (CDS) and intergenic regions (IGS) of six *Tripterospermum* plastomes was evaluated using DnaSP v.6.12.01 software [50].

4.4. Repeat Sequence and Codon Usage Analysis

The simple sequence repeats (SSRs) were determined by MISA (https://webblast.ipk-gatersleben.de/misa/, accessed on 22 November 2023) [51] with 10, 5, 4, 3, 3, and 3 nucleotide repeats set for mononucleotides (mono-), dinucleotides (di-), trinucleotides (tri-), tetranucleotides (tetra-), pentanucleotides (penta-), and hexanucleotides (hexa-), respectively. Four long repeat types in six plastid genomes, F (forward), P (palindrome), R (reverse), and C (complement) repeats, were detected by the REPuter program (https://bibiserv.cebitec.uni-bielefeld.de/reputer, accessed on 6 November 2023) [52] with a minimum repeat size of 30 bp, an edit distance of three, and 90% similarity. Codon usage and relative synonymous codon usage (RSCU) values were calculated using CodonW v1.4.4 software [53]. Repeat sequences of protein-coding regions were eliminated from the codon usage calculations to avoid sampling errors.

4.5. Phylogenetic Analyses

In this study, we used the 81 plastomes of the subtribe Gentianinae to construct phylogenetic trees, including two *Crawfurdia*, sixty-six *Gentiana*, three *Kuepferia*, two *Metagentiana*, two *Sinogentiana,* and six *Tripterospermum*. Five Swertiinae (*Comastoma pulmonarium*, *Gentianopsis paludosa*, *Halenia elliptica*, *Lomatogoniopsis alpina*, and *Swertia erythrosticta*) were set as the outgroups (Supplementary Table S6). CDSs were extracted from the 86 plastomes using PhyloSuite v.1.2.2 [54]. By removing repetitive sequences, the final 79 CDS regions were used to construct the phylogenetic tree. The sequences were aligned by MAFFT v7.490 [55] with auto parameters and then concatenated using PhyloSuite v1.2.2 [54].

The phylogenetic trees were inferred by maximum likelihood (ML) and Bayesian inference (BI). ML analyses were performed in IQ-TREE v2.0.3 [56], incorporating the SH-aLRT test and ultrafast bootstrap (UFBoot) feature, with the model identified by ModelFinder implemented in IQ-TREE (-alrt 2000 -B 2000 -m MFP). BI analysis was performed with MrBayes v3.2.7 [57] using best-fit models, which were selected with MrModeltest 2.4 [58] using AIC. The following settings were used: ngen = 5,000,000; samplefreq = 1000; burninfrac = 0.25.

5. Conclusions

This study assembled and annotated two *Tripterospermum* plastomes, comparing them with those of other *Tripterospermum* species to investigate plastid genome differences within the genus. The plastomes of *Tripterospermum* displayed a typical quadripartite structure, encompassing an LSC region, an SSC region, and a pair of IRs. The plastome lengths of *Tripterospermum* ranged from 150,929 bp (*T. luteoviride*) to 151,350 bp (*T. championii*) and encoded 134 genes, including 86 protein-coding genes, 37 tRNA genes, eight rRNA genes, and three pseudogenes (*infA*, *rps19*, and *ycf1*). Among these, the CDS region of *psaI* and the IGS region of *rrn4.5-rrn5* exhibited the highest variability. Our phylogenetic analysis revealed that matrices constructed using CDS and CDS + IGS provided stronger support

within the genus *Tripterygium*. However, for Gentianae, the CDS + IGS combined matrix yielded superior results. This study contributes novel evidence supporting the utility of plastid genomes in elucidating the phylogeny of *Tripterospermum*, thereby aiding in resolving classification challenges within this genus.

References

1. Murata, J. *Synopsis of Tripterospermum (Gentianaceae). Journal of the Faculty of Science*; Section III: Botany; University of Tokyo: Bunkyo City, Japan, 1989.
2. POWO. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. 2023. Available online: http://www.plantsoftheworldonline.org/ (accessed on 28 November 2023).
3. Ho, T.N.; Pringle, J.S. Gentianaceae. In *Flora of China*; Wu, Z.Y., Raven, P.H., Eds.; Science Press: Beijing, China; Missouri Botanic Garden Press: St. Louis, MO, USA, 1995; Volume 16, p. 7.
4. Chen, C.-H.; Lin, J.-Y.; Lin, C.-N.; Hsu, S.-Y. Inhibition of angiotensin-I-converting enzyme by tetrahydroxyxanthones isolated from *Tripterospermum lanceolatum*. *J. Nat. Prod.* **1992**, *55*, 691–695. [CrossRef]
5. Hsu, M.-F.; Raung, S.-L.; Tsao, L.-T.; Lin, C.-N.; Wang, J.-P. Examination of the inhibitory effect of norathyriol in formylmethionyl-leucyl-phenylalanine-induced respiratory burst in rat neutrophils. *Free Radic. Biol. Med.* **1997**, *23*, 1035–1045. [CrossRef]
6. Otsuka, H.; Kijima, K. An iridoid gentiobioside, a benzophenone glucoside and acylated flavone C-glycosides from *Tripterospermum japonicum* (Sieb. et Zucc.) Maxim. *Chem. Pharm. Bull.* **2001**, *49*, 699–702. [CrossRef]
7. Zhu, K.C.; Ma, C.H.; Ye, G.; Fan, M.S.; Huang, C.G. Two new secoiridoid glycosides from *Tripterospermum chinense*. *Helv. Chim. Acta* **2007**, *90*, 291–296. [CrossRef]
8. Fang, J.-J.; Ye, G. Flavonoids and xanthones from *Tripterospermum chinense*. *Chem. Nat. Compd.* **2008**, *44*, 514–515. [CrossRef]
9. Zhang, T.; Ma, C.-J.; Wei, Y.-L.; Si, J.-G.; Fu, L.; Dong, J.-X.; Zou, Z.-M. The novel indole glucoalkaloid and secoiridoid glucoside from *Tripterospermum chinense*. *Phytochem. Lett.* **2020**, *35*, 191–196. [CrossRef]
10. Favre, A.; Yuan, Y.-M.; Küpfer, P.; Alvarez, N. Phylogeny of subtribe *Gentianinae* (*Gentianaceae*): Biogeographic inferences despite limitations in temporal calibration points. *Taxon* **2010**, *59*, 1701–1711. [CrossRef]
11. Favre, A.; Matuszak, S.; Sun, H.; Liu, E.; Yuan, Y.-M.; Muellner-Riehl, A.N. Two new genera of *Gentianinae* (*Gentianaceae*): *Sinogentiana* and *Kuepferia* supported by molecular phylogenetic evidence. *Taxon* **2014**, *63*, 342–354. [CrossRef]
12. Blume, C.L. *Tripterospermum. Bijdrangen tot de Flora van Nederlandsch Indië*; Lands Drukkerij: Batavia, Indonesia, 1826; p. 849.
13. Wallich, N. *Tentanien Florae Napalensis Illustratae, Part 2*; Asiatic Lithographicae Press: Serampore/Calcutta, India, 1826; pp. 37–64.
14. Ho, T.N.; Chen, S.L.; Liu, S.W. *Metagentiana*, a new genus of Gentianaceae1. *Bot. Bull. Acad. Sin.* **2002**, *43*, 83–91.
15. Clarke, C.B. Gentianaceae. In *The Flora of British India*; Hooker, J.D., Ed.; Reeve: London, UK, 1885; Volume 4, pp. 93–132.
16. Ridley, H.N. Gentianaceae. In *The Flora of the Malay Peninsula*; L. Reeve & Co.: London, UK, 1923; Volume 2, pp. 432–437.
17. Ubolcholaket, A. Gentianaceae. In *Flora of Thailand*; Smitinand, T., Larsen, K., Eds.; Chutima Press: Bangkok, Thailand, 1987; Volume 5, pp. 72–92.
18. Marquand, C.V.B. New Asiatic Gentians: II. *Bull. Misc. Inf. R. Bot. Gard. Kew* **1931**, *1931*, 68–88. [CrossRef]

19. Marquand, C.V.B. The gentians of China. *Bull. Misc. Inf. R. Bot. Gard. Kew* **1937**, *1937*, 134–180. [CrossRef]

20. Chen, S.; Xia, T.A.O.; Wang, Y.; Liu, J.; Chen, S. Molecular systematics and biogeography of *Crawfurdia*, *Metagentiana* and *Tripterospermum* (Gentianaceae) based on nuclear ribosomal and plastid DNA sequences. *Ann. Bot.* **2005**, *96*, 413–424. [CrossRef] [PubMed]

21. Chen, S.-Y.; Chen, S.-L.; Xia, T.; Wang, Y.-J. Phylogeny of *Metagentiana* and related genera (*Gentianaceae*) inferred from nuclear ribosomal ITS sequences. *J. Syst. Evol.* **2005**, *43*, 491. [CrossRef]

22. Matuszak, S.; Muellner-Riehl, A.N.; Sun, H.; Favre, A. Dispersal routes between biodiversity hotspots in Asia: The case of the mountain genus *Tripterospermum* (Gentianinae, Gentianaceae) and its close relatives. *J. Biogeogr.* **2016**, *43*, 580–590. [CrossRef]

23. Patwardhan, A.; Ray, S.; Roy, A. Molecular markers in phylogenetic studies-a review. *J. Phylogenetics Evol. Biol.* **2014**, *2*, 131.

24. Clegg, M.T.; Gaut, B.S.; Learn, G.H., Jr.; Morton, B.R. Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6795–6801. [CrossRef] [PubMed]

25. Tu, X.-D.; Liu, D.-K.; Xu, S.-W.; Zhou, C.-Y.; Gao, X.-Y.; Zeng, M.-Y.; Zhang, S.; Chen, J.-L.; Ma, L.; Zhou, Z. Plastid phylogenomics improves resolution of phylogenetic relationship in the *Cheirostylis* and *Goodyera* clades of *Goodyerinae* (*Orchidoideae*, *Orchidaceae*). *Mol. Phylogenetics Evol.* **2021**, *164*, 107269. [CrossRef]

26. Zeb, U.; Dong, W.L.; Zhang, T.T.; Wang, R.N.; Shahzad, K.; Ma, X.F.; Li, Z.H. Comparative plastid genomics of *Pinus* species: Insights into sequence variations and phylogenetic relationships. *J. Syst. Evol.* **2020**, *58*, 118–132. [CrossRef]

27. Jansen, R.K.; Ruhlman, T.A. Plastid genomes of seed plants. *Genom. Chloroplasts Mitochondria* **2012**, *35*, 103–126.

28. Zhou, T.; Wang, J.; Jia, Y.; Li, W.; Xu, F.; Wang, X. Comparative chloroplast genome analyses of species in *Gentiana* section *Cruciata* (Gentianaceae) and the development of authentication markers. *Int. J. Mol. Sci.* **2018**, *19*, 1962. [CrossRef]

29. Zhang, X.; Sun, Y.; Landis, J.B.; Lv, Z.; Shen, J.; Zhang, H.; Lin, N.; Li, L.; Sun, J.; Deng, T. Plastome phylogenomic study of Gentianeae (Gentianaceae): Widespread gene tree discordance and its association with evolutionary rate heterogeneity of plastid genes. *BMC Plant Biol.* **2020**, *20*, 340. [CrossRef]

30. Fu, P.C.; Sun, S.S.; Twyford, A.D.; Li, B.B.; Zhou, R.Q.; Chen, S.L.; Gao, Q.B.; Favre, A. Lineage-specific plastid degradation in subtribe *Gentianinae* (*Gentianaceae*). *Ecol. Evol.* **2021**, *11*, 3286–3299. [CrossRef] [PubMed]

31. Zhang, Y.; Yu, J.; Xia, M.; Chi, X.; Khan, G.; Chen, S.; Zhang, F. Plastome sequencing reveals phylogenetic relationships among *Comastoma* and related taxa (Gentianaceae) from the Qinghai–Tibetan Plateau. *Ecol. Evol.* **2021**, *11*, 16034–16046. [CrossRef] [PubMed]

32. 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]

33. Tu, X.-D.; Zhao, Z.; Zhou, C.-Y.; Zeng, M.-Y.; Gao, X.-Y.; Li, M.-H.; Liu, Z.-J.; Chen, S.-P. Comparative Analysis of Plastomes in Elsholtzieae: Phylogenetic Relationships and Potential Molecular Markers. *Int. J. Mol. Sci.* **2023**, *24*, 15263. [CrossRef] [PubMed]

34. Quax, T.E.F.; Claassens, N.J.; Söll, D.; van der Oost, J. Codon bias as a means to fine-tune gene expression. *Mol. Cell* **2015**, *59*, 149–161. [CrossRef]

35. Sharp, P.M.; Cowe, E.; Higgins, D.G.; Shields, D.C.; Wolfe, K.H.; Wright, F. Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res.* **1988**, *16*, 8207–8211. [CrossRef]

36. McCauley, D.E. The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends Ecol. Evol.* **1995**, *10*, 198–202. [CrossRef]

37. Schaal, B.A.; Hayworth, D.A.; Olsen, K.M.; Rauscher, J.T.; Smith, W.A. Phylogeographic studies in plants: Problems and prospects. *Mol. Ecol.* **1998**, *7*, 465–474. [CrossRef]

38. 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]

39. Bryan, G.J.; McNicoll, J.; Ramsay, G.; Meyer, R.C.; De Jong, W.S. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. *Theor. Appl. Genet.* **1999**, *99*, 859–867. [CrossRef]

40. Ishii, T.; McCouch, S.R. Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. *Theor. Appl. Genet.* **2000**, *100*, 1257–1266. [CrossRef]

41. Chumley, T.W.; Palmer, J.D.; Mower, J.P.; Fourcade, H.M.; Calie, P.J.; Boore, J.L.; Jansen, R.K. The complete chloroplast genome sequence of *Pelargonium × hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* **2006**, *23*, 2175–2190. [CrossRef]

42. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.

43. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]

44. Jin, J.-J.; Yu, W.-B.; Yang, J.-B.; Song, Y.; DePamphilis, C.W.; Yi, T.-S.; Li, D.-Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2020**, *21*, 241. [CrossRef] [PubMed]

45. Qu, X.-J.; Moore, M.J.; Li, D.-Z.; Yi, T.-S. PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **2019**, *15*, 50. [CrossRef]

46. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]

47. 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]

48. Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [CrossRef] [PubMed]

49. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32* (Suppl. S2), W273–W279. [CrossRef] [PubMed]

50. 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]

51. 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]

52. 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] [PubMed]

53. Peden, J.F. Analysis of Codon Usage. Ph.D. Thesis, University of Nottingham, Nottingham, UK, 2000.

54. Zhang, D.; Gao, F.; Jakovlić, I.; Zou, H.; Zhang, J.; Li, W.X.; Wang, G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **2020**, *20*, 348–355. [CrossRef]

55. 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]

56. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]

57. Ronquist, F.; Teslenko, M.; Van Der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, 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]

58. Nylander, J.A.A. MrModeltest v2. Program distributed by the author, Volume 2. In *Evolutionary Biology Centre*; Uppsala University: Uppsala, Sweden, 2004.