

Article

# The Complete Chloroplast Genome of *Hypoestes forskolii* (Vahl) R.Br: Insights into Comparative and Phylogenetic Analyses within the Tribe Justiceae

Samaila Samaila Yaradua <sup>1,2,\*</sup> and Kowiyou Yessoufou <sup>1</sup>

<sup>1</sup> Department of Geography, Environmental Management and Energy Studies, APK Campus, University of Johannesburg, Johannesburg 2006, South Africa

<sup>2</sup> Department of Biology, Umaru Musa Yaradua University, Katsina 820102, Nigeria

\* Correspondence: dryaradua@gmail.com

**Abstract:** *Hypoestes forskolii* is one of the most important species of the family Acanthaceae, known for its high economic and medicinal importance. It is well distributed in the Arab region as well as on the African continent. Previous studies on ethnomedicine have reported that *H. forskolii* has an anti-parasitic effect as well as antimalarial and anthelmintic activities. Previous studies mainly focused on the ethnomedicinal properties, hence, there is no information on the genomic architecture and phylogenetic positions of the species within the tribe Justiceae. The tribe Justiceae is the most taxonomically difficult taxon in Acanthoideae due to its unresolved infratribal classification. Therefore, by sequencing the complete chloroplast genome (cp genome) of *H. forskolii*, we explored the evolutionary patterns of the cp genome and reconstructed the phylogeny of Justiceae. The cp genome is quadripartite and circular in structure and has a length of 151,142 bp. There are 130 genes (86 coding for protein, 36 coding for tRNA and 8 coding for rRNA) present in the plastome. Analyses of long repeats showed only three types of repeats: forward, palindromic and reverse were present in the genome. Microsatellites analysis revealed 134 microsatellites in the cp genome with mononucleotides having the highest frequency. Comparative analyses within Justiceae showed that genomes structure and gene contents were highly conserved but there is a slight distinction in the location of the genes in the inverted repeat and single copy junctions. Additionally, it was discovered that the cp genome includes variable hotspots that can be utilized as DNA barcodes and tools for determining evolutionary relationships in the Justiceae. These regions include: *atpH-atpI*, *trnK-rps16*, *atpB-rbcL*, *trnT-trnL*, *psbI-trnS*, *matK*, *trnH-psbA*, and *ndhD*. The Bayesian inference phylogenetic tree showed that *H. forskolii* is a sister to the *Dicliptera* clade and belongs to Diclipterinae. The result also confirms the polyphyly of *Justicia* and inclusion of Diclipterinae within justicioid. This research has revealed the phylogenetic position of *H. forskolii* and also reported the resources that can be used for evolutionary and phylogenetic studies of the species and the Justiceae.

**Keywords:** chloroplast genome; *Hypoestes forskolii*; justiceae; phylogenomics



**Citation:** Yaradua, S.S.; Yessoufou, K. The Complete Chloroplast Genome of *Hypoestes forskolii* (Vahl) R.Br: Insights into Comparative and Phylogenetic Analyses within the Tribe Justiceae. *Genes* **2022**, *13*, 2259. <https://doi.org/10.3390/genes13122259>

Academic Editors: Wajid Zaman, Hakim Manghwar and Zhiqiang Wu

Received: 10 November 2022

Accepted: 28 November 2022

Published: 30 November 2022

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



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

## 1. Introduction

*Hypoestes forskolii* (Vahl) R.Br is a herbaceous plant and one of the most important Acanthaceae species mainly distributed in some part of Africa and the Arab region [1]. The plant has been used in folk medicine in the treatment of various diseases like cancer, gonorrhea, heart and chest diseases [2,3]. The species is also reported to have several biological properties which includes antifungal, atrypanosomal, antileishmanial, antiplasmodial and cytotoxic properties [2,4]. A leaf decoction from the plant is used by Herdsmen to kill insect and parasite on sheep [5,6]. Fresh leaves of *H. forskolii* are used to heal wounds and are also reported to have insecticidal activity [7]. Despite the traditional medicinal value of the plant, its complete chloroplast genome (cp genome) has not been sequenced, thus the genome evolution remains unexplored.

The tribe Justiceae is the largest tribe in Acanthoideae (Acanthaceae) with ca. 2000 species distributed worldwide and exhibit a great diversity in morphological features. The tribe is considered the most taxonomically difficult taxon in Acanthoideae due to its unresolved tribal classification. The taxonomic problems in the tribe began after the infrafamilial classification of Acanthaceae by Lindau (in 1895). In his work, he classified *Justicia* L. and its assumed sister species in a tribe that is characterized by the presence of two stamens and two or three aperturate pollen. The problem of this classification is that he included species that don't have this character in Justiceae and placed other species with two stamen in other tribes. Since then, several studies [8–12] have been conducted to address the infratribal classification. Their work was based on morphological data, but they did not analyze their data phylogenetically. However, their work contributed immensely to providing knowledge of infrafamilial and infratribal relationships within the Acanthaceae. The major setback of the Bremekamp infrafamilial classification is that he did not account for the generic composition of his tribes [11]. McDade and her colleagues (in 2000) were the first students of Acanthaceae who employed phylogenetic analysis using the *nrITS* and the *trnL-F* region from 55 species to resolve the Justiceae tribal classification. Their results recognized four lineages and these lineages are characterized by tricolporate hexapseudocolpate pollen and also provided a base for evaluating the phylogenetic relationship of the Justiceae species. Since then, ref. [13] worked on *Isoglossinae*, refs. [14–16] worked on the Tetmerium lineage. The recent study that tries to evaluate the tribal classification of Justiceae, particularly the justicoid lineages is [17]. In their study, they used six DNA regions from 178 samples of Justiceae and concluded that *Justicia* is paraphyletic and called for further investigation to address the infratribal classification of the tribe.

Understanding the evolutionary links among plant species at various taxonomic hierarchies can be achieved with the help of genetic data. This is due to the presence of various protein-coding and transfer RNA genes that play a crucial role in the cells of plants. The major role of the chloroplast organelle in a plant is the conversion of carbon dioxide and water into simple sugars [18]. The chloroplast genome (cp genome) is the most informative and conserved genome in angiosperms compared with nuclear and mitochondria. The cp genome contains sufficient information for inferring phylogenetic relationships, studies of taxa diversification and comparative analyses [19]. The cp genome of land plants usually have a circular and a quadripartite structure which include a single copy region (70–80 kb), small single copy region (15–28 kb) and a pair of inverted repeat (18–30 kb) [20]. The length of the cp genome of angiosperms ranges from 140 kb to 180 kb and they contain about 132 unique genes (ca. 86 coding for protein, 4 coding for ribosomal ribonucleic acid and 32 coding for transfer ribonucleic acid) [21]. Despite the conserved nature of the cp genome, there is some degree of variation in gene content and genome structure as well as mutations [22]. This causes the sequences of different species to diverge, which may be utilized to analyse the phylogenetic relationships of various taxonomic groups [23]. Only the chloroplast genomes of six taxa of the Justiceae have been sequenced, despite the significance of the cp genomes in contemporary systematics.

Here, we have reported and characterized the cp genome of *H. forskaolii* using the Illumina Hiseq 2500 system. We have combined the newly sequenced *H. forskaolii* with the 12 cp genomes of Justiceae sampled from the public Genbank database to explore the evolutionary circumstance of Justiceae. We have also analyzed the structure of the cp genome with other Justiceae to identify regions to be used for the species authentication and genetic diversity studies of the species including population genetics studies. We have reconstructed the phylogenetic relationship of the species using the plastome sequences to infer its phylogenetic positions and attempt to resolve the infratribal classification of Justiceae.

## 2. Materials and Methods

A typical specimen of *H. forskaolii* was collected from Batagarawa town, Katsina, Katsina State, Nigeria. The leaves from the specimen were collected and stored in a zip lock

bag containing silica gel and the specimen was taken to Herbarium Center for Biodiversity and Conservation, Umaru Musa Yaradua University for identification. The specimen was authenticated by the Curator and was given accession number UMH0145. The dried leaf material was used to extract the DNA using a Qiagen genomic DNA extraction kit, following the manufacturer's instructions.

### 2.1. Library Construction, Sequencing and Assembly

For the DNA sample preparations, 1.0 µg of DNA was used as input material. Following the manufacturer's instructions, the NEBNext DNA Library Prep Kit was used to create the sequencing libraries, followed by addition of indices to the sample. The genomic DNA was sheared at random into segments of 350 bp, which were then prepared for further PCR enrichment and sequencing. The resulting PCR products were subject to purification (using the AMPure XP system), then the resultant libraries were measured using real-time PCR and evaluated for size distribution using an Agilent 2100 Bio analyzer. After pooling, the qualified libraries based on projected data volume and effective concentration, were fed onto Illumina sequencers. Using PRINSEQlite v0.20.4 [24], clean reads sequences (5.2 GB) were obtained from the raw data through filtering, which were then subjected to assembly using NOVOPlasty 4.2. [25] using kmer (K-mer = 39) and the cp genome of *Justicia flava* (NC 044862.1) was used as the seed and reference.

### 2.2. Gene Annotation

The online tool GeSeq [26] was used to annotate all the genes present in the cp genome of *H. forskolii* using the plastome sequence of *Justicia flava* as reference. The sequin program was used to correct the errors in the genes that were not successfully annotated in the GeSeq by adjusting the position of the codons. The cp genome architecture was drawn in OGDRAW (Organelle Genome DRAW) [26].

### 2.3. Codon Usage Analysis

The software MEGA version 6 [27] was used to calculate the relative synonymous codon usage values (RSCU), codon usage and base composition.

### 2.4. Repeat Analysis

MicroSATellite (MISA) (<http://webblast.ipk-gatersleben.de/misa/index.php> accessed on 24 November 2022) was used to identify Simple Sequence Repeats (SSRs) in the *H. forskolii* cp genome [28]. For mononucleotides, dinucleotides, trinucleotides, and tetra, penta, and hexa nucleotides SSR motifs eight, five, four, and three repetitions units were used, respectively. The online software REPuter [29] was used to identify the types of long repeats present in the cp genome.

### 2.5. Genome Comparison

Using the annotation of *H. forskolii* as a reference in the Shuffle-LAGAN mode [30], the cp genomes of seven species of Justiceae were compared using the tool mVISTA [31]. The online software IR scope (<https://irscope.shinyapps.io/irapp/> accessed on 1 November 2022) was used to compare the border regions of the cp genome.

### 2.6. Characterization of Substitution Rate

To identify the genes that are under selection pressure, DNAsp version 6 [32] was used to examine synonymous (dS) and nonsynonymous (dN) substitution rate and dN/dS ratio.

### 2.7. Sequence Divergence

To determine sequence divergence and identify the variable regions in the cp genome, sliding window analysis was used to evaluate the diversity of nucleotide in the cp genome using DnaSP version 6 with the following parameters: 600 bp for window length and 200 bp for step size.

## 2.8. Phylogenetic Analysis

The cp genome of seventeen Acanthoideae species as well as four species, *Erythranthe lutea* (Phrymaceae), *Scrophularia dentata* (Scrophulariaceae), *Lysionotus pauciflorus* (Gesneriaceae) and *Tanaecium tetragonolobum* (Bignoniaceae) were downloaded from Genbank. The Acanthoideae species are the in group while the other four were used as the out group. Using MAFFT version 7 [33], the newly sequenced cp genome of *H. forskaolii* was aligned with the downloaded cp genomes. Using MrBayes version 3.2.6, a Bayesian inference strategy was used to reconstruct the phylogenetic tree [34]. The appropriate model was chosen using jModelTest version 3.7 [35].

## 3. Results and Discussion

### 3.1. Characteristics of *H. forskaolii* Chloroplast Genome

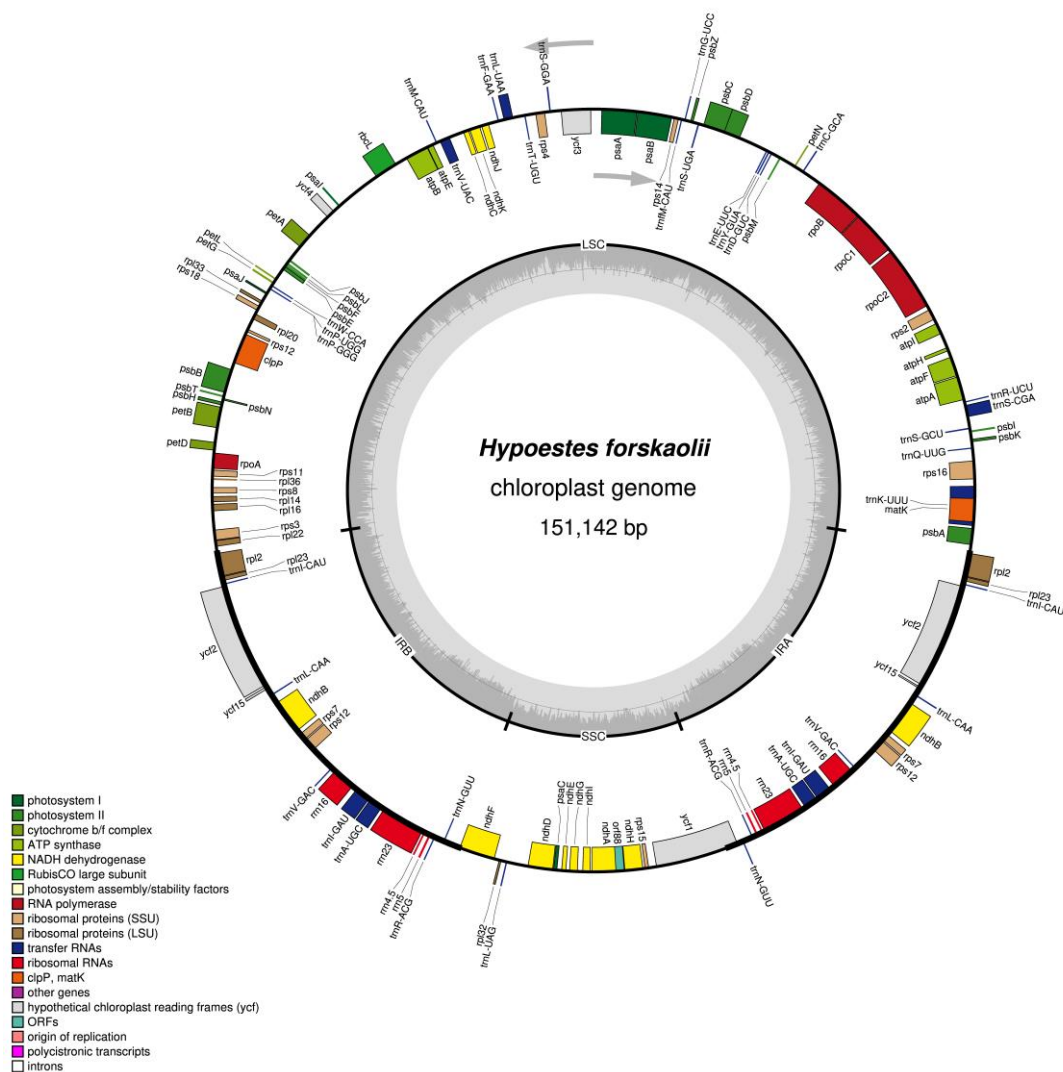
Previous findings revealed that the cp genome of flowering plants is extremely well-preserved in both the gene content and structural organization; however, expansion and contraction in the inverted repeat and single copy junctions are among the evolutionary events leading to variations in the cp genomes [36,37]. The cp genome of *H. forskaolii* is quadripartite and circular in structure and has a total length of 151,142 bp. The cp genome consists of a pair of inverted repeats (IRb and IRa) 25,477 bp, Small Single Copy (SSC) 17,012 and Large Single Copy 83,176 bp (Figure 1). A total of 72,852 bp make up the genome's non-coding region which is comprised of introns and intergenic spacers, and the 78,290 bp code for protein genes. The GC percentage of the LSC and SSC regions was 36.0% and 32.3%, respectively, whereas the GC content of the inverted repeats IRa and IRb was 43.4% and 43.3%, respectively (Table 1). It is discovered that the IRs have a larger proportion of GC than SSC and LSC regions.

**Table 1.** Nucleotide composition in the *H. forskaolii* cp genome.

Region	T(U) (%)	C (%)	A (%)	G (%)	Total (bp)
cp genome	31.2	19.4	30.8	18.7	151,142
LSC	32.4	18.5	31.16	17.5	83,176
SSC	34.0	16.8	34.0	15.2	17,012
IRA	28.4	22.5	28.3	20.9	25,477
IRB	28.2	20.8	28.4	22.5	25,477
1st Position	31.2	19.4	30.1	19.0	50,381
2nd Position	31.0	19.4	31.3	18.3	50,381
3rd Position	31.0	19.2	30.9	18.6	50,381

The annotation of *H. forskaolii* cp genome revealed a total of 130 genes (86 protein-coding genes, 36 tRNAs genes and 8 rRNAs genes), 96 genes are present in the LSC (82 protein coding genes and 21 tRNA genes) and SSC (14 protein coding genes and 1 tRNA) while the remaining 17 genes (7 tRNAs, 4 rRNAs and 7 protein-coding genes) are repeated in the IRa and IRb regions (Table 2 and Figure 1). Almost all the protein-coding genes in the cp genome have ATG as their start codon and few of them have alternative start codon, GTG, ACG and ATC; this has been reported to occur in the cp genome of angiosperms [38–40].

Some of the coding genes present in the cp genome of *H. forskaolii* have introns. Introns are reported to occur in some of the protein-coding and tRNAs genes of flowering plants cp genomes [38,39]. Out of the 130 coding genes, 16 are characterized with one or two introns (Table 3). Among the 16 genes, six are tRNAs and 11 are protein coding genes. Ten of the intron-containing genes are located in the LSC, one gene in the SSC while the remaining five are in the inverted repeat region. ATP dependent protease subunit p gene (*clpP*) and one of the Photosystem I gene (*ycf3*) possessed two introns while the remaining 14 genes have only one intron. The tRNA, *trnK-UUU* is the gene with longest intron which is due to the inclusion of *matK* in the gene.



**Figure 1.** The structure of the *H. forskaoilii* cp genome. Genes inside the circles are transcribed clockwise, while those outside the circles are transcribed counterclockwise. The colorful bar displays genes that are known to be functioning. The inner circle's dark grey and light grey colors, respectively, designate the GC and AT contents.

The frequency of a codon that encodes for a specific amino acid was compared using codon usage analysis [41]. A codon is a sequence of trinucleotides that encodes for specific amino acids that are used in protein synthesis [42]. Because of bias in mutation, codon use is a factor influencing the development of the chloroplast genome [43] and it differs between species [44]. The nucleotide sequence of the protein-coding genes (78,290 bp) was used to calculate the frequency of the codons present in the cp genome. The relatively synonymous codon usage (RSCU) of the genes in the cp genome is shown in (Table 4, Figure 2). The codon usage analysis revealed that 26,095 codons encode for the genes in the cp genome. All the 20 known amino acids are encoded by 61 codons (Figure 2). Codons coding for the amino acid leucine are more prevalent in the cp genome while codons that code for Cysteine are the less common (Table 4). The Cytosine (C) and Guanine (G) endings are more prevalent than the Thymine (T) and Adenine (A) endings; the cp genomes of other angiosperms have Thymine and Adenine endings occurring more frequently [45–47]. The result of the analysis (Table 4) revealed that codon usage bias is low in the cp genome of *H. forskaoilii*. The RSCU values for 31 codons were less than 1, and are characterized with G/C endings, whereas the RSCU values for 30 codons were greater than 1, with A/T

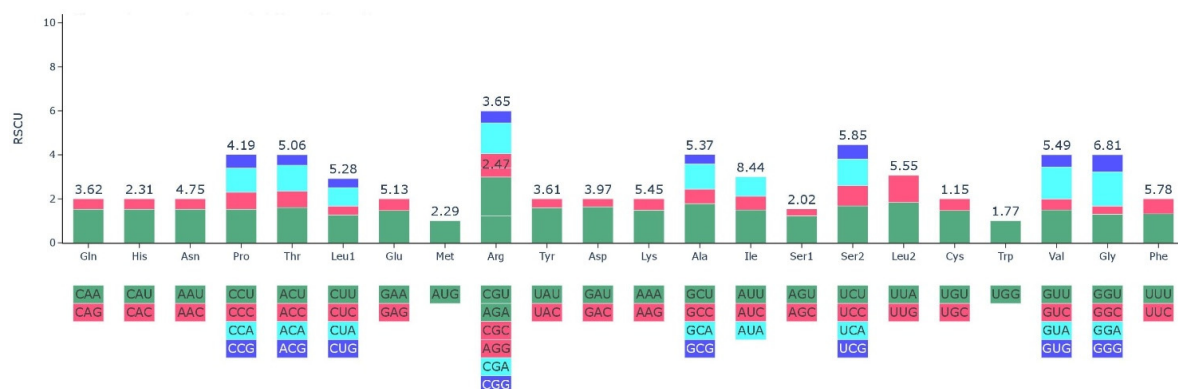


endings. Methionine and Tryptophan are the only two amino acids without codon bias with RSCU values of 1.

**Table 2.** Genes present in the chloroplast genome of *H. forskali*.

Category	Group of Genes	Name of Genes
RNA genes	ribosomal RNA genes (rRNA)	<i>rrn5, rrn4.5, rrn16, rrn23</i>
	Transfer RNA genes (tRNA)	<i>trnL-UAG</i> , <i>trnK-UUU</i> <sup>+</sup> , <i>trnS-GCU</i> , <i>trnQ-UUG</i> , <i>trnV-UAC</i> <sup>+</sup> , <i>trnR-UCU</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnC-GCA</i> , <i>trnM-CAU</i> , <i>trnY-GUA</i> , <i>trnT-GGU</i> , <i>trnG-GCC</i> , <i>trnI-M-CAU</i> , <i>trnS-UGA</i> , <i>trnT-UGU</i> , <i>trnL-UAA</i> <sup>+</sup> , <i>trnF-GAA</i> , <i>trnW-CCA</i> , <i>trnP-UGG</i> , <i>trnP-GGG</i> , <i>trnL-CAA</i> <sup>a</sup> , <i>trnV-GAC</i> <sup>a</sup> , <i>trnI-GAU</i> <sup>+,a</sup> , <i>trnA-UGC</i> <sup>+,a</sup> , <i>trnR-ACG</i> <sup>a</sup> , <i>trnN-GUU</i> <sup>a</sup> , <i>trnH-GUG</i> , <i>trnG-UCC</i> , <i>trnS-GGA</i>
Ribosomal proteins	Small subunit of ribosome	<i>rps19, rps8, rps14, rps7</i> <sup>a</sup> , <i>rps2, rps11, rps2, rps14, rps18, rps16</i> <sup>+</sup> , <i>rps15, rps12</i> <sup>a</sup>
Transcription	Large subunit of ribosome	<i>rpl33, rpl16, rpl20, rpl32, rpl23</i> <sup>a</sup> , <i>rpl22, rpl14, rpl36, rpl2</i> <sup>+,a</sup>
	DNA dependent RNA polymerase	<i>rpoC2, rpoB rpoA, rpoC1</i> <sup>+</sup> ,
Protein genes Other genes	Photosystem I	<i>psaI, psaB, psaj, psaA, psaC, ycf3</i> <sup>++</sup>
	Photosystem II	<i>psbZ, psbK, psbF, psbD, psbL, psbN, psbH, psbT, psbJ, psbB, psbE, psbM, psbC, psbI, psbA</i>
	Subunit of cytochrome	<i>petL, petB, petN, petG, petA, petD</i>
	Subunit of synthase	<i>atpH, atpF, atpE, atpB</i> <sup>+</sup> , <i>atpA, atpI</i>
	Chloroplast envelope membrane protein	<i>cemA</i>
	NADH dehydrogenase	<i>ndhK, ndhH, ndhI, ndhD, ndhB</i> <sup>+,a</sup> <i>ndhG, ndhF, ndhE, ndhC, ndhJ, ndhA</i> <sup>+</sup>
	Large subunit of rubisco	<i>rbcL</i>
	Subunit acetyl-coA carboxylase	<i>accD</i>
	ATP dependent protease subunit P	<i>clpP</i> <sup>++</sup>
	Maturase	<i>matK</i>
	C-type cytochrome synthesis	<i>ccsA</i>
	Translational initiation factor	<i>infA</i>
	Component of TIC complex	<i>ycf1</i> <sup>a</sup>
	Hypothetical proteins	<i>ycf2</i> <sup>a</sup> , <i>ycf4, ycf15</i> <sup>a</sup>

<sup>a</sup> Duplicated genes, <sup>+</sup> Gene with one intron and <sup>++</sup> Gene with two introns.



**Figure 2.** The 20 amino acids and stop codons' relative synonymous codon usage (RSCU) in the chloroplast genome of *H. forskali*'s protein-coding genes.

**Table 3.** Genes with introns in the *H. forskaolii* chloroplast genome and length of introns and exons.

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
<i>trnK-UUU</i>	LSC	36	2453	36		
<i>rps16</i>	LSC	224	912	35		
<i>trnS-CGA</i>	LSC	31	661	59		
<i>atpF</i>	LSC	470	649	140		
<i>rpoC1</i>	LSC	1635	784	433		
<i>ycf3</i>	LSC	152	710	228	683	127
<i>trnL-UAA</i>	LSC	36	513	49		
<i>trnV-UAC</i>	LSC	36	584	37		
<i>clpP</i>	LSC	226	631	299	749	69
<i>petB</i>	LSC	5	687	653		
<i>rpl2</i>	IR	435	658	397		
<i>ndhB</i>	IR	755	679	776		
<i>rps12</i>	IR	25	542	231		
<i>trnI-GAU</i>	IR	41	939	34		
<i>trnA-UGC</i>	IR	37	820	34		
<i>ndhA</i>	SSC	538	989	552		

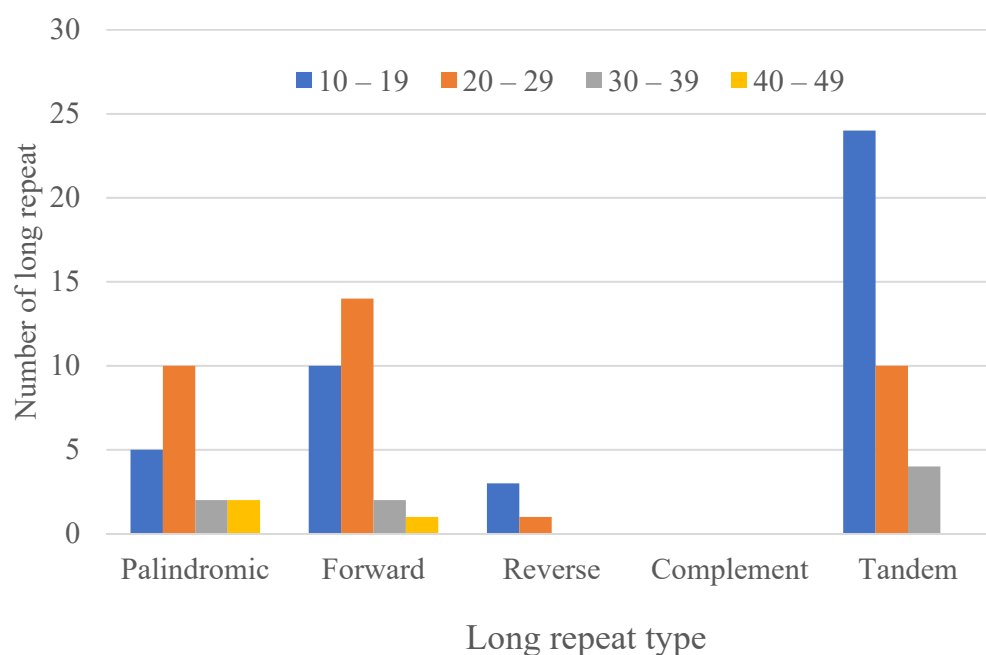
**Table 4.** Codon usage of the *H. forskaolii* chloroplast genome.

Codon	Amino Acid	RSCU	tRNA	Codon	Amino Acid	RSCU	tRNA
UUC	Phe	0.67	<i>trnF-GAA</i>	UAU	Tyr	1.61	<i>trnY-GUA</i>
UUU	Phe	1.33		UAC	Tyr	0.39	
CUG	Leu	0.42		CAG	Gln	0.47	
UUA	Leu	1.84	<i>trnL-UAA</i>	UAA	Stop	1.55	
CUA	Leu	0.83		CAA	Gln	1.53	<i>trnQ-UUG</i>
CUU	Leu	1.27	<i>trnL-UAG</i>	CAU	His	1.54	<i>trnH-GUG</i>
CUC	Leu	0.4		CAC	His	0.46	
UUG	Leu	1.24	<i>trnL-CAA</i>	UAG	Stop	0.93	
AUG	Met	1	<i>trnM-CAU</i>	AAG	Lys	0.51	
AUU	Ile	1.5	<i>trnI-GAU</i>	AAU	Asn	1.54	<i>trnN-GUU</i>
GUC	Val	0.49		GAC	Asp	0.36	
AUA	Ile	0.89	<i>trnI-CAU</i>	AAA	Lys	1.49	<i>trnK-UUU</i>
AUC	Ile	0.61		AAC	Asn	0.46	
GUU	Val	1.5	<i>trnV-GAC</i>	GAU	Asp	1.64	<i>trnD-GUC</i>
UCU	Ser	1.68	<i>trnS-GGA</i>	UGU	Cys	1.46	<i>trnC-GCA</i>
GUA	Val	1.45		GAA	Glu	1.47	<i>trnE-UUC</i>
GUG	Val	0.56	<i>trnV-UAC</i>	GAG	Glu	0.53	
UCA	Ser	1.21		UGA	Stop	0.52	
UCC	Ser	0.92		UGC	Cys	0.54	
CCA	Pro	1.1		CGA	Arg	1.41	
UCG	Ser	0.67	<i>trnS-UGA</i>	UGG	Trp	1	<i>trnW-CCA</i>
CCU	Pro	1.51	<i>trnP-UGG</i>	CGU	Arg	1.2	<i>trnR-ACG</i>
CCC	Pro	0.78		CGC	Arg	0.41	<i>trnR-UCU</i>
GCA	Ala	1.14		GGA	Gly	1.58	
CCG	Pro	0.61		CGG	Arg	0.55	
ACU	Thr	1.61		AGA	Arg	1.78	
ACC	Thr	0.73		AGG	Arg	0.65	
GCC	Ala	0.67		GGC	Gly	0.37	
ACG	Thr	0.47	<i>trnT-UGU</i>	AGC	Ser	0.3	
GCU	Ala	1.75	<i>trnA-UGC</i>	GGU	Gly	1.27	
ACA	Thr	1.19	<i>trnT-GGU</i>	AGU	Ser	1.22	<i>trnS-GCU</i>
GCG	Ala	0.44		GGG	Gly	0.78	<i>trnG-UCC</i>

### 3.2. Repeat Analyses

#### 3.2.1. Long Repeats

Long repeats sequences present in the cp genome of *H. forskaolii* were identified using the program REPuter; from the results it was discovered that tandem repeats and three types of long repeats (forward, reverse and palindromic) were present in the plastome *H. forskaolii* (Figure 3). In total, there are 88 repeats in the cp genome of *H. forskaolii* (19 palindromic repeats, 27 forward repeats, 4 reverse repeats and 38 tandem repeats). Most of the palindromic and forward repeats sizes are between 20–29 bp, followed 10–19 bp. The length of repeated sequences in *H. forskaolii* cp genome ranges from 10 to 44 bp, are analogous to the lengths in other angiosperm cp genomes [48–50].



**Figure 3.** Types of tandem and long repeat distribution and their frequency in the chloroplast genomes of *H. forskaolii*.

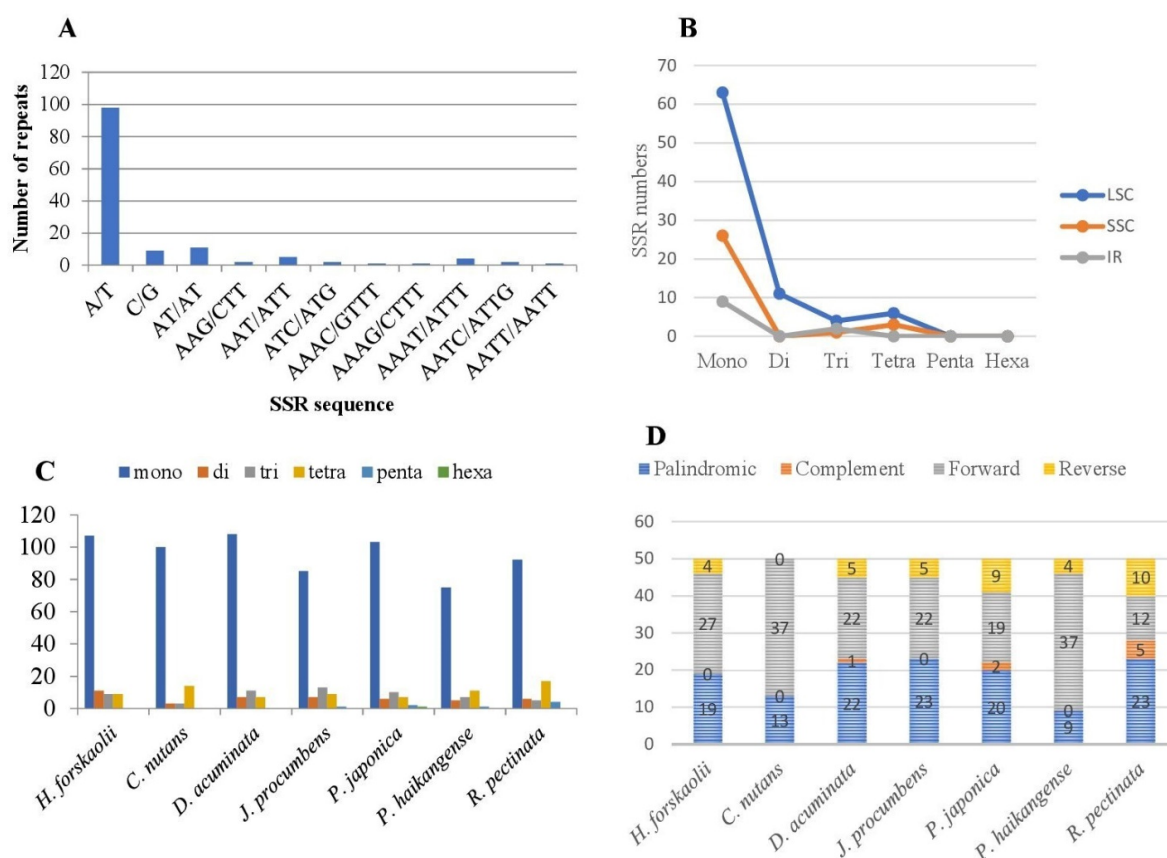
We compared the frequency of repeats among seven Justiceae cp genomes, and the result indicated that only three species, namely, *D. acuminata*, *P. japonica* and *R. pectinata* contained complement, forward, reverse and palindromic repeats (Figure 4D). *C. nutans* and *P. haikangense* have the highest frequency of forward repeats (37) while *R. pectinata* has the lowest (12). *D. acuminata* and *J. procumbens* have the same number of reverse repeats, five each. *R. pectinata* has the highest number of reverse repeats (9) while *H. forskaolii* has the lowest (4). Complement repeats are found to be the less numerous type of repeat across the genome with *D. acuminata*, *P. japonica* and *R. pectinata* having one, two and five, respectively.

#### 3.2.2. Simple Sequence Repeats (SSRs)

Simple sequence repeats (SSRs) are short repeats of sequences usually 1–6 bp that are very useful at evaluating genetic variation among species. These SSRs are present in cp genomes of angiosperms and are uniparentally inherited. They are therefore employed as molecular markers in developmental research such as genetic heterogeneity investigations, and they also aid in the identification of species [51–53]. This study discovered 136 microsatellites in the cp genome of *H. forskaolii* (Table 5). Mononucleotides are the most frequent SSRs in the cp genome, constituting about 78.67%, of which majority are polythymine (41.91%) and polyadenine (30.14%); this is consistent with previous studies [53]. Among the dinucleotide only AT/AT is found in the genome. Reflecting series complementary, only two trinucleotide AAT/ATT and ATC/ATG, five tetra AAAC/GTTT, AAAG/CTTT, AAAT/ATTT, AATC/ATTG, AATT/AATT were present in the cp genome.



Penta nucleotide and hexa nucleotide were not discovered in the cp genome (Figure 4A). The LSC region harbored most of the microsatellites, followed by SSC (Figure 4B).



**Figure 4.** Repeat analyses of the sampled plastome of Justiceae. (A): Frequency of different SSR motifs in different repeat types in *H. forskoolii* chloroplast genome. (B): Distribution of SSR in LSC, SSC, and IR regions. (C): Number of different SSR types in the chloroplast genome of seven Justiceae. (D): Number of different repeats in chloroplast genomes of seven species of Justiceae.

The frequency of SSRs among the cp genomes of seven species of Justiceae was analyzed (Figure 4C); the results revealed that mononucleotides repeats are the most frequent across all the cp genomes. *D. acuminata* and *H. forskoolii* are the species with the highest frequency of mononucleotide; 108 and 107, respectively. Pentanucleotides were not present in the cp genomes of *H. forskoolii*, *D. nutans* and *D. acuminata* while hexanucleotide was only present in *P. japonica*.

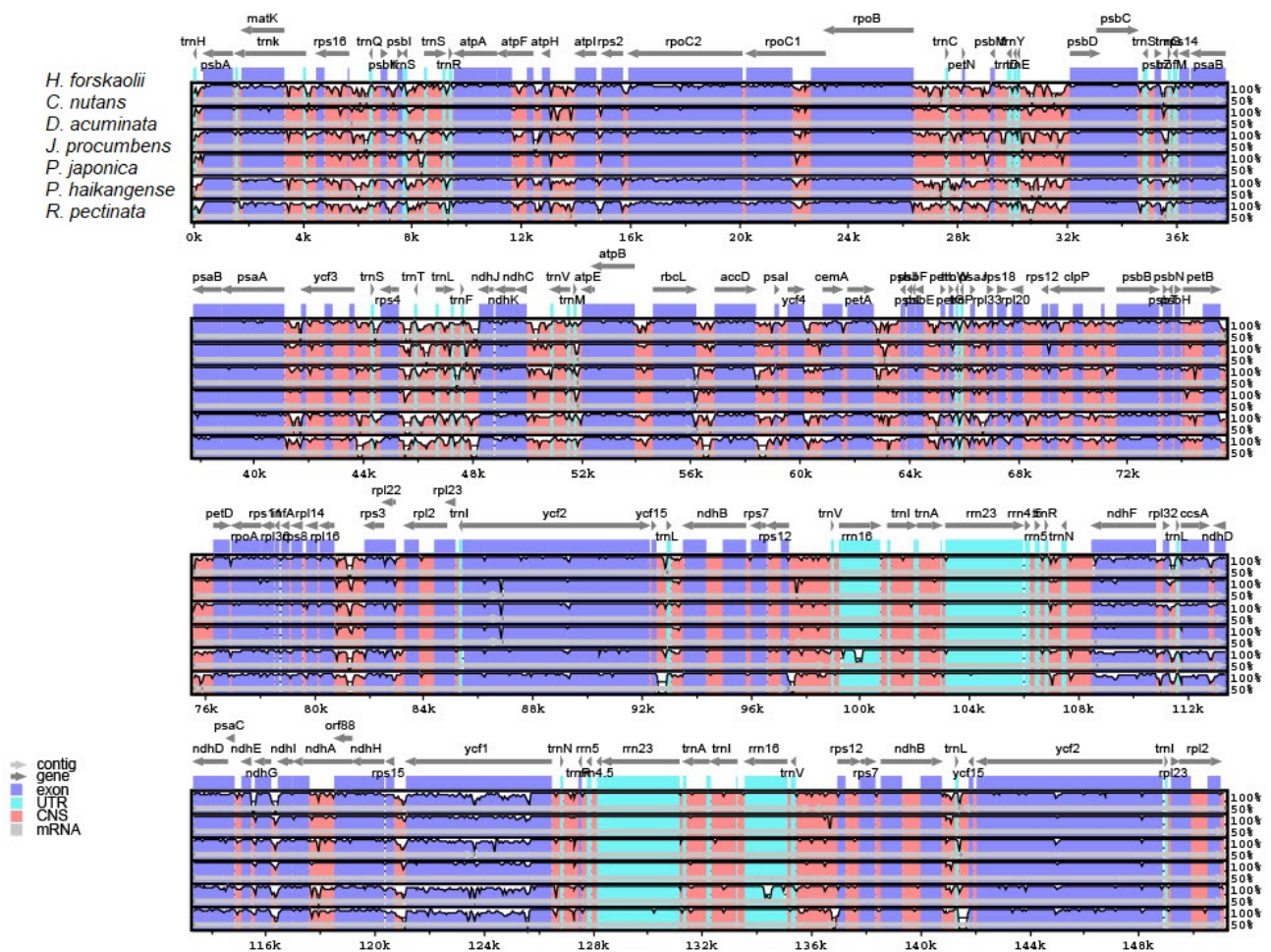
### 3.3. Comparative Analysis of Justiceae Species Cp Genome

To evaluate the level of genome divergence in Justiceae, the newly sequenced cp genome *H. forskoolii* was compared with six Justiceae species cp genomes downloaded from the GenBank. The cp genomes were aligned using mVISTA with the annotation of *H. forskoolii* as a reference. The result of the analysis revealed that the compared cp genomes are well-preserved in terms of genome structure and gene content; however, there was some level of variation. The protein-coding regions were found to be more conserved than the introns and intergenic spacers. In terms of the four regions, the IRa and IRb were more conserved than the SSC and LSC (Figure 5). This has been reported to occur in certain taxa cp genomes in earlier studies [54,55]. The most divergent non-coding regions among the seven cp genomes are *trnH-psbA*, *trnK-rps16*, *rps16-trnQ-UUG*, *trnE-UUC-psbD*, *atpH-atpI*, *trnT-trnL*, *ndhC-trnV*, *accD-psaI*, *petA-psbJ*, *atpB-rbcL*, *rps12* and *trnL-rpl32*. A slight sequence variation was observed in the following genes *psbM*, *matK*, *ycf1*, *trnA-UGC*, *ndhH*, and *rrn16*. These regions can be used as a source of potential barcode

for identification/authentication of Justiceae species as well as resources for inferring phylogenetic relationships of the Acanthoideae.

**Table 5.** Microsatellite in the chloroplast genome of *H. forskoolii*.

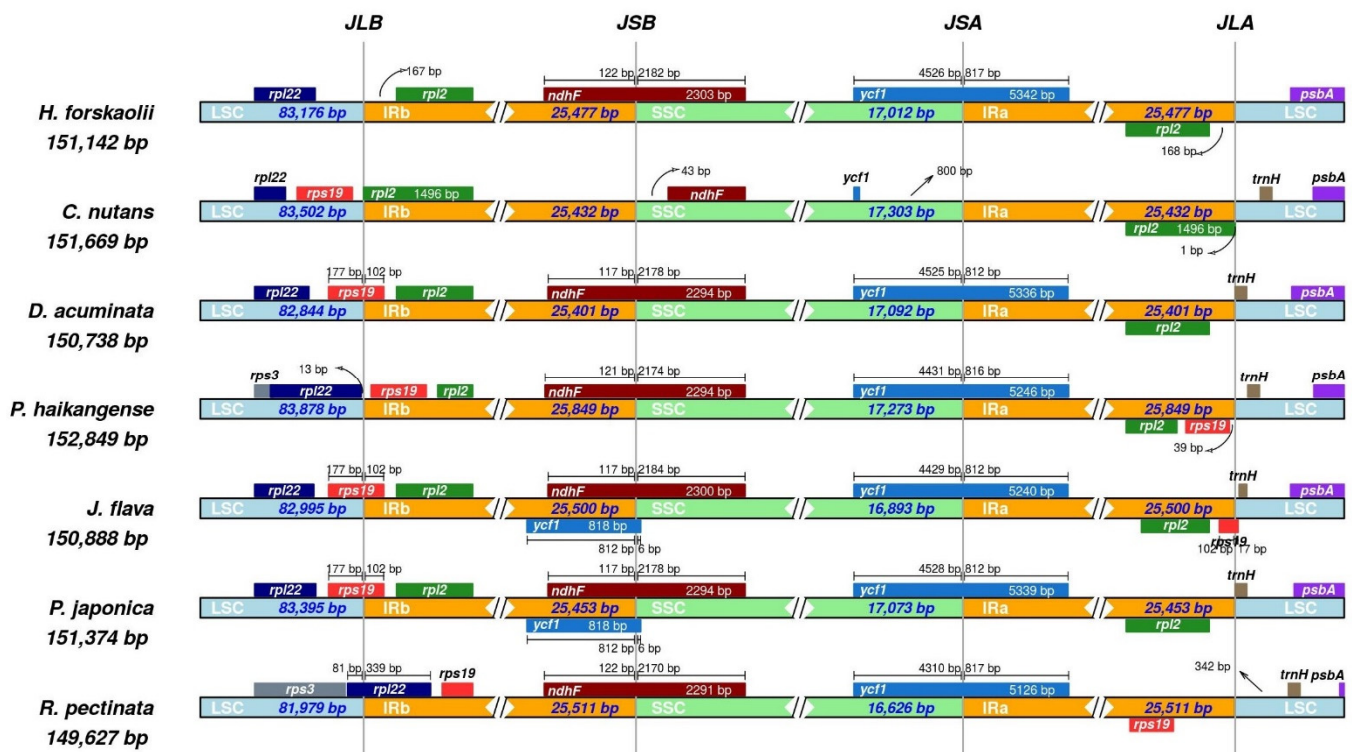
cpSSR ID	Repeat Motif	Length (bp)	No. of Repeats	SSR Start Position
1	(A) 8	8	23	4087; 8109; 9324; 14,945; 15,707; 18,082; 22,395; 41,631; 64,140; 80,149; 95,787; 110,230; 111,185; 112,039; 112,263; 114,062; 114,556; 118,479; 125,329; 136,542; 151,007; 151,074
2	(A) 9	9	9	7444; 11,725; 26,746; 43,488; 43,722; 66,819; 73,976; 88,658; 151,041
3	(A) 10	10	5	7650; 12,601; 14,734; 27,880; 132,799
4	(A) 12	12	1	15,729
5	(A) 13	13	1	7927
6	(A) 16	16	1	127,320
7	(A) 17	17	1	13,063
8	(C) 11	11	1	4477
9	(C) 12	12	1	131,580
10	(G) 8	8	2	4422; 57,979
11	(G) 9	9	1	5964
12	(G) 10	10	1	116,259
13	(G) 11	11	1	74,737
14	(G) 12	12	1	102,700
15	(G) 16	16	1	66,828
16	(T) 8	8	27	7431; 8012; 25,690; 30,519; 32,669; 35,551; 42,497; 45,300; 50,150; 59,760; 72,331; 74,672; 75,843; 81,161; 82,593; 82,893; 83,210; 83,277; 97,742; 109,264; 111,758; 112,374; 113,566; 123,714; 123,762; 123,859; 138,497
17	(T) 9	9	14	16,042; 17,941; 28,559; 31,561; 42,560; 65,592; 77,186; 83,242; 109,326; 115,581; 123,423; 123,827; 123,846; 145,625
18	(T) 10	10	9	11,804; 12,514; 30,933; 53,860; 68,645; 70,546; 101,483; 123,976; 125,541
19	(T) 11	11	3	59,201; 123,498; 123,792;
20	(T) 12	12	2	54,320; 124,361
21	(T) 13	13	1	9522
22	(T) 16	16	1	106,956
23	(AT) 5	5	2	20,342; 46,477
24	AT (7)	7	2	7285; 13,415
25	AT (9)	9	1	75,422
26	(TA) 5	5	3	19,311; 45,521; 45,967
27	TA (6)	6	3	30,631; 45,576; 82,952
28	(ATA) 4	4	2	64,698; 149,097
29	(TTC) 4	4	1	34,430
30	(AAT) 4	4	1	63,219
31	(TTA) 4	4	1	81,350
32	(TAT) 4	4	1	85,183
33	(TGA) 4	4	1	90,439
34	(TCT) 5	5	1	124,766
35	(ATC) 4	4	1	143,840
36	(ATAA) 3	3	1	45,662
37	(TAAA) 3	3	1	66,210
38	(AAAC) 3	3	1	67,218
39	(AAAT) 3	3	1	74,480
40	(AATA) 3	3	1	113,175
41	(AATC) 3	3	1	118,310
42	(AATT) 3	3	1	122,706



**Figure 5.** Variable regions in the cp genome of seven Justiceae species. The top arrow represents the direction of transcription; the colors blue and pink denote protein coding and conserved non-coding sequence, respectively; light green denotes tRNAs and rRNAs. The cp genome coordinates are shown by the x-axis, and the percentage identity ranges from 50% to 100% on the y-axis.

The structure and size of the chloroplast genome is often retained by angiosperms [53]; however, due to evolutionary processes, including genome contraction and expansion, there can be subtle variations in the size and location of the boundaries of inverted repeats and single copy regions [56,57]. We compared the JLB, JSB, JSA and JLA boundaries of the seven cp genomes of Justiceae and the results (Figure 6) showed some degree of similarity and variation among the compared cp genomes. The length of the seven cp genomes ranged from 149,627 bp (*R. pectinata*) to 152,849 bp (*P. haikangense*). Three species *J. flava*, *P. japonica* and *D. acuminata* have their JLB within the gene *rps19* with 102 bp overlapping into the IR. The JLB was bordered by *rps19* (LSC) and *rpl2* in *C. nutans* due to the contraction of the IR. In contrast, due to the IR expansion, the JLB of the *R. pectinata* was located within the *rpl22* while *P. haikangense* was between the *rpl22* (LSC) and *rps19* (IR). The JSB was found within *ndhF* with 2170–2184 bp overlapping into the IR except for the *C. nutans*. The *trnH* is located at the junction of the LSC/IRa border of all the compared cp genomes with the exception of *H. forskoolii*, which might be due to an expansion which leads to the loss of *rps19*. The cp genome of *C. nutans* varied with the other cp genomes by having the *ndhF* gene located in the SSC region. The *ycf1* pseudogene extended through the SSC and IRa with about 4300 bp in the SSC and 812 bp in the IRa in all the compared genomes except for *C. nutans*. The JSA and JSB had the most conserved borders among the compared cp genomes. The cp genome of *R. pectinata* was unique by having *rps3* in LSC/IRb border. The cp genome of *R. pectinata* had the smallest LSC region 81,979 bp while *P. haikangense* had

the longest 83,878 bp. *H. forskalii* was the only species that lost the *rps19*, which is located LSC/IR border.



**Figure 6.** Structural variation in the junction of inverted repeat and single copy regions among the seven cp genomes of Justiceae. (JSA: Junction of the SSC and the IRA; JLB: Junction of the LSC and the IRb; JSB: Junction of the SSC and the IRb).

### 3.4. Divergence of Protein Coding Genes Sequence

To determine the genes that were undergoing selective pressure in the cp genome of *H. forskalii*, DNAsp was used to calculate the dN/dS ratio, nonsynonymous (dN) and synonymous (dS) rates. The results showed that the dN/dS ratio was less than one in almost all of the paired genes except *atpF* and *clpP* (Figure 7), indicating that most of the genes were under negative selection with the exception of *clpP* and *atpF*. The synonymous (dS) values range from 0.0053 to 0.1628 in all the protein-coding genes (Figure 7). Twenty-nine genes, including *ycf15*, *ycf3*, *rps18*, *rps14*, *rpl36*, *rpl32*, *psbM*, *psbN*, *psbJ*, *psaI*, *petG*, *ndhJ*, *ndhC*, *infA* and *atpH* showed no nonsynonymous changes occurring in the plastome of *H. forskalii*; comparable findings were reported for other cp genomes [58–60].

### 3.5. Identification of Sequence Divergence

Variable regions of the chloroplast genome are very useful in inferring phylogenetic relationships and identification of species at the lowest taxonomic rank. These regions also play a vital role in providing information that helps in detecting differences between species and revealing the changes in the population structure [61,62]. The plastome sequence of *H. forskalii* was found to be similar with that of related Justiceae species. The calculated pi values ranged from 0 to 0.79 (Figure 8), which shows slightly variation among the chloroplast genome and are relatively conserved. This pattern of variation was reported to occur in the plastomes of angiosperm [40]. Comparing the sequence divergence in the single copy regions and the inverted repeat region, the single copy region showed a higher variability than the inverted repeat. Highly variable protein-coding genes in the plastomes include *psbM*, *ndhH* and *ycf1*. Similarly, two intergenic spacers, *trnE-UUC-psbD* and *rps16-trnQ-UUG* were found to be highly variable. These results agree with the



mVISTA divergence analysis and show that the regions could be used in the identification and authentication of Justiceae species.

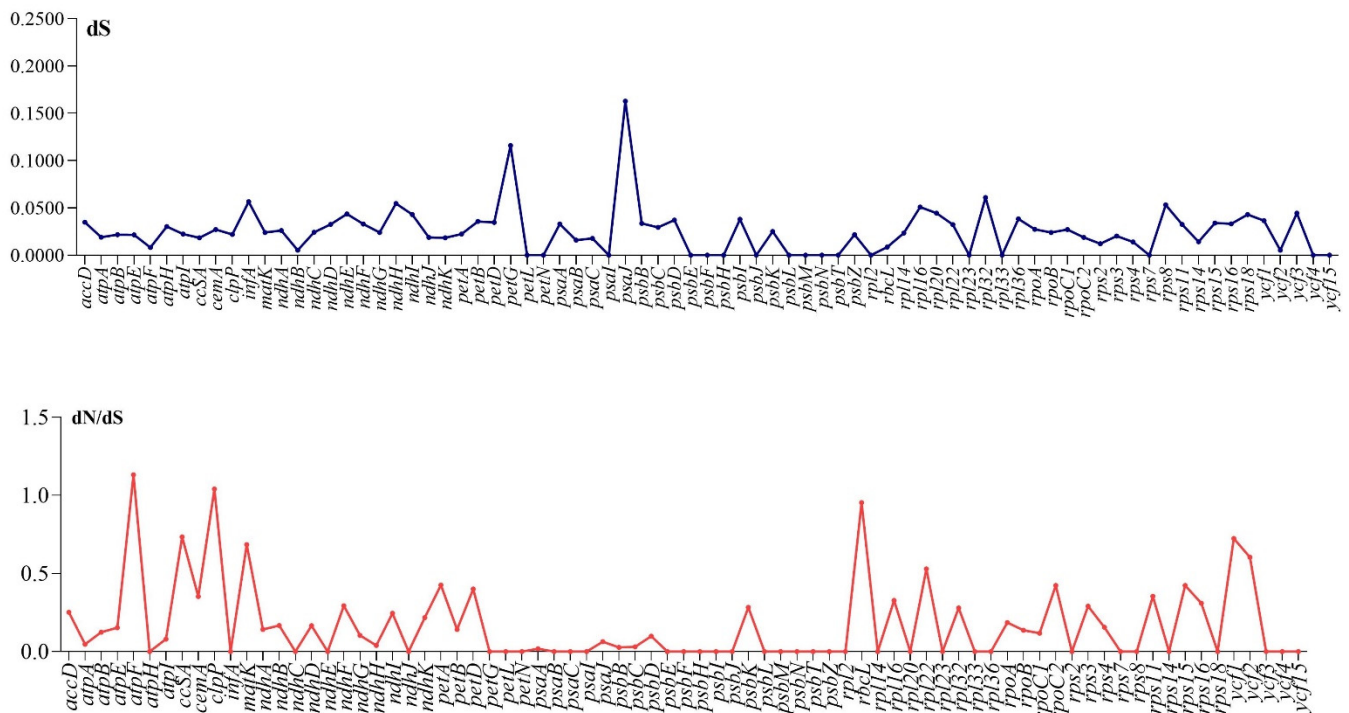


Figure 7. dN/dS ratio and synonymous values of the genes in the *H. forskalii* chloroplast genome.

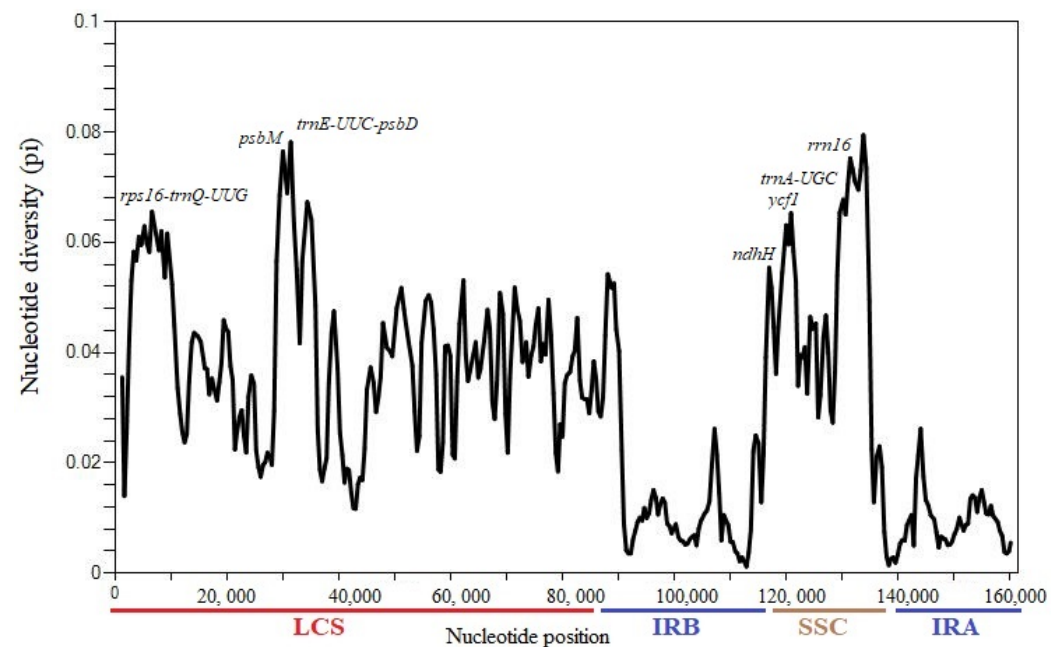
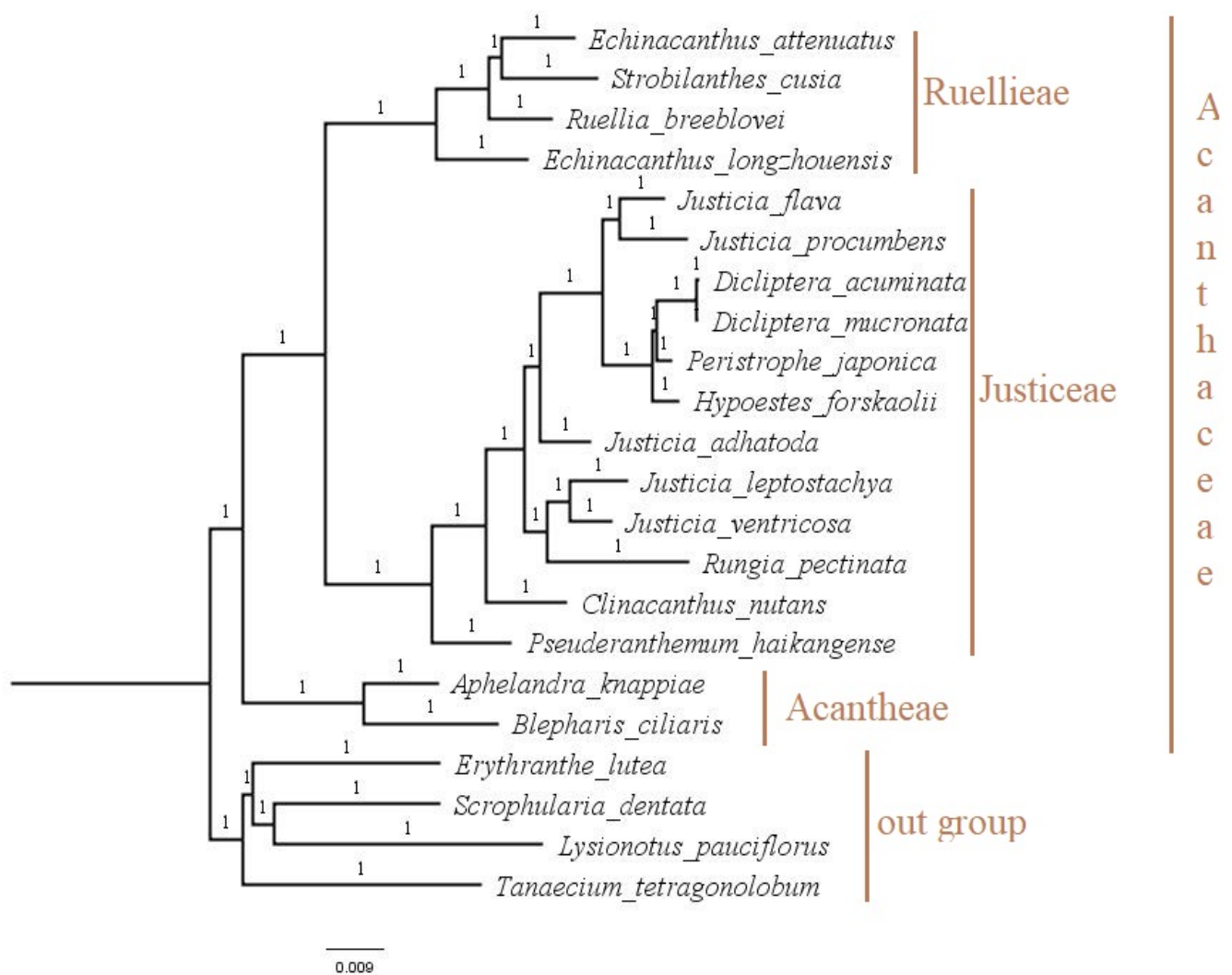


Figure 8. Sliding window analysis of nucleotide variability among the seven Justiceae species cp genomes (windowlength: 600 bp; step size: 200 bp).

### 3.6. Phylogenetic Analysis

The cp genome has been well utilized in inferring phylogenetic relationship due to its conserved nature and the presence of informative sites [63–65]. Various taxonomic complexes at different taxonomic levels have been resolved using phylogenetic trees reconstructed from cp genomes [66,67]. To reconstruct the phylogeny of Justiceae and infer the

phylogenetic position of *H. forskalii*, the cp genome sequences of 21 taxa were downloaded from Genbank. MAFFT version 7 was used to align the cp genome of *H. forskalii* and all the downloaded cp genomes. Bayesian inference was used in reconstructing the phylogenetic tree. The results in (Figure 9), show the lineage of justicioid is strongly supported (1.0 PP) with all the twelve species belonging to the tribe Justiceae clustering in one clade (monophyletic). This is congruent with previous studies using molecular, *nrITS* and some of the chloroplast genes [0,17]. Within the clade, there are six major sub clades, all with strongly supported posterior probability [PP] = 1.00. *H. forskalii* forms a sister relationship with a sub clade containing *Peristrophe* and *Dicliptera* (Diclipterinae) with a strong support. A similar tree was reported by [17] and this confirmed that the species is a member of the Diclipterinae subtribe as reported by Bremekamp in his earlier classification. The core Diclipterinae (*Dicliptera*, *Hypoestes* and *Peristrophe*) is strongly supported in this study and members of this genera are reported to be united by some of the inflorescence part [68]. The sister relationship between *Peristrophe* and *Dicliptera* needs to be revisited by looking at their phylogenetic positions, though [69,70] suggested that the two taxa should be treated as different genera due to their differences in capsule dehiscence. The New World *Justicia* species forms a sister relationship with Diclipterinae with strong support. The phylogenetic tree clearly showed that *Justicia* is paraphyletic, as reported previously by [0,17], and there is a need to combine the phylogenetic approach with a morphological study to resolve this complexity.



**Figure 9.** Bayesian inference phylogenetic tree reconstructed using the cp genome of 22 taxa. The branch nodes' numbers represent the posterior probabilities values (PP).



#### 4. Conclusions

In this study, we have sequenced and reported on the cp genome of *H. forskalii* to provide a valuable plastome genomic resources for the species. The plastomes have a typical gymnosperm cp genome structure they are comparable to another p genome of Acanthaceae. Simple sequence repeats used for evolutionary studies within the species were identified. The genome comparative analyses of seven Justiceae species revealed variable hotspot that could be used to develop a DNA barcode for the identification of the species. These hotspots will also be useful in phylogenetic relationship studies of the family Acanthaceae. The study has also revealed that only a few genes are under positive selection. The findings of the study have reported and confirmed the tribal position of major genera within Justiceae and has called for a further phylogenetic approach and morphological study to resolve the taxonomic complexities of the polyphyletic *Justicia* and the Justiceae.

**Author Contributions:** S.S.Y. collected the plant material, designed and performed the experiment, S.S.Y. and K.Y. analyzed the data and drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the University of Johannesburg postdoctoral funding scheme.

**Data Availability Statement:** The complete chloroplast genome sequence of *H. forskalii* is deposited in the GenBank with I.D no: ON398071.

**Acknowledgments:** The authors acknowledge the University of Johannesburg for the postdoctoral fellowship position.

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

#### References

1. Ellis, B.W. *Taylor's Guide to Annuals: How to Select and Grow More Than 400 Annuals, Biennials, and Tender Perennials*; Houghton Mifflin Co.: Boston, MA, USA, 1999; pp. 1–448.
2. Almehdar, H.; Abdallah, H.M.; Osman, A.-M.M.; Abdel-Sattar, E.A. In Vitro cytotoxic screening of selected Saudi medicinal plants. *J. Nat. Med.* **2012**, *66*, 406–412. [[CrossRef](#)] [[PubMed](#)]
3. Shen, C.C.; Ni, C.L.; Huang, Y.L.; Huang, R.L.; Chen, C.C. Furanolabdane diterpenes from *Hypoestes purpurea*. *J. Nat. Prod.* **2004**, *67*, 1947–1949. [[CrossRef](#)] [[PubMed](#)]
4. Mothana, R.A.; Al-Musayib, N.M.; Matheussen, A.; Cos, P.; Maes, L. Assessment of the in vitro antiprotozoal and cytotoxic potential of 20 selected medicinal plants from the island of Soqatra. *Molecules* **2012**, *17*, 14349–14360. [[CrossRef](#)] [[PubMed](#)]
5. Balkwill, K.; Norris, F.G. Taxonomic studies in the Acanthaceae; The genus *Hypoestes* in southern Africa. *S. Afr. J. Bot.* **1985**, *51*, 133–144. [[CrossRef](#)]
6. Padysakova, E.; Bartos, M.; Tropek, R.; Janecek, S. Generalization versus specialization in pollination systems: Vistors, thieves, and pollinators of *hypoestes aristate* (Acanthaceae). *PLoS ONE* **2013**, *8*, e59299. [[CrossRef](#)]
7. Lindau, G. Acanthaceae. In *Die natürlichen Pflanzenfamilien, IV(3b)*; Engler, A., Prantl, K., Eds.; Engelmann: Leipzig, Germany, 1895; pp. 274–354.
8. Van Tieghem, P. Structure du pistil et de l'ovule du fruit et de la grane des Acathacees. *Ann. Sci. Nat. Serie 9. Botanique.* **1908**, *7*, 1–24.
9. Bremekamp, C.E.B. Notes on some acanthaceous genera of controversial position. *Acta Bot. Neerl.* **1955**, *4*, 644–655. [[CrossRef](#)]
10. Bremekamp, C.E.B. Delimitation and subdivision of the Acanthaceae. *Bull. Bot. Surv. India.* **1965**, *7*, 21–30.
11. Scotland, R.W.; Vollesen, K. Classification of Acanthaceae. *Kew Bull.* **2000**, *55*, 513–589. [[CrossRef](#)]
12. McDade, L.A.; Daniel, T.F.; Masta, S.E.; Riley, K.M. Phylogenetic relationships within the tribe Justiceae (Acanthaceae): Evidence from molecular sequences, morphology, and cytology. *Ann. Missouri Bot. Gard.* **2000**, *87*, 435–458. [[CrossRef](#)]
13. Kiel, C.A.; McDade, L.A.; Daniel, T.F.; Champluvier, D. Phylogenetic delimitation of Isoglossinae (Acanthaceae: Justiceae) and relationships among constituent genera. *Taxon* **2006**, *55*, 683–694. [[CrossRef](#)]
14. Daniel, T.F.; McDade, L.A.; Manktelow, M.; Kiel, C.A. The “Tetramerium Lineage” (Acanthaceae: Acanthoideae: Justiceae): Delimitation and intra-lineage relationships based on cp and nrITS sequence data. *Syst. Bot.* **2008**, *33*, 416–436. [[CrossRef](#)]
15. Kiel, C.A.; McDade, L.A. The Mirandea clade (Acanthaceae, Justiceae, Tetramerium Lineage): Phylogenetic signal from molecular data and micromorphology makes sense of taxonomic confusion caused by remarkable diversity of floral form. *Syst. Bot.* **2014**, *39*, 950–964. [[CrossRef](#)]
16. Cortes, A.L.; Rapini, A.; Daniel, T.F. The Tetramerium lineage (Acanthaceae: Justiceae) does not support the Pleistocene Arc hypothesis for South American seasonally dry forests. *Amer. J. Bot.* **2015**, *102*, 992–1007. [[CrossRef](#)]

17. Kiel, C.A.; Daniel, F.T.; Darbyshire, I.; McDade, L.A. Unraveling relationships in the morphologically diverse and taxonomically challenging “justicioid” lineage (Acanthaceae: Justicieae). *Taxon* **2017**, *66*, 645–674. [\[CrossRef\]](#)
18. Neuhaus, H.; Emes, M. Nonphotosynthetic metabolism in plastids. *Ann. Rev. Plant Biol.* **2000**, *51*, 111–140. [\[CrossRef\]](#)
19. Grevich, J.J.; Daniell, H. Chloroplast Genetic Engineering: Recent Advances and Future Perspectives. *Crit. Rev. Plant Sci.* **2005**, *24*, 83–107. [\[CrossRef\]](#)
20. Wicke, S.; Schneeweiss, G.M.; Depamphilis, C.W.; Kai, F.M.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297.
21. Shaw, J.; Shafer, H.L.; Leonard, O.R.; Kovach, M.J.; Schorr, M.; Morris, A.B. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: Thetortoise and the hare IV. *Am. J. Bot.* **2014**, *101*, 1987–2004. [\[CrossRef\]](#)
22. Guisinger, M.M.; Chumley, T.W.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. Implications of the plastid genome sequence of typha (Typhaceae, poales) for understanding genome evolution in poaceae. *J. Mol. Evol.* **2010**, *70*, 149–166. [\[CrossRef\]](#)
23. Yang, J.B.; Tang, M.; Li, H.T.; Zhang, Z.R.; Li, D.Z. Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analysis. *BMC Evol. Biol.* **2013**, *13*. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. Geseq: Versatile and accurate annotation of organelle genomes. *Nucl. Acids Res.* **2017**, *45*, W6–W11. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2016**, *45*, e18.
26. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Mol. Biol. Evol.* **2015**, *30*, 2725–2729.
28. Thiel, T.; Michalek, W.; Varshney, R.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [\[CrossRef\]](#) [\[PubMed\]](#)
29. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Mayor, C.; Brudno, M.; Schwartz, J.R.; Poliakov, A.; Rubin, E.M.; Frazer, K.A.; Pachter, L.S.; Dubchak, I. VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **2000**, *16*, 1046–1047. [\[CrossRef\]](#)
31. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, 273–279. [\[CrossRef\]](#)
32. Librado, P.; Rozas, J. DnaSP v6: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [\[CrossRef\]](#)
33. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Fredrik, R.; Maxim, T.; Paul, V.M.; Daniel, L.A.; Aaron, D.; Sebastian, H.; Bret, L.; Liang, L.; Mar, A.S.; John, P.H. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic* **2012**, *61*, 539–542.
35. Posada, D. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **2008**, *25*, 1253–1259. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Chen, H.; Shao, J.; Zhang, H.; Jiang, M.; Huang, L.; Zhang, Z.; Yang, D.; He, M.; Ronaghi, M.; Luo, X.; et al. Sequencing and analysis of *Strobilanthes cusia* (Nees) Kuntze chloroplast Genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in Angiosperm. *Front. Plant Sci.* **2018**, *9*, 324. [\[CrossRef\]](#)
37. Chang, C.C.; Lin, H.C.; Lin, I.P.; Chow, T.Y.; Chen, H.H.; Chen, W.H.; Cheng, C.H.; Lin, C.Y.; Liu, S.M.; Chang, C.C.; et al. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol. Biol. Evol.* **2006**, *23*, 279–291. [\[CrossRef\]](#)
38. Samaila, S.Y.; Shah, M.M. The complete chloroplast genome of *Lantana camara* L. (Verbenaceae). *Mitochondrial DNA part B* **2020**, *1*, 919.
39. Park, I.; Kim, W.J.; Yeo, S.-M.; Choi, G.; Kang, Y.-M.; Piao, R.; Moon, B.C. The complete chloroplast genome sequences of *Fritillaria ussuriensis* maxim. In addition, *Fritillaria cirrhosa* D. Don, and comparative analysis with other *Fritillaria* species. *Molecules* **2017**, *282*, 22.
40. 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. [\[CrossRef\]](#)
41. Campbell, W.H.; Gowri, G. Codon usage in higher plants, green algae, and cyanobacteria. *Plant Physiol.* **1990**, *92*, 1–11. [\[CrossRef\]](#)
42. Liu, Q.; Dou, S.; Ji, Z.; Xue, Q. Synonymous codon usage and gene function are strongly related in *Oryza sativa*. *Biosystems* **2005**, *80*, 123–131. [\[CrossRef\]](#)
43. Li, B.; Lin, F.R.; Huang, P.; Guo, W.Y.; Zheng, Y.Q. Complete chloroplast Genome sequence of *Decaïsnea insignis*: Genome organization, Genomic resources and comparative analysis. *Sci. Rep.* **2017**, *7*, 10073. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Srivastava, D.; Shanker, A. Identification of simple sequence repeats in chloroplast genomes of Magnoliids through bioinformatics approach. *Interdiscip. Sci. Comput. Life Sci.* **2016**, *8*, 327–336. [\[CrossRef\]](#) [\[PubMed\]](#)

45. Zhou, J.; Chen, X.; Cui, Y.; Sun, W.; Li, Y.; Wang, Y. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species. *Int. J. Mol. Sci.* **2017**, *18*, 1839. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Jiang, D.; Zhao, Z.; Zhang, T.; Zhong, W.; Liu, C.; Yuan, Q.; Huang, L. The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. *Genes* **2017**, *8*, 227. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Zhou, J.; Cui, Y.; Chen, X.; Li, Y.; Xu, Z.; Duan, B.; Li, Y.; Song, J.; Yao, H. Complete chloroplast genomes of *Papaver rhoeas* and *Papaver orientale*: Molecular structures, comparative analysis, and phylogenetic analysis. *Molecules* **2018**, *23*, 437. [\[CrossRef\]](#)
48. Li, Y.G.; Xu, W.Q.; Zou, W.T.; Jiang, D.Y.; Liu, X.H. Complete chloroplast genome sequences of two endangered *Phoebe* (Lauraceae) species. *Bot. Stud.* **2017**, *58*, 37. [\[CrossRef\]](#)
49. Greiner, S.; Wang, X.; Rauwolf, U.; Silber, M.V.; Mayer, K.; Meurer, J.; Haberer, G.; Herrmann, R.G. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. sequence evaluation and plastome evolution. *Nucleic Acids Res.* **2008**, *36*, 2366–2378. [\[CrossRef\]](#)
50. Song, Y.; Wang, S.; Ding, Y.; Xu, J.; Li, M.F.; Zhu, S.; Chen, N. Chloroplast Genomic Resource of Paris for Species Discrimination. *Sci. Rep.* **2017**, *7*, 3427. [\[CrossRef\]](#)
51. Bryan, G.J.; McNicol, J.W.; Meyer, R.C.; Ramsay, G.; De Jong, W.S. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. *Theor. Appl. Genet.* **1999**, *99*, 859–867. [\[CrossRef\]](#)
52. Provan, J. Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana*. *Mol. Ecol.* **2000**, *9*, 2183–2185. [\[CrossRef\]](#)
53. Ebert, D.; Peakall, R. Chloroplast simple sequence repeats (cpSSRs): Technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol. Ecol. Resour.* **2009**, *9*, 673–690. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Jacinta, N.M.; Xiang, D.; Jia-Xin, Y.; Elijah, M.M.; Vincent, O.W.; Millicent, A.O.; Josphat, K.S.; Paulm, M.M.; Guang-Wan, H. Complete chloroplast genome of *Chlorophytum comosum* and *Chlorophytum gallabatense*: Genome structures, comparative and phylogenetic analysis. *Plants* **2020**.
55. Dhafer, A.A.; Samaila, S.Y.; Enaj, J.A.; Abidina, A. Complete chloroplast genome sequence of *Barleria prionitis*, Comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genom.* **2020**, *21*, 393.
56. Philippe, H.; Delsuc, F.; Brinkmann, H.; Lartillot, N. Phylogenomics, Annual Review of Ecology. *Evol. Sys.* **2005**, *36*, 541–562. [\[CrossRef\]](#)
57. Raubeson, L.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boorem, J.L.; Jansen, R.K. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genom.* **2007**, *8*, 174–201. [\[CrossRef\]](#)
58. Zhou, T.; Chen, C.; Wei, Y.; Chang, Y.; Bai, G.; Li, Z.; Kanwal, Z.; Zhao, G. Comparative transcriptome and chloroplast genome analyses of two related *Dipteronia* Species. *Front. Plant Sci.* **2016**, *7*, 1512. [\[CrossRef\]](#)
59. Rousseau-Gueutin, M.; Bellot, S.; Martin, G.E.; Boutte, J.; Chelaifa, H.; Lima, O.; Michon-Coudouel, S.; Naquin, D.; Salmon, A.; Ainouche, K.; et al. The chloroplast genome of the hexaploid *Spartina maritima* (Poaceae, Chloridoideae): Comparative analyses and molecular dating. *Mol. Phylogenet. Evol.* **2015**, *93*, 516. [\[CrossRef\]](#)
60. Xu, J.-H.; Liu, Q.; Hu, W.; Want, T.; Xue, Q.; Messing, J. Dynamics of chloroplast genome in green plants. *Genomics* **2015**, *106*, 221–231. [\[CrossRef\]](#)
61. Hebert, P.D.N.; Ratnasingham, S.; de Waard, J.R. Barcoding animal life: Cytochrome C oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B Biol. Sci.* **2003**, *270*, S96–S99. [\[CrossRef\]](#)
62. Hebert, P.D.N.; Stoeckle, M.Y.; Zemlak, T.S.; Francis, C.M.; Charles, G. Identification of birds through DNA barcodes. *PLoS Biol.* **2004**, *2*, e312. [\[CrossRef\]](#)
63. Borsch, T.; Quandt, D. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Syst. Evol.* **2009**, *282*, 169–199.
64. Dong, W.P.; Liu, J.; Yu, J.; Wang, L.; Zhou, S.L. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **2012**, *7*, e35071. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Tong, W.; Kim, T.S.; Park, Y.J. Rice chloroplast genome variation architecture and phylogenetic dissection in diverse *Oryza* species assessed by whole-genome resequencing. *Rice* **2016**, *9*, 57. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Dong, W.P.; Liu, H.; Xu, C.; Zuo, Y.J.; Chen, Z.J.; Zhou, S.L. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. *BMC Genet.* **2014**, *15*, 138. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Du, Y.P.; Bi, Y.; Yang, F.P.; Zhang, M.F.; Chen, X.Q.; Xue, J.; Zhang, X.H. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **2017**, *7*, 5751. [\[CrossRef\]](#)
68. Darbyshire, I.; Vollesen, K. The transfer of the genus *Peristrophe* to *Dicliptera* (Acanthaceae), with a new species described from eastern Africa. *Kew Bull.* **2007**, *62*, 119–128.
69. Balkwill, K.F. Taxonomic studies in the tribe Justicieae of the family Acanthaceae. [Unpublished]. Ph.D. Thesis, University of Natal, Pietermaritzburg, South Africa, 1985.
70. Balkwill, K.; Gettiffe-Norris, F.; Balkwill, M.J. Taxonomic studies in the Acanthaceae: *Dicliptera* in southern Africa. *Kew Bull.* **1996**, *51*, 1–61. [\[CrossRef\]](#)