

## Article

# Comparative Plastomes of *Curcuma alismatifolia* (Zingiberaceae) Reveal Diversified Patterns among 56 Different Cut-Flower Cultivars

Jie Wang<sup>1,2,†</sup>, Xuezhu Liao<sup>2,†</sup>, Yongyao Li<sup>2</sup>, Yuanjun Ye<sup>3</sup>, Guoming Xing<sup>1</sup>, Shenglong Kan<sup>2</sup> , Liyun Nie<sup>2</sup>, Sen Li<sup>1</sup>, Luke R. Tembrock<sup>4,\*</sup>  and Zhiqiang Wu<sup>2,\*</sup> 

- <sup>1</sup> College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China; wj18434764189@163.com (J.W.); xingguoming@163.com (G.X.); saulisen@163.com (S.L.)
- <sup>2</sup> Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China; liaoxuezhu@caas.cn (X.L.); liyongyao@caas.cn (Y.L.); kanshenglong@caas.cn (S.K.); nieliyun18@163.com (L.N.)
- <sup>3</sup> Guangdong Provincial Key Lab of Ornamental Plant Germplasm Innovation and Utilization, Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China; yeyuanjun199910@126.com
- <sup>4</sup> Department of Agricultural Biology, Colorado State University, Fort Collins, CO 80523, USA
- \* Correspondence: luke.tembrock@colostate.edu (L.R.T.); wuzhiqiang@caas.cn (Z.W.)
- † These authors contributed equally to this work.

**Abstract:** *Curcuma alismatifolia* (Zingiberaceae) is an ornamental species with high economic value due to its recent rise in popularity among floriculturists. Cultivars within this species have mixed genetic backgrounds from multiple hybridization events and can be difficult to distinguish via morphological and histological methods alone. Given the need to improve identification resources, we carried out the first systematic study using plastomic data wherein genomic evolution and phylogenetic relationships from 56 accessions of *C. alismatifolia* were analyzed. The newly assembled plastomes were highly conserved and ranged from 162,139 bp to 164,111 bp, including 79 genes that code for proteins, 30 tRNA genes, and 4 rRNA genes. The A/T motif was the most common of SSRs in the assembled genomes. The *Ka/Ks* values of most genes were less than 1, and only two genes had *Ka/Ks* values above 1, which were *rps15* (1.15), and *ndhl* (1.13) with *petA* equal to 1. The sequence divergence between different varieties of *C. alismatifolia* was large, and the percentage of variation in coding regions was lower than that in the non-coding regions. Such data will improve cultivar identification, marker assisted breeding, and preservation of germplasm resources.

**Keywords:** plastomes; comparative analyses; phylogeny; ornamental horticulture; *Curcuma alismatifolia*; siam tulip



**Citation:** Wang, J.; Liao, X.; Li, Y.; Ye, Y.; Xing, G.; Kan, S.; Nie, L.; Li, S.; Tembrock, L.R.; Wu, Z. Comparative Plastomes of *Curcuma alismatifolia* (Zingiberaceae) Reveal Diversified Patterns among 56 Different Cut-Flower Cultivars. *Genes* **2023**, *14*, 1743. <https://doi.org/10.3390/genes14091743>

Academic Editor: Jacqueline Batley

Received: 27 July 2023

Revised: 20 August 2023

Accepted: 28 August 2023

Published: 31 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*C. alismatifolia* is a species in the Zingiberaceae family and native to Southeast Asia [1]. Also known as the “summer tulip”, this species was introduced into China in the 1990s [2]. *C. alismatifolia* has been well known in Asia through its long history as a medicinal plant [3,4]. The showy part of most Zingiberaceae species are bracts, and because of the numerous and aesthetically attractive bracts generated by various cultivars, species in *Curcuma* have recently gained popularity among floriculturists [5]. The ornamental cultivars of *Curcuma* are often of hybrid origin, with polyploidy common among many wild and cultivated lineages [6,7]. The traditional diagnostic characteristics used to identify *Curcuma* are insufficient and difficult to utilize because floral morphology, rhizome color, bract shape and color, and position of inflorescences in *Curcuma* are neither universal or unique to all species [8]. In recent years, increased interest in *Curcuma* has motivated researchers to

employ molecular techniques to better resolve relationships between different cultivars and species [9]. The use of molecular techniques is further justified by the fact that morphological characteristics used to identify *Curcuma* lineages often do not resolve closely related lineages. Additionally, the lack of high-quality nuclear reference genomes and the presence of hybrid lineages in *Curcuma* make the use of plastomic techniques especially applicable for the resolution of relationships in the immediate term. The species *C. alismatifolia*, which is one of the largest and most showy in the genus *Curcuma*, contains many named cultivars with diverse hybrid origins. Such complex pedigrees can make the tracing of phylogenetic relationships difficult and potentially hinder breeding efforts. While some plastome data have been generated for *Curcuma* [6,10], no previous work has used this type of data to study the relationships between *C. alismatifolia* accessions. Such efforts will enrich our understanding of the history and diversity of *C. alismatifolia* cultivation as well as improve future breeding efforts through avoidance of incompatible crosses.

Through the swift advancement of sequencing technology and related cost reductions, the plastid genome or plastome, which is simpler and cheaper to assemble than the nuclear genome, is now regularly employed in phylogenomic and population genomic studies [11–13]. In nearly every case, plastids in land plants, which are typically between 100 and 200 kb in length, have a typically circular structure, including two inverted repeats (IRa/b), a large single-copy (LSC) region, and a small single-copy (SSC) region between them [14–16]. Plastomes are inherited through the maternal line and lack recombination in nearly all angiosperm lineages [17,18]. The small genome size, lack of recombination, and conserved genomic structure of the plastome are the main factors contributing to the simplicity and accuracy of plastome genome assembly and analysis [19,20]. For precise species identification and phylogenetic inference, plastome genome sequences have been extensively exploited [21,22]. Comparative plastome genomics can also uncover mutational hotspots, which provide robust signals in species differentiation, phylogenetic analysis, and population genetic studies, as well as crucial information on the study of plastome evolution, such as patterns of gene loss and IR border variability [23,24]. While mutational hotspots are widely used in barcoding projects, entire plastomes are now being employed as super DNA barcodes given the relative ease in generating such data [25–27]. Comprehensive characterizations of numerous individual plastomes within a species form the basis of functional genomic investigations and can guide efforts in chloroplast transformation in efforts to improve metabolic function [28]. Additionally, with further genomic classification and comparison, the plastome can be used as a substantial repository of molecular markers for mapping, phylogenetic study, population-level research, and genetic analysis [29–31]. Lastly, the resolution of plastome lineages is necessary for plant breeders in characterizing plastome–nuclear–genome incompatibilities such that inviable crosses can be avoided in hybrid cultivar development [32,33].

Because of the reticulate history and widespread morphological diversity found in *C. alismatifolia*, this study aimed to (1) contribute full-length plastomes from previously uncharacterized lineages of *C. alismatifolia* cultivars, (2) perform comparative analyses from these plastomes to more effectively comprehend the evolution of genome architecture at the intraspecific level, (3) reconstruct the maternal phylogenetic relationships in *C. alismatifolia* using plastome evidence and compare this to previous taxonomic delimitations including named cultivars, and (4) propose novel DNA markers to discriminate *C. alismatifolia* maternal lineages. We expected to find plastomic introgression from species outside of *C. alismatifolia* if interspecific crosses were involved in the history of the cultivated lineages. Furthermore, we expected to find high levels of genomic diversity within *C. alismatifolia* given the high levels of morphological diversity between different cultivars.

## 2. Materials and Methods

### 2.1. Sampling and DNA Extraction, Sequencing

Fresh plant leaves from about 56 accessions of *C. alismatifolia* were collected for DNA extraction (Table S1). We placed the collected fresh plant leaves of the samples in  $-80^{\circ}\text{C}$ . The

material for these extractions was collected from plants growing at Guangdong Academy of Agricultural Sciences. An established cetyl trimethyl ammonium bromide (CTAB) method was utilized to extract total genomic DNA before Illumina paired-end sequencing [34].

## 2.2. Plastome Sequencing, Assembly, and Annotation

We first extracted DNA from libraries with an insert length of 500 bp and sequenced the paired-end read length of 150 bp using MGISEQ-T7. We used Fastp v0.20.1 [35] for quality control filtering of the raw data, applying the following criteria, which produced around 6 Gb of clean reads for each accession: filtered reads with adapters, reads with N bases greater than 10%, and reads with low-quality bases greater than 50%. The chloroplast database was used to match all paired-end clean reads via the bwa v0.7.17-r1188 [36] software. Plastome-specific reads were then chosen using Picard v2.20.3 [37]. The selections of plastome reads were assembled using SPAdes v3.14.0 [38] with default parameter settings, and the resultant scaffolds (GFA files) were imported using Bandage v0.8.1 [38] to generate the complete plastome for each accession. The *C. phaeocaulis* plastome was used as a template to give a complete reference. All genes were manually delimited. The Plastid Genome Annotator (PGA) [39] was then used to perform a complete re-annotation of every species, followed by manual revisions where necessary. The Draw Organelle Genome Maps online software (OGDRAW version 1.3.1) [40] was used to implement the depiction of plastome structure.

## 2.3. Comparative Genomic Analysis

REPuter typical settings were used to identify four distinct repetition types, including forward, reverse, palindrome, and completion [22,41]. The Perl script MISA was used to find simple sequences repetition (SSRs), with 10, 6, 5, 5, 5, and 5 repeat units set for mono-, di-, tri-, tetra-, penta-, and hexa-motif microsatellites as the minimum threshold, respectively [42]. Using the Ca1 plastome as a reference, genome sequence diversity analysis was conducted using mVISTA to compare 11 representative plastomes using the Shuffle-LAGAN [43] alignment tool. To determine the nonsynonymous ( $K_a$ ), synonymous ( $K_s$ ), and the ratio of nonsynonymous to synonymous nucleotide substitutions ( $K_a/K_s$ ) for each gene, we utilized CODEML in PAML v4.9 [44].

## 2.4. Phylogenetic Analysis

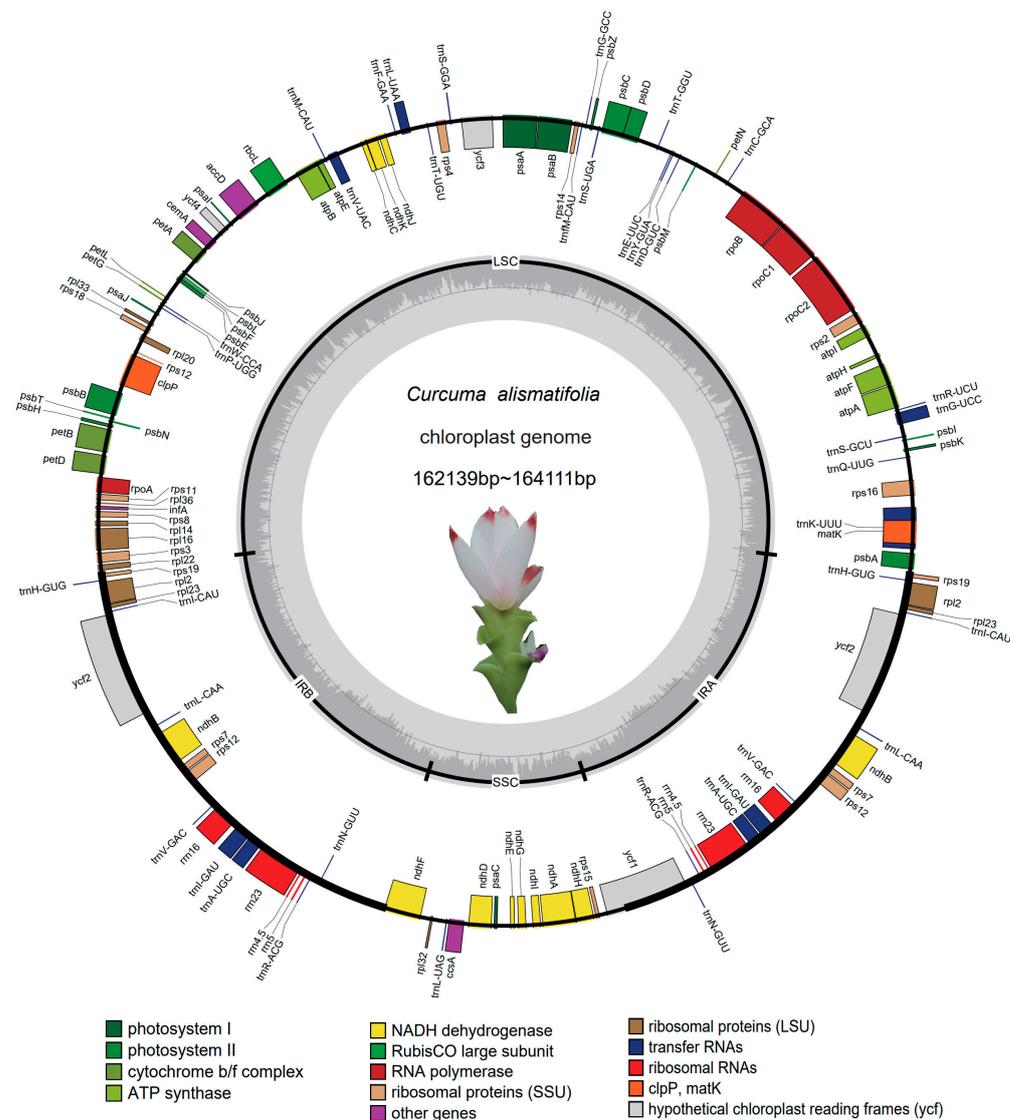
All 56 plastomes and two outgroup species (*Curcuma longa*, NC\_042886.1 and *Curcuma phaeocaulis*, NC\_045242.1) were included for analysis. Using MAFFT v7.464 [45,46], a whole plastome sequence alignment of 56 plastomes and two outgroup species was produced. The misaligned locations were trimmed using TrimAL v1.3 [47]. According to the annotation files, we extracted and retrieved the longest CDS sequences of 79 genes that code for proteins from the genome sequence of each plastome and compared them using MAFFT [45,46]. To achieve our objectives, we concatenated the alignments of the nucleotide sequence with 79 genes that code for proteins. With the use of 1000 ultrafast bootstrap runs and IQ-TREE v2.0 [44], this data set was also used to reconstruct the phylogenetic tree and evaluate branching support in the FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree>, accessed on 12 March 2023) tree display program.

# 3. Results

## 3.1. General Features of *C. alismatifolia* Plastomes

Complete plastome lengths for 56 *C. alismatifolia*-cultivated accessions ranged from 162,139 bp to 164,111 bp (Figure 1), Similar to other studies, all the plastomes possessed a typical quadripartite structure including a large single copy (LSC) region (ranging from 86,860–88,364 bp), a small single copy (SSC) region (ranging from 15,692–15,863 bp), and two inverted repeat (IR) regions (ranging from 29,855–30,255 bp). In general, the plastome sequences of the *C. alismatifolia* accessions were similar in length and structure. The average

overall GC content among the sequenced plastomes was 36.0% with a 40.9% GC content within IRs, 33.8% within the LSC, and 29.6% within the SSC (Table S1).



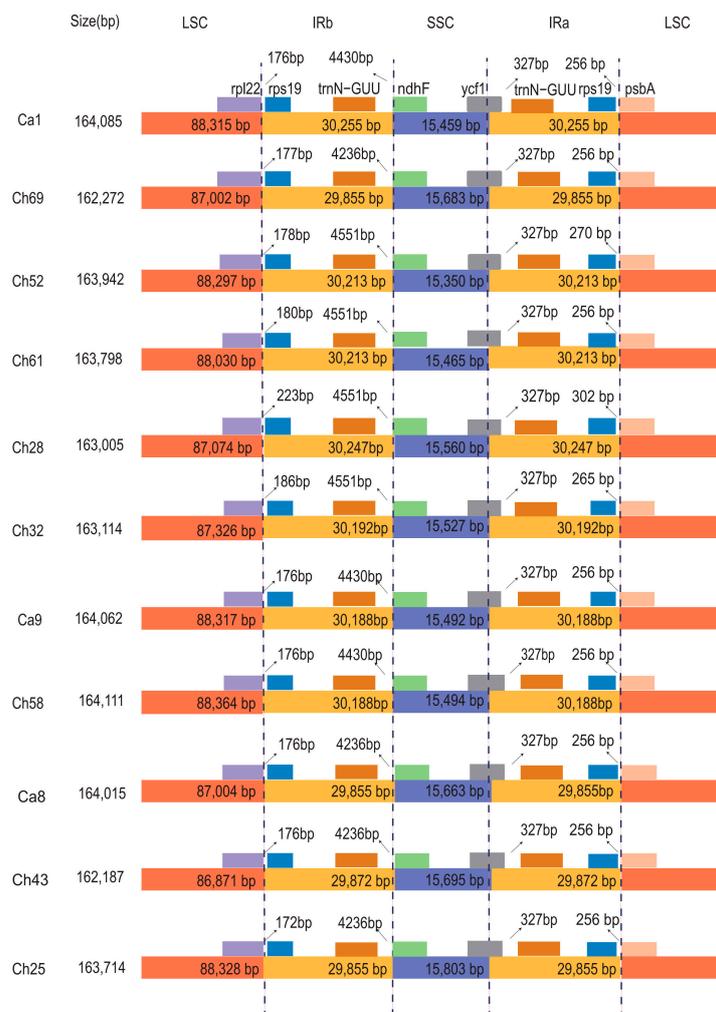
**Figure 1.** The general plastome structure from 56 *C. alismatifolia* accessions. The inverted repeat regions (IRA and IRB) are identified with thick lines on the outer entire circle. The direction of transcription for the genes outside the circle is clockwise, whereas the genes inside are expressed in a counterclockwise direction. Colors are assigned to genes depending on their functional groupings. Darker/lighter grey bars in the inner ring represent GC/AT content.

The 56 *C. alismatifolia* accessions were not only similar in length and structure, but also contained the same number of genes. Each accession contained 113 functional genes, with 79 genes that code for proteins, 30 tRNA genes, and four rRNA genes in the plastomes (Tables S2 and S3). Among these 113 unique genes, one gene was found that crossed different repeat junctions. The gene *ycf1* extended across the junction of IRA and SSC. Of the 113 genes, the LSC region contained a total of 82 genes including 21 tRNAs and 61 genes that code for proteins. Both inverted repeats contained 20 genes including eight genes that code for proteins, eight tRNAs, and four rRNAs. The SSC region contained 13 genes with 11 protein coding genes, one tRNA, and one rRNA. A total of two introns were found in three genes, including *rps12*, *clpP*, and *ycf3*, and one intron in each of the remaining 14 genes, including *trnL-UAA*, *trnA-GUC*, *trnG-UCC*, *trnV-UAC*, *trnI-GAU*, *trnK-UUU*, *rpl2*, *rpoC1*, *atpF*, *rpl16*, *petB*, *ndhA*, *petD*, and *ndhB*. The longest intron of *trnK-UUU* contains the *matK*

gene. Trans-splicing of the *rps12* gene was predicted to occur with a single 5' end exon in the LSC and a repetitive 3' end duplication in two IRs (Table S3).

### 3.2. Contraction and Expansion of IRs

The total length of the plastome can constrict and expand as a result of differences in the single-copy and IR region sizes at the borders. A total of 11 representative *C. alismatifolia* accessions were compared at the LSC/IR/SSC borders. Despite the four units being very well conserved, the boundary regions of the LSC, IR, and SSC still showed minor differences. For example, the intergenic region linking the IRB and SSC junction displayed a significant range in length among the 11 plastomes tested. Among *trnN-GUU* and *ndhF*, there were approximately 4236 to 4551 bp between them. *Ycf1* spanned the SSC and IRA junction in all 11 plastomes, but there was no gene spanning the IRB and SSC junction. The intergenic regions between the IR and LSC junctions varied significantly among the 11 plastomes with the distance between *rpl22* and *rps19* ranging from 172 bp to 233 bp between the IRB and LSC regions and from 256 bp to 302 bp between *rps19* and *psbA* at the IRA and LSC junction. The main contribution to the difference in IR and LSC boundaries was that the *rps19* gene in accession Ch28 was more distant from *rpl22* than in any other plastome, although the total length of this plastome was shorter than all but 2 of the 11 plastomes compared here (Figure 2).



**Figure 2.** Analyzing 11 *C. alismatifolia* accessions at junctions throughout the LSC, SSC, and IR regions. Figure is not scaled (LSC: large single copy, IRa/b: inverted repeat, SSC: small single copy). The distance between nearby genes is shown via the number displayed next to the straight line at the junction.

### 3.3. Sequence Repeats in the Complete Plastome

Nucleotide repeats in plastomes such as SSRs can be particularly helpful markers to recognize species and populations considering the high levels of mutation in these regions. A comparative analysis of repetitive sequences between all 56 plastomes found that the general distribution, amounts of repeats, and types share a lot of similarities among the *C. alismatifolia* and relative species. A total of four SRR types were identified in 56 *C. alismatifolia* plastomes; among these SSRs, 82.7% (3858/4666) were single nucleotide A/T motifs, 15.1% (702/4666) were AT/AT motifs, 0.8% (40/4666) were AAT/ATT motifs, and 1.4% (66/4666) were C/G motifs. Between 55 and 75 A/T SSRs were discovered in the 56 plastomes. Ch63 and Ch47 were found to have fewer than other accessions with only 55 found, indicating that for these accessions, long A/T segments tend to have more indels. The 56 plastomes contained between 10 and 17 AT/AT SSRs, with Ch43 having the highest number. The 56 plastomes were discovered to contain a total of 0–2 C/G SSRs, with Ch52 having none. There were between zero and two AAT/ATT SSRs found in the 56 plastomes. There were 16 accessions found to contain no AAT/ATT SSRs (Figure 3a). The SSR analysis revealed that certain genomic regions can be used to determine the lineages of *C. alismatifolia* if there are data available for various types of SSR motifs and their length variations are combined into a research hierarchy. Through the analysis of SSRs, it has been determined that specific genomic regions can effectively identify the lineages of *C. alismatifolia*. This is achieved by combining data on various types of SSR motifs and their corresponding length variations. In the future, these SSRs might be utilized to provide molecular markers for genetics in population studies and species identification.

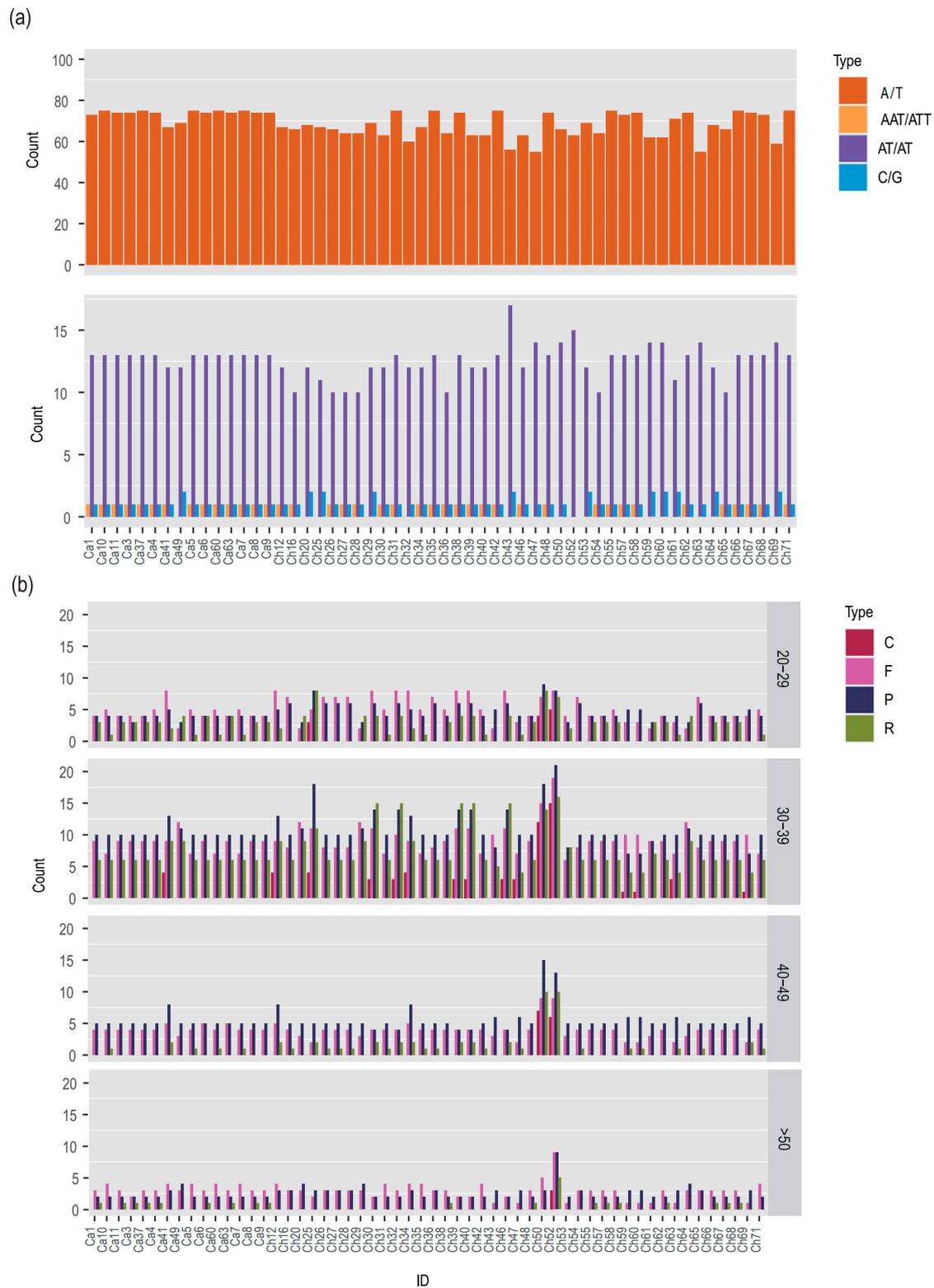
All 56 plastomes were analyzed using Reputer to find repeats of a length of at least 8 bp. Four distinct repeats were taken into consideration: forward (F), complement (C), reverse (R), and palindromic (P). The total number for each motif type was determined by grouping them based on sequence length. Most of the repetitive sequences ranged in length from 30 to 39 bp, followed by the 20–29 bp and 40–49 bp ranges, with the >50 bp range having the fewest. In the 20–29 bp group, only three accessions contained all four nucleotide repeats. A total of 43 accessions had three nucleotide repeats not including C repeats, and 10 accessions contained only F and P. In the 30–39 bp group, only 16 accessions contained all four nucleotide repeats. The remaining 40 accessions had three nucleotide repeats not including C. In the 40–49 bp group, nearly half of the 56 accessions contained only F and P. Ch50 and Ch52 contained all four nucleotide repeats, and C repeats were only found in Ch50 and Ch52. In the >50 bp group, 15 accessions contained three nucleotide repeats, and only Ch52 contained four nucleotide repeats (Figure 3b). Based on the variations in the type and quantity of repeats, a creation of markers which might be employed to identify different species or lineages is possible. In previous research, others have found that variations in the location, abundance, and type of repeated sequences found in a genome can serve as reliable indicators to recognize species or populations.

### 3.4. Evolutionary Rates among Protein Coding Genes

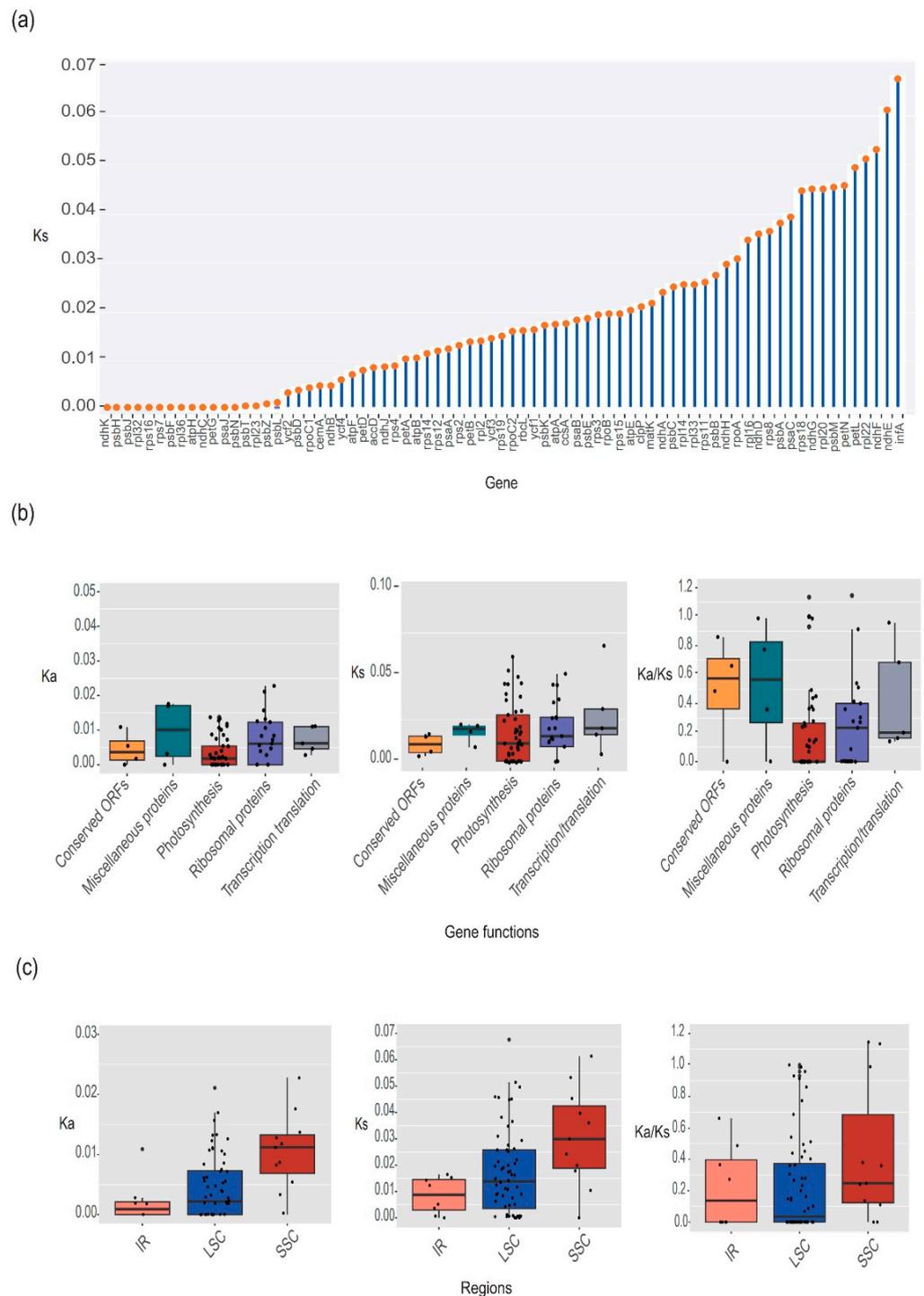
To further verify diversity and evolution among functional sequences in *C. alismatifolia* accessions, we estimated the  $K_s$  for each of 79 genes that code for proteins to compare the rates of evolution between different *C. alismatifolia* plastomes. The most rapidly evolving genes as quantified by  $K_s$  were *infA* (0.0676), *ndhE* (0.0613), *ndhF* (0.0532), *rpl22* (0.0514), and *petL* (0.0496). By contrast, nearly 20 genes, such as *ndhK*, *psbH*, *rpl32*, *ndhC*, *ycf2*, and *psbF*, the genes which evolved slowly, had  $K_s$  values close to zero (Figure 4a).

To calculate the evolutionary velocities of various genomic units, these genes were divided into functional or locational (including LSC, SSC, and IR) groupings. When classifying according to their function, transcription genes showed the highest  $K_s$ ; photosynthesis genes showed the lowest  $K_a$  (Figure 4b). When genes were analyzed based on genomic location, the SSC had the highest rates and the IRs had the lowest (Figure 4c). Most genes had  $K_a/K_s$  values less than 1, indicating that they were susceptible to selective purification

(Figure 4b,c). Only two genes had  $K_a/K_s$  values above 1, which were *rps15* (1.15) and *ndhl* (1.13) with *petA* equal to 1 (Table S4).



**Figure 3.** (a) The number of distinct types of simple sequences repetition (SSRs) from 56 *C. alismatifolia* plastomes, including A/T, AT/AT, AAT/ATT, and C/G type SSRs. (b) Variances in repeat quantity and type in 56 *C. alismatifolia* plastomes including forward (F), complement (C), reverse (R), and palindromic (P) type repeats.

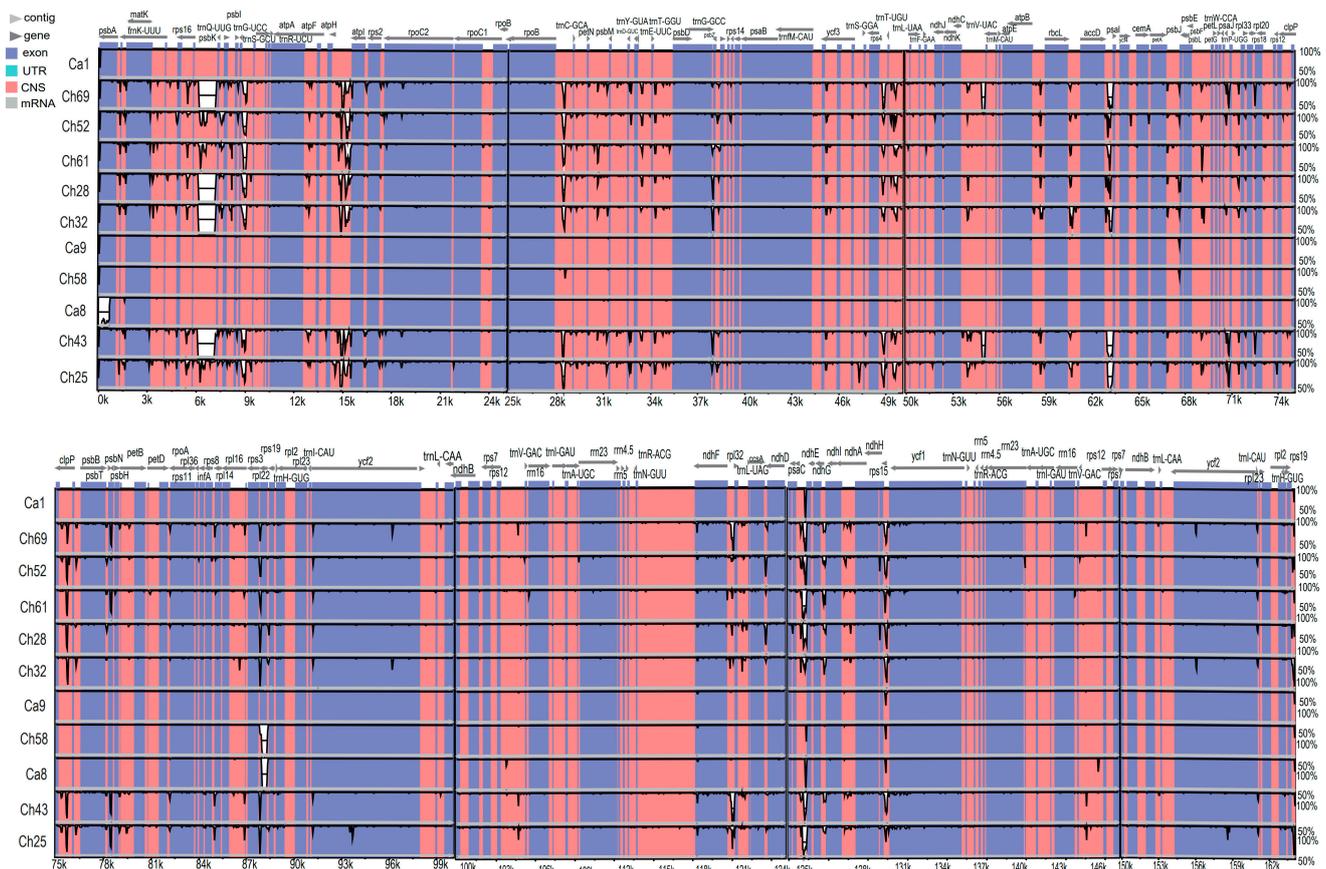


**Figure 4.** This study examined the selection patterns and intensity of 79 genes that code for proteins in 56 *C. alismatifolia* plastomes. (a)  $K_s$  values of 79 genes that code for proteins ranked by  $K_s$ . (b)  $K_a$ ,  $K_s$ , and  $K_a/K_s$  ratios in genes with different functional classifications, and (c)  $K_a$ ,  $K_s$ , and  $K_a/K_s$  ratios in LSC, SSC, and IR regions.

### 3.5. Genome Sequence Divergence

A comparison between 11 representative *C. alismatifolia* accessions was conducted to identify mutational hotspots in the plastome. We found that the sequence divergence between different accessions of *C. alismatifolia* was largest in noncoding regions of the plastome. The LSC exhibited a higher level of sequence divergence, and IRs exhibited the lowest. The 11 accessions could be divided into two main types according to their

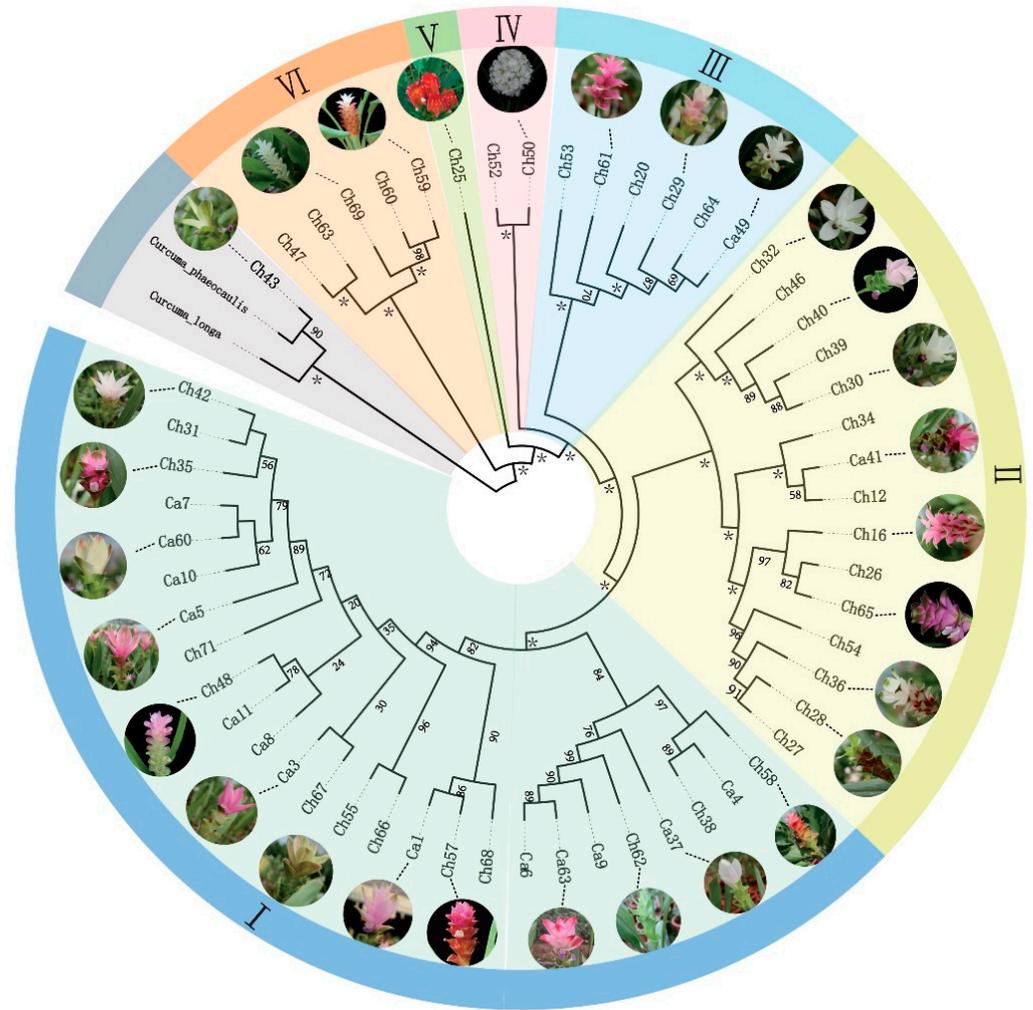
shared plastome variations (Figure 5). However, some hypervariable regions could be detected within each type. For example, the sequences of Ch69, Ch52, Ch61, Ch28, Ch32, and Ch43 are similar in genome sequence divergence, but in the coding regions, fifteen genes possessed greater variability: *rps16*, *psbK*, *atpI*, *rpoC2*, *rpoC1*, *rpl20*, *clpP*, *psbT*, *psbN*, *rpl14*, *ycf2*, *ndhF*, *ndhG*, *ycf2*, and *rps19*. Seventeen intergenic regions also showed higher levels of variations: *trnK-UUU-rps16*, *rps16-trnQ-UUG*, *rps4-trnT-UGU*, *atpI-rps2*, *petN-psbM*, *psbM-trnD-GUC*, *trnY-GUA-trnE-UUC*, *rps4-trnT-UGU*, *ndhC-trnV-UAC*, *rbcL-accD*, *accD-psaL*, *psbE-trnW-CCA*, *psbT-psbN*, *rps12-trnV-GAC*, *ndhF-rpl32*, *ccsA-ndhD*, and *rps15-ycf1*.



**Figure 5.** Comparing the sequence divergence of 11 newly assembled *C. alismatifolia* plastomes using mVISTA and Ca1 as reference. The range of sequence identity (50–100%) is shown on the Y-axis. The genes for tRNA and rRNA are not shown in this picture.

### 3.6. Phylogenetic Analyses and Molecular Marker Identification

We completed a phylogenetic analysis utilizing 56 *C. alismatifolia* plastomes to evaluate the divergence of the plastome in the context of evolution and uncover synapomorphies (and eventually barcodes) for certain lineages, with *C. phaeocaulis* (NC\_045242.1) and *C. longa* (NC\_042886.1) set as outgroups. A phylogenetic tree of *C. alismatifolia* based on all CDSs concatenated the resolved Ch43 with the outgroup *C. phaeocaulis*, suggesting that this accession was either misidentified or of hybrid origin. A phylogenetic tree based on concatenated CDSs suggests that the Ch43 accession of *C. alismatifolia* was either misidentified or of hybrid origin, as it resolved with the outgroup *C. phaeocaulis* (Figure S1). The remaining 55 accessions could be further divided into seven clades according to the phylogeny. Ch25 was resolved in a distinct early diverging clade with more support. The color and phenotype of different accessions within clades is diverse, suggesting that morphology and maternal lineage are apparently discordant, but closer examination using the lineages defined here may provide previously undetected morphological similarities (Figure 6).



**Figure 6.** Phylogenetic relationships of 56 *C. alismatifolia* plastomes according to complete plastome sequences. The rapid bootstrap levels produced using IQ-TREE are represented by scores at the nodes; \* at nodes represents a bootstrap value of 100%.

To further discriminate the 56 *C. alismatifolia* accessions into discrete groupings, we searched for regions that were abundant in SNPs and INDELS to locate possible barcode loci. From the alignment, 25 SNP loci and 6 INDEL loci were discovered in group VI, 145 SNP loci and 69 INDEL loci were discovered in group V, 111 SNP loci and 50 INDEL loci were discovered in group IV, 70 SNP loci and 49 INDEL loci were discovered in group III, 54 SNP loci and 33 INDEL loci were discovered in group II, and 2 SNP loci were discovered in group I (Tables S5 and S6). A super DNA barcode advent might be taken into consideration utilizing the complete plastome to recognize *C. alismatifolia* lineages using shotgun sequencing data given the prevalence of SNPs and INDELS throughout the whole genome (Table 1).

**Table 1.** Number of potential molecular markers for different branches of the evolutionary tree.

Group	INDELS	SNPs
I	55	100
II	33	54
III	49	70
IV	50	111
V	69	145
VI	6	25

#### 4. Discussion

In this study, we reported 56 complete plastome sequences from well-known cultivars of *C. alismatifolia*. By assembling genomes and annotating genes, we obtained more particularized information on plastome evolution in *C. alismatifolia* and presented a comparative analysis. From these analyses, it is clear that the overall structure of *C. alismatifolia* plastomes is conserved but that sufficient molecular evolution has occurred such that different groups can be identified.

Contraction and expansion of plastome genomic units and repositioning of IR junctions can result in important evolutionary processes, similar modifications in plastome size, gene duplication, the creation of pseudogenes, or the reduction of many copies of a gene to a single copy [48,49]. Such large boundary differences can be distinct markers for lineage identification [18,24]. We selected 11 representative accessions from different branches of the phylogenetic tree to study IR junctions and sequence evolution. Considerable repositioning in the IR and LSC boundary was found in Ch28 compared to the other accessions (Figure 2). We discovered that there were variations among accessions in IR junction positioning with similar sequences, reflecting the substantial differences within *C. alismatifolia*. Given these length differences at the IR boundaries, molecular markers could be developed at these regions, and through an SNP analysis, we also found some sites that could be used as DNA barcodes [50].

Synonymous mutations are generally considered to not be subjected to natural selection, and subsequently,  $K_s$  represents the foundational base replacement speed of the evolutionary procedure [51–54]. We determined the  $K_s$  for each of the 79 genes that code for proteins in the 56 *C. alismatifolia* accessions to comprehend the evolutionary history of the species (Figure 4). Among them, the *infA* gene had the fastest  $K_s$  (0.0672). While the majority of the slower-evolving genes were involved in self-replication and photosynthesis [24]. In most cases, the rates of evolution in some plastome genes are species-specific. For example, the *clpP* gene is highly conserved ( $K_a/K_s$  0.02) in *C. alismatifolia*, but in other angiosperm lineages, it is the most variable gene [16,55]. We discovered that the *clpP* gene contains two introns, which is consistent with previous research and may contribute to the low  $K_a/K_s$  [10,55–58]. More than half of the  $K_a/K_s$  values of the genes in the *C. alismatifolia* accessions were less than 1, suggesting that they were subject to purifying selection (Figure 4b,c). Only two genes had  $K_a/K_s$  values above 1, which were *rps15* (1.15) and *ndhI* (1.13) with *petA* equal to 1.

In this work, all 56 plastomes were examined to determine the types, quantities, and distribution of repeat sequences. The regions were found to be variable and had the potential to be useful in identifying haplogroups. The phylogenetic tree clades were not clearly associated with flower color, indicating that extensive hybridization of *C. alismatifolia* may have decoupled some morphological characters from maternal inheritance. Interestingly, some phylogenetic patterns derived from functional gene alignments were also reflected in non-functional molecular evolution such as in Ch50 and Ch52, which were resolved on the same branch and possessed a greater number of repeats; they are currently classified as unique *C. alismatifolia* accessions based on these shared, derived sequences (Figure 6). An analysis of the sequence divergence revealed that the intraspecific variation between *C. alismatifolia* accessions was high, suggesting rapid evolution and/or a long time since divergence from sister lineages as well as the possibility that unrecognized cryptic lineages are present. Some morphological characteristics were found to be consistent within a given maternal lineage. Our results demonstrate the potential of plastomic data in resolving maternal relationships in the face of hybridization and polyploidization, which are common in *Curcuma*. The morphological characteristics of different *C. alismatifolia* accessions are very similar to that of other species of *Curcuma*, and from accessions like Ch43, it would appear that interspecific introgression may be contributing to such a morphological homogenization. When nuclear data are available, the data presented in this paper will be of great value in understanding the patterns of maternal introgression in the history of hybrid *Curcuma* cultivation as well as identifying cytonuclear incompatibles.

**Supplementary Materials:** The following support materials are available for download at <https://www.mdpi.com/article/10.3390/genes14091743/s1>, Figure S1: Phylogenetic tree of *C. alismatifolia* constructed on the CDS of 79 genes that code for proteins; Table S1: Sample information and feature of the *C. alismatifolia*; Table S2: Genes contained in the plastomes of 56 *C. alismatifolia* accessions; Table S3: Number of genes in plastid genomes from *C. alismatifolia*; Table S4: The  $K_s$ ,  $K_a$ , and  $K_a/K_s$  ratios for 79 genes that code for proteins of *C. alismatifolia*; Table S5: Variation of SNP locations and nucleotide among the several clades of *C. alismatifolia*; Table S6: Variation of INDEL locations and nucleotide among the several clades of *C. alismatifolia*.

**Author Contributions:** Z.W., L.R.T. and S.L. designed the study and revised the manuscript. J.W. and X.L. drafted the first-round manuscript and prepared the tables and figures. Y.Y. collected part of the samples and provided the photos and completed part of the analysis. J.W., X.L., Y.L., G.X. and S.K. performed the comparative genome analysis of the whole species. J.W., X.L., S.K. and L.N. performed the phylogenetic analysis. All authors discussed the results and commented on the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Guangdong Pearl River Talent Program (grants 2021QN02N792), the Science Technology and Innovation Commission of Shenzhen Municipality (RCYX20200714114538196), the National Natural Science Foundation of China (31970244 and 32170238), the Chinese Academy of Agricultural Sciences Elite Youth Program (110243160001007), the Scientific Research Foundation for Principal Investigators and Kunpeng Institute of Modern Agriculture at Foshan (KIMA-QD2022004) to Zhiqiang Wu. We thank the help from members of the Wu Lab.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Taheri, S.; Abdullah, T.L.; Noor, Y.M.; Padil, H.M.; Sahebi, M.; Azizi, P. Data of the first de novo transcriptome assembly of the inflorescence of *Curcuma alismatifolia*. *Data Brief* **2018**, *19*, 2452–2454. [PubMed]
2. Paisooksantivatana, Y.; Kako, S.; Seko, H. Genetic diversity of *Curcuma alismatifolia* Gagnep. (*Zingiberaceae*) in Thailand as revealed by allozyme polymorphism. *Genet. Resour. Crop Evol.* **2001**, *48*, 459–465.
3. Ghani, A.; Haque, S.; Alam, M.A.; Hossain, M.M.; Majumder, M.M.; Siddiqua, S.A.; Hasan, S.M.R.; Akter, R. Evaluation of Analgesic and Antioxidant Potential of the Leaves of *Curcuma alismatifolia* Gagnep. *Stamford J. Pharm. Sci.* **1970**, *1*, 3–9.
4. Taheri, S.; Abdullah, T.L.; Rafii, M.Y.; Hari Krishna, J.A.; Werbrouck, S.P.O.; Teo, C.H.; Sahebi, M.; Azizi, P. De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (*Zingiberaceae* family) through Illumina sequencing. *Sci. Rep.* **2019**, *9*, 3047. [PubMed]
5. Liao, X.; Ye, Y.; Zhang, X.; Peng, D.; Hou, M.; Fu, G.; Tan, J.; Zhao, J.; Jiang, R.; Xu, Y.; et al. The genomic and bulked segregant analysis of *Curcuma alismatifolia* revealed its diverse bract pigmentation. *ABioTech* **2022**, *3*, 178–196. [PubMed]
6. Gui, L.; Jiang, S.; Xie, D.; Yu, L.; Huang, Y.; Zhang, Z.; Liu, Y. Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution. *Gene* **2020**, *732*, 144355.
7. Mallet, J. Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **2005**, *20*, 229–237.
8. Kress, W.J.; Prince, L.M.; Williams, K.J. The phylogeny and a new classification of the gingers (*Zingiberaceae*): Evidence from molecular data. *Am. J. Bot.* **2002**, *89*, 1682–1696.
9. Deng, J.; Liang, H.; Zhang, L.; Zhang, W.; Zhang, G.; Luo, X.; Yang, R.; Shafique Ahmad, K. Evaluation on genetic relationships among China's endemic *Curcuma* herbs by SRAP markers. *Plant* **2021**, *9*, 16. [CrossRef]
10. Liang, H.; Zhang, Y.; Deng, J.; Gao, G.; Ding, C.; Zhang, L.; Yang, R. The complete chloroplast genome sequences of 14 *Curcuma* species: Insights into genome evolution and phylogenetic relationships within Zingiberales. *Front. Genet.* **2020**, *11*, 802.
11. Gao, C.; Wu, C.; Zhang, Q.; Zhao, X.; Wu, M.; Chen, R.; Zhao, Y.; Li, Z. Characterization of chloroplast genomes from two salvia medicinal plants and gene transfer among their mitochondrial and chloroplast genomes. *Front. Genet.* **2020**, *11*, 574962. [PubMed]
12. Li, H.T.; Yi, T.S.; Gao, L.M.; Ma, P.F.; Zhang, T.; Yang, J.B.; Gitzendanner, M.A.; Fritsch, P.W.; Cai, J.; Luo, Y.; et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* **2019**, *5*, 461–470. [PubMed]
13. Li, H.T.; Luo, Y.; Gan, L.; Ma, P.F.; Gao, L.M.; Yang, J.B.; Cai, J.; Gitzendanner, M.A.; Fritsch, P.W.; Zhang, T.; et al. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol.* **2021**, *19*, 232.
14. Jansen, R.K.; Ruhlman, T.A. Plastid Genomes of Seed Plants. In *Genomics of Chloroplasts and Mitochondria*; Springer: Berlin, Germany, 2012; pp. 103–126.

15. Rogalski, M.; do Nascimento Vieira, L.; Fraga, H.P.; Guerra, M.P. Plastid genomics in horticultural species: Importance and applications for plant population genetics, evolution, and biotechnology. *Front. Plant Sci.* **2015**, *6*, 586.
16. Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134.
17. Ruhlman, T.A.; Zhang, J.; Blazier, J.C.; Sabir, J.S.M.; Jansen, R.K. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *Am. J. Bot.* **2017**, *104*, 559–572. [[PubMed](#)]
18. Zhou, J.; Zhang, S.; Wang, J.; Shen, H.; Ai, B.; Gao, W.; Zhang, C.; Fei, Q.; Yuan, D.; Wu, Z.; et al. Chloroplast genomes in *Populus* (*Salicaceae*): Comparisons from an intensively sampled genus reveal dynamic patterns of evolution. *Sci. Rep.* **2021**, *11*, 9471.
19. Alwadani, K.G.; Janes, J.K.; Andrew, R.L. Chloroplast genome analysis of box-ironbark *Eucalyptus*. *Mol. Phylogenet. Evol.* **2019**, *136*, 76–86.
20. Gitzendanner, M.A.; Soltis, P.S.; Wong, G.K.S.; Ruhfel, B.R.; Soltis, D.E. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *Am. J. Bot.* **2018**, *105*, 291–301. [[PubMed](#)]
21. Yang, Z.; Ma, W.; Yang, X.; Wang, L.; Zhao, T.; Liang, L.; Wang, G.; Ma, Q. Plastome phylogenomics provide new perspective into the phylogeny and evolution of Betulaceae (*Fagales*). *BMC Plant Biol.* **2022**, *22*, 611.
22. Wu, Z.Q.; Ge, S. The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Mol. Phylogenet. Evol.* **2012**, *62*, 573–578. [[CrossRef](#)]
23. Lu, R.-S.; Yang, T.; Chen, Y.; Wang, S.-Y.; Cai, M.-Q.; Cameron, K.M.; Li, P.; Fu, C.-X. Comparative plastome genomics and phylogenetic analyses of Liliaceae. *Bot. J. Linn. Soc.* **2021**, *196*, 279–293. [[CrossRef](#)]
24. Wang, J.; Fu, G.; Tembrock, L.R.; Liao, X.; Ge, S.; Wu, Z. Mutational meltdown or controlled chain reaction: The dynamics of rapid plastome evolution in the hyperdiversity of Poaceae. *J. Syst. Evol.* **2023**, *61*, 328–344. [[CrossRef](#)]
25. Hollingsworth, P.M.; Li, D.Z.; van der Bank, M.; Twyford, A.D. Telling plant species apart with DNA: From barcodes to genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2016**, *371*, 20150338. [[CrossRef](#)] [[PubMed](#)]
26. Magdy, M.; Ou, L.; Yu, H.; Chen, R.; Zhou, Y.; Hassan, H.; Feng, B.; Taitano, N.; Van der Knaap, E.; Zou, X.; et al. Pan-plastome approach empowers the assessment of genetic variation in cultivated *Capsicum* species. *Hortic. Res.* **2019**, *6*, 108. [[CrossRef](#)]
27. Ge, S.; Guo, Y. Evolution of genes and genomes in the genomics era. *Sci. China Life Sci.* **2020**, *63*, 602–605. [[CrossRef](#)]
28. Sugiura, M. History of chloroplast genomics. *Photosynth. Res.* **2003**, *76*, 371–377. [[CrossRef](#)]
29. Raubeson, L.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boore, 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. [[CrossRef](#)]
30. Lu, R.S.; Hu, K.; Zhang, F.J.; Sun, X.Q.; Chen, M.; Zhang, Y.M. Pan-Plastome of Greater Yam (*Dioscorea alata*) in China: Intraspecific Genetic Variation, Comparative Genomics, and Phylogenetic Analyses. *Int. J. Mol. Sci.* **2023**, *24*, 3341. [[CrossRef](#)] [[PubMed](#)]
31. Wang, J.; Liao, X.; Gu, C.; Xiang, K.; Wang, J.; Li, S.; Tembrock, L.R.; Wu, Z.; He, W. The Asian lotus (*Nelumbo nucifera*) pan-plastome: Diversity and divergence in a living fossil grown for seed, rhizome, and aesthetics. *Ornam. Plant Res.* **2022**, *2*, 2. [[CrossRef](#)]
32. Yao, J.L.; Cohen, D. Multiple gene control of plastome-genome incompatibility and plastid DNA inheritance in interspecific hybrids of *Zantedeschia*. *Theor. Appl. Genet.* **2000**, *101*, 400–406. [[CrossRef](#)]
33. Fishman, L.; Sweigart, A.L. When Two Rights Make a Wrong: The Evolutionary Genetics of Plant Hybrid Incompatibilities. *Annu. Rev. Plant Biol.* **2018**, *69*, 707–731. [[CrossRef](#)] [[PubMed](#)]
34. Porebski, S.; Bailey, L.G.; Baum, B.R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol.* **1997**, *15*, 8–15. [[CrossRef](#)]
35. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)] [[PubMed](#)]
36. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)] [[PubMed](#)]
37. Nurk, S.; Bankevich, A.; Antipov, D.; Gurevich, A.A.; Korobeynikov, A.; Lapidus, A.; Prjibelski, A.D.; Pyshkin, A.; Sirotkin, A.; Sirotkin, Y.; et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* **2013**, *20*, 714–737. [[CrossRef](#)]
38. Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* **2015**, *31*, 3350–3352. [[CrossRef](#)]
39. 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](#)]
40. 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](#)]
41. 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](#)]
42. Beier, S.; Thiel, T.; Munch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]

43. Brudno, M.; Do, C.B.; Cooper, G.M.; Kim, M.F.; Davydov, E.; Program, N.C.S.; Green, E.D.; Sidow, A.; Batzoglou, S. LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **2003**, *13*, 721–731. [[CrossRef](#)] [[PubMed](#)]
44. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **1997**, *13*, 555–556. [[CrossRef](#)] [[PubMed](#)]
45. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)] [[PubMed](#)]
46. 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](#)]
47. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [[CrossRef](#)] [[PubMed](#)]
48. Abdullah; Mehmood, F.; Shahzadi, I.; Waseem, S.; Mirza, B.; Ahmed, I.; Waheed, M.T. Chloroplast genome of *Hibiscus rosa-sinensis* (*Malvaceae*): Comparative analyses and identification of mutational hotspots. *Genomics* **2020**, *112*, 581–591. [[CrossRef](#)] [[PubMed](#)]
49. Zheng, G.; Wei, L.; Ma, L.; Wu, Z.; Gu, C.; Chen, K. Comparative analyses of chloroplast genomes from 13 *Lagerstroemia* (*Lythraceae*) species: Identification of highly divergent regions and inference of phylogenetic relationships. *Plant Mol. Biol.* **2020**, *102*, 659–676. [[CrossRef](#)]
50. Leliaert, F.; Verbruggen, H.; Vanormelingen, P.; Steen, F.; López Bautista, J.M.; Zuccarello, G.C.; De Clerck, O. DNA-based species delimitation in algae. *Eur. J. Phycol.* **2014**, *49*, 179–196. [[CrossRef](#)]
51. Weng, M.L.; Blazier, J.C.; Govindu, M.; Jansen, R.K. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* **2014**, *31*, 645–659. [[CrossRef](#)]
52. Zhu, A.; Guo, W.; Gupta, S.; Fan, W.; Mower, J.P. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **2016**, *209*, 1747–1756. [[CrossRef](#)] [[PubMed](#)]
53. Monroe, J.G.; Srikant, T.; Carbonell Bejerano, P.; Becker, C.; Lensink, M.; Exposito Alonso, M.; Klein, M.; Hildebrandt, J.; Neumann, M.; Kliebenstein, D.; et al. Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature* **2022**, *602*, 101–105. [[CrossRef](#)] [[PubMed](#)]
54. Barthet, M.M.; Pierpont, C.L.; Tavernier, E.K. Unraveling the role of the enigmatic *matk* maturase in chloroplast group IIA intron excision. *Plant Direct* **2020**, *4*, e00208. [[CrossRef](#)] [[PubMed](#)]
55. Williams, A.M.; Friso, G.; van Wijk, K.J.; Sloan, D.B. Extreme variation in rates of evolution in the plastid *Clp* protease complex. *Plant J.* **2019**, *98*, 243–259. [[CrossRef](#)]
56. Wang, J.; He, W.; Liao, X.; Ma, J.; Gao, W.; Wang, H.; Wu, D.; Tembrock, L.R.; Wu, Z.; Gu, C. Phylogeny, molecular evolution, and dating of divergences in *Lagerstroemia* using plastome sequences. *Hortic. Plant J.* **2023**, *9*, 345–355. [[CrossRef](#)]
57. Abdel Ghany, S.E.; LaManna, L.M.; Harroun, H.T.; Maliga, P.; Sloan, D.B. Rapid sequence evolution is associated with genetic incompatibilities in the plastid *Clp* complex. *Plant Mol. Biol.* **2022**, *108*, 277–287. [[CrossRef](#)]
58. Williams, A.M.; Carter, O.G.; Forsythe, E.S.; Mendoza, H.K.; Sloan, D.B. Gene duplication and rate variation in the evolution of plastid *ACCase* and *Clp* genes in angiosperms. *Mol. Phylog. Evol.* **2022**, *168*, 107395. [[CrossRef](#)]

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