Gene 873 (2023) 147479

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Chloroplast genome analysis of three *Acanthus* species reveal the adaptation of mangrove to intertidal habitats

Dongna Ma [a,b,1], Qiansu Ding [d,1], Zhizhu Zhao [e], Xiao Han [a], Hai-Lei Zheng [c,*]

[a] National Engineering Research Center of Cereal Fermentation and Food Biomanufacturing, State Key Laboratory of Food Science and Technology, School of Food Science and Technology, Jiangnan University, Wuxi 214122, China
[b] Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China
[c] Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Xiamen 361102, China
[d] Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA
[e] School of Biological Sciences, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Edited by: Tzen-Yuh Chiang

Keywords:
Acanthus
Chloroplast genome
Nucleotide variability
Positive selection
Phylogenetic tree

ABSTRACT

Acanthus is a distinctive genus that covers three species with different ecological niches including *Acanthus mollis* (arid terrestrial), *Acanthus leucostachyus* (damp forest) and *Acanthus ilicifolius* (coastal intertidal). It is an intriguing question how these species evolved from terrestrial to coastal intertidal. In the present study, we assembled chloroplast genomes of *A. ilicifolius*, *A. leucostachyus* and *A. mollis*, which exhibited typical quadripartite structures. The sizes were 150,758, 154,686 and 150,339 bp that comprised a large single copy (LSC, 82,963, 86,461 and 82,612 bp), a small single copy (SSC, 17,191, 17,511 and 17,019 bp), and a pair of inverted repeats (IRs, 25,302, 25,357 and 25,354 bp), respectively. Gene annotation revealed that *A. ilicifolius*, *A. leucostachyus* and *A. mollis* contained 113, 112 and 108 unique genes, each of which contained 79, 79 and 74 protein-coding genes, 30, 29 and 30 tRNAs, and 4 rRNA genes, respectively. Differential gene analysis revealed plenty of *ndhs* gene deletions in the terrestrial plant *A. mollis*. Nucleotide diversity analysis showed that the *psbK*, *ycf1*, *ndhG*, and *rpl22* have the highest nucleotide variability. Compared to *A. leucostachyus* and *A. mollis*, seven genes in *A. ilicifolius* underwent positive selection. Among them, the *atpF* gene showed a strong positive selection throughout terrestrial to marine evolution and was important for adaptation to coastal intertidal habitats. Phylogenetic analysis indicated that *A. ilicifolius* has a closer genetic relationship with *A. leucostachyus* than *A. mollis* which further confirmed the evolutionary direction of *Acanthus* going from terrestrial to coastal intertidal zones.

1. Introduction

Mangrove plants are native to tropical and subtropical tidal-accessible coastal zones, and their exceptional stress tolerance makes them important materials for studying plant tolerance to abiotic stresses, such as salinity, flooding, high light, and high-temperature (Ball, 1988). How mangrove plants acquired this ability during their evolution is not fully understood. The mangrove *Acanthus ilicifolius* is a shrub of the genus *Acanthus* in the family Acanthaceae, with its close relatives in the same genus, the *Acanthus leucostachyus*, growing in tropical terrestrial forests, and the *Acanthus mollis*, which grows in temperate or subtropical arid regions (Daniel, 2003). These three closely related species in different ecological niches are ideal materials for studying the evolution of plants from terrestrial to coastal and from adaptation to drought to saline flooding. In addition, *A. ilicifolius* is a medicinal plant, and the whole plants or roots can be used as traditional Chinese medicine, which has the effects of heat-clearing and detoxifying, reducing swelling and dissipating knots, and relieving cough and asthma (Singh and Aeri, 2013).

Chloroplast is an important organelle of plant cells and is the site of photosynthetic reactions (Pfannschmidt et al., 1999). It is a semi-autonomous organelle with a matrilineal mode of inheritance (Palmer et al., 1983). The higher plant chloroplast genome is typically 120 to 160 kb in size of circular double-stranded DNA and usually consists of

Abbreviations: Bp, base pair(s); kb, kilobase(s) or 1000 bp; rRNA, ribosomal RNA.

* Corresponding author.
E-mail address: zhenghl@xmu.edu.cn (H.-L. Zheng).

[1] These authors contributed equally to this article.

https://doi.org/10.1016/j.gene.2023.147479
Received 4 March 2023; Received in revised form 24 April 2023; Accepted 8 May 2023
Available online 12 May 2023

four parts: a large single copy (LSC) region, a small single copy (SSC) region, and a pair of inverted complementary repeat (IRs) regions. Generally speaking, the chloroplast genome encodes 110 to −130 genes approximately. Despite its relatively stable structure, some mutational events, including insertions, deletions, substitutions, and translocations, inevitably occurred during the evolution resulting in changes in the size and the gene content of the chloroplast genome (Lei et al., 2016). Chloroplast genomes are of great value in revealing the origins, evolution and relationships of different species. They have been widely used in molecular identification, population genetics, plastid engineering, and phylogenetic analysis (Sun et al., 2021, Yang et al., 2020).

In this study, we performed *de novo* assembly of chloroplast genomes of three species of *Acanthus* using high-throughput sequencing technology. The composition, structure and phylogenetic relationships of the chloroplast genomes were analyzed to clarify evolutionary relationships and phylogenetic positions of them. Molecular evidence was provided for verifying the hypothesis that mangrove plants migrated from inland to the coast, and the sequence information was also made available for the molecular identification of the *Acanthus* plant.

2. Materials and methods

2.1. Leaves collections of Acanthus

In this study, the leaves material of three species of *Acanthus* were collected separately. *A. ilicifolius* were collected from the Zini mangrove forest, south of the Jiulong River Estuary, Fujian Province, China. *A. mollis* were collected from mature plants provided by the Shanghai Institute of Garden. *A. leucostachyus* were collected from the Xishuangbanna Tropical Botanical Garden in Xishuangbanna, Yunnan Province, China. Fresh leaves collected from various locations were dried and preserved in silica gel.

2.2. DNA extraction, library preparation and sequencing

DNA was extracted by a modified CTAB method, DNA quality was determined by 1.0% agarose gel electrophoresis, and DNA concentration was quantified by Qubit 3.0 (Thermo Fisher Scientific, USA) fluorescence quantification. The total genomic DNA of each sample was prepared to construct sequencing libraries according to Illumina's standard protocol. Firstly, the quality of genomic DNA was tested and then a series of procedures were carried out to form sequencing libraries, including sonication, end repairs, 3′-terminal plus A, the addition of adaptors, selection of the size of fragments, and PCR amplification. After the inspection of quality, the constructed libraries were sequenced by Illumina Hiseq 2500 platform.

2.3. Chloroplast genome assembly

We used NGSQCToolkit (v2.3.3) (Patel and Jain, 2012) software to remove adaptors and to filter out the reads whose Q-score ≤ 5 accounted for above 50% of the bases to obtain high-quality clean reads. Thereafter, SPAdes (Bankevich et al., 2012) software was used to assemble contigs from the clean reads. The contigs were then concatenated using SSPACE (Boetzer et al., 2011) software to form scaffolds, and finally, a complete chloroplast genome was obtained using Gapfiller (Boetzer and Pirovano, 2012) software by complementing the scaffold sequences.

2.4. Structural annotation of chloroplast genome

The well-assembled chloroplast genomes were annotated using PGA (Qu et al., 2019) software, and the annotation results were corrected in Geneious (v10.22) (Kearse et al., 2012) software. The gene annotation information for rRNA was obtained by comparing the chloroplast genomes with that in the NCBI database using hmmer (v3.3.2) (Mistry et al., 2013) software. The Aragorn (Laslett and Canback, 2004) software was used to predict the tRNA of the chloroplast genomes. The representation of a circular map of three chloroplast genomes was depicted by the online program Organellar Genome DRAW (OGDRAW) (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (Lohse et al., 2013).

2.5. Comparative analysis of chloroplast genome

We performed variable site analysis of protein-coding and intergenic regions of the three *Acanthus* chloroplast genomes using DnaSP (Librado and Rozas, 2009). DAMBE (Xia, 2018) software was used to acquire the relative synonymous codon value (RSCU). The RSCU is equal to the ratio of the usage frequency of a particular codon divided by the expected frequency of that codon. If the RSCU ≥ 1, the codon is used more frequently than expected; while if the RSCU 1, the codon is used less frequently. Simple sequence repeats (SSRs) were identified by Misa (v2.1) (Beier et al., 2017). We defined a criterion of minimum repeat size ≥ 10 bases for mononucleotides, at least six repeats for dinucleotides, and five repeats for trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides. A distance threshold of 100 bases between two SSRs was set for compound microsatellites.

2.6. Phylogenetic analysis

We selected the chloroplast genomes of 17 closely genetically related species of the family Acanthaceae for phylogenetic analysis and the *Vitis vinifera* chloroplast genome served as an outgroup. The chloroplast genome sequences of 14 plants of the family Acanthaceae were downloaded from NCBI. We selected IRs, LSC, SSC, CDS, and complete chloroplast genome sequences for phylogenetic analysis. Firstly, we performed multiple sequence alignment by MAFFT (Katoh and Standley, 2013) software. And then, Maximum likelihood (ML) and Maximum parsimony (MP) analysis in MEGA7 (Kumar et al., 2016) software was applied to infer phylogenetic relationships within the evolutionary branch of the family Acanthaceae with 1000 repetitions.

2.7. Positive selection analysis

To analyze whether CDS regions in chloroplast genome sequences were subjected to positive selection during the evolution of *Acanthus* species, the synonymous_calc.py (https://github.com/tanghaibao/bio-pipeline) with the Nei-Gojobori method was used to calculate the rate of nonsynonymous (Ka) and synonymous substitutions (Ks) for each gene (Nei and Gojobori, 1986). And in this study, if $Ka/Ks > 1$, species evolution is considered to have a positive selection effect on genes (the gene is positively selected). If $Ka/Ks = 1$, the neutral selection is considered to exist (the gene is considered to be neutral during evolution). If $Ka/Ks < 1$, a purifying selection effect is considered (the gene is negatively selected).

3. Results

3.1. Assembly and structural characteristics of the chloroplast genomes of three Acanthus species

By Illumina sequencing, 653,697, 591,954, and 942,249 paired-end reads were generated for *A. ilicifolius*, *A. leucostachyus* and *A. mollis*, respectively. The average organelle coverage for *A. ilicifolius*, *A. leucostachyus* and *A. mollis* reached 1350, 1217, and 2205, correspondingly. The total genome lengths were 150,758, 154,686 and 150,339 bp, comprising a LSC (82,963, 86,461 and 82,612 bp), a small single copy region (SSC) (17,191, 17,511 and 17,019 bp) and inverted complementary repeats (IRA and IRB) (25,302, 25,357 and 25,354 bp), respectively (Fig. 1).

The GC content of the chloroplast genomes of the three *Acanthus* species were 38.41%, 37.42% and 38.12%, respectively, with a mean GC

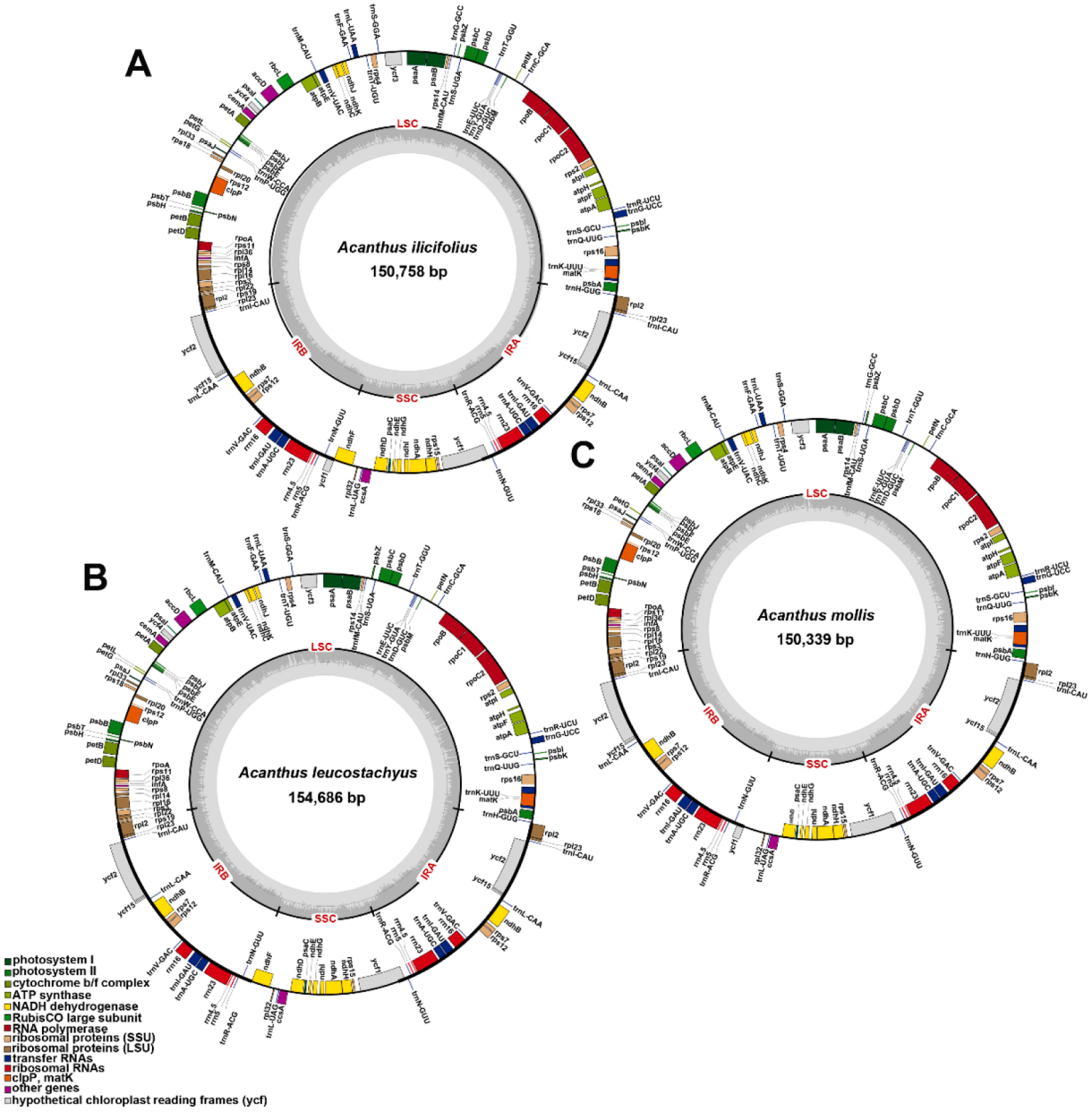

Fig. 1. Chloroplast genome maps of three *Acanthus* species. **A,** *A. ilicifolius* chloroplast genome. **B,** *A. leucostachyus* chloroplast genome. **C,** *A. mollis* chloroplast genome. Genes shown outside the circle are transcribed clockwise and those inside counterclockwise. Genes belonging to different functional groups are color-coded.

content of 37.98%. The GC content of the LSC region (mean 35.87%) and the SSC region (mean 31.94%) were lower than that of the IRs region (mean 43.55%) in *Acanthus* species (Table 1). The higher GC content in the IRs region compared to the SSC region was probably due to the presence of rRNA genes in the IRs region.

Gene annotation revealed that the chloroplast genomes of *A. ilicifolius*, *A. leucostachyus* and *A. mollis* contained 113, 112 and 108 unique genes, of which 79, 79 and 74 were protein-coding, 30, 29 and 30 were tRNA, and 4 were rRNA genes, correspondingly (Table 1, 2). There were 18 duplicated genes in the *Acanthus* species, including 4 rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*) and 14 other genes (*ndhB*, *rpl2*, *rpl23*, *rps12*, *rps7*, *trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*, *ycf15*, and *ycf2*) that were repeated once, while the *ycf1* gene was only repeated in *A. ilicifolius* and *A. mollis* (Table S1). Furthermore, 20 intron-containing genes were identified in *Acanthus* species, including 15 different genes (*atpF*, *ndhB*, *petB*, *petD*, *rpl16*, *rpl2*, *rpoC1*, *rps12*, *rps16*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-*

Table 1
Summary statistics for the assembly of three *Acanthus* species chloroplast genomes.

Genome features	A. *ilicifolius*	A. *leucostachyus*	A. *mollis*
Genome length	150,758	154,686	150,339
LSC length	82,963	86,461	82,612
SSC length	17,191	17,511	17,019
IR length	25,302	25,357	25,354
Number of genes	133 (113)	131 (112)	131 (108)
Protein-coding gene (unique)	87 (79)	86 (79)	82 (74)
tRNA genes (unique)	37 (30)	36 (29)	37 (30)
rRNA genes (unique)	8 (4)	8 (4)	8 (4)
Duplicated genes in IR	18	17	18
GC content (%)	38.41	37.42	38.12
GC content in LSC (%)	36.43	35.05	36.12
GC content in SSC (%)	32.35	31.71	31.75
GC content in IR (%)	43.71	43.44	43.50
Aligned paired-end reads	653,697	591,954	942,249
Average organelle coverage	1350	1217	2205

Table 2
List of annotated genes in the *A. ilicifolius*, *A. leucostachyus*, and *A. mollis* chloroplast genomes.

Category	Group of gene	Name of gene
Photosynthetic	Subunits of photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Subunits of photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ* (*A. ilicifolius*)
	Subunits of NADH dehydrogenase	*ndhA* (*A. mollis*), *ndhH* (*A. mollis*), *ndhI* (*A. mollis*), *ndhK* (*A. mollis*), *ndhB**(2), *ndhC*, *ndhD*, *ndhE*, *ndhF* (*A. mollis*), *ndhG*, *ndhJ*
	Subunits of cytochrome *b*/f complex	*petA*, *petB**, *petD**, *petG*, *petL* (*A. mollis*), *petN*
	Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*
	Large subunit of rubisco	*rbcL*
Self-replication	Proteins of large ribosomal subunit	*rpl14*, *rpl16**, *rpl2**(2), *rpl20*, *rpl22*, *rpl23*(2), *rpl32*, *rpl33*, *rpl36*
	Proteins of small ribosomal subunit	*rps11*, *rps12***(2), *rps14*, *rps15*, *rps16**, *rps18*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7*(2), *rps8*
	Subunits of RNA polymerase	*rpoA* (*A. leucostachyus*), *rpoB*, *rpoC1**, *rpoC2*
	Ribosomal RNAs	*rrn16*(2), *rrn23*(2), *rrn4.5*(2), *rrn5*(2)
	Transfer RNAs	*trnA-UGC** (2), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG-GCC* (*A. leucostachyus*), *trnG-UCC**, *trnH-GUG*, *trnI-CAU* (2), *trnI-GAU**(2), *trnK-UUU**, *trnL-CAA* (2), *trnL-UAA**, *trnL-UAG*, *trnM-CAU*, *trnN-GUU* (2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG* (2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC* (2), *trnV-UAC**, *trnW-CCA*, *trnY-GUA*, *trnfM-CAU*
Biosynthesis	Maturase	*matK*
	Protease	*clpP***
	Envelope membrane protein	*cemA*
	Acetyl-CoA carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
	Translation initiation factor	*infA*
Unknown function	Conserved hypothetical chloroplast Reading Frames	*ycf1*(2), *ycf15*(2), *ycf2*(2), *ycf3***, *ycf4*

UAA, and *trnV-UAC*) containing one intron and *ndhA* in *A. ilicifolius* and *A. leucostachyus*. The *trnH-GUG* gene in *A. ilicifolius* and *A. mollis* also contained one intron. *ClpP* and *ycf3* genes had two introns (Table S2). These 12 introns were located in the LSC region; one intron was in the SSC region, and five introns were in the IR region (Table S2). We speculated that the plastid-encoded genes *ndhI*, *ndhF*, *ndhH*, *ndhA*, *ndhK*, and *petL* might be lost in *A. mollis*. The plastid-encoded *rpoA* gene has been lost in *A. leucostachyus*, and *psbZ* has been lost in *A. ilicifolius* (Table 2). The complete and well-annotated chloroplast genomes have been submitted to NCBI under GenBank accession numbers MW752129 for *A. ilicifolius*, OM022237 for *A. leucostachyus*, and OM022238 for *A. mollis*.

3.2. Analyses of long repeats and simple sequence repeats (SSRs)

We used REPuter software to identify the number of long repeats in the three chloroplast genomes of *Acanthus*, including four types (forward repeats, reverse repeats, complement repeats, and palindromic repeats). A total of 147 long repeats were identified in the *Acanthus* species, including 48 forward repeats, 31 reverse repeats, 8 complement repeats, and 60 palindromic repeats (Table S3). This indicated that the percentage of palindromic repeat sequences was the highest among the four repeat types (40.82%), while complement repeats were the lowest (5.44%) (Fig. 2A). The chloroplast genomes of *A. ilicifolius*, *A. leucostachyus* and *A. mollis* had 15, 11 and 22 forward repeats, 11, 15 and 5 reverse repeats, 3, 4 and 1 complement repeats, and 20, 19 and 21 palindromic repeats, respectively (Table S3). Among them, *A. mollis* have the largest number of palindromic repeats and forward repeats (Fig. 2B). The length distribution of long repeats in *A. ilicifolius* and *A. mollis* were mainly 20–30 bp, while long repeats with 31–40 bp were found to be the most common in the *A. leucostachyus* chloroplast genome (Fig. 2C). In conclusion, the number, length and distribution of long repeats among the *Acanthus* species were unevenly distributed.

SSRs annotation showed that a total of 331 SSRs were found in the species of the *Acanthus*. There were 98, 138 and 95 SSRs for *A. ilicifolius*, *A. leucostachyus* and *A. mollis*, respectively (Table S3). The average number of SSRs among the three *Acanthus* species was 110, with the largest number of SSRs in *A. leucostachyus* and the lowest SSRs in *A. ~~ilicifolius~~ mollis*. We identified six types of SSRs, including mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide. Of these SSRs, only the chloroplast genome of *A. leucostachyus* exhibited all six types of repeats. No tetranucleotide and hexanucleotide repeats were found in the chloroplast genome of *A. mollis* and *A. ilicifolius*. In the *A. ilicifolius*, *A. leucostachyus* and *A. mollis* chloroplast genomes, the most frequent mononucleotide were found by 81, 89 and 78, respectively, accounting for 82.65%, 64.49% and 82.11% of all SSRs. This was followed by dinucleotide and trinucleotide repeats with 9, 17 and 7 (9.18%, 12.32% and 7.37%), and 4, 18 and 6 (4.08%,13.04% and 6.32%) (Fig. 3A; Table S3), correspondingly. Further analysis revealed that the majority of SSRs were located in the LSC regions (55, 67 and 47 loci, 66.27, 66.33 and 61.04%) than in the SSC regions (14, 11 and 13 loci, 16.87, 10.89 and 16.88%) and IRs regions (14, 23 and 17 loci, 16.87, 22.77 and 22.08%) of the three *Acanthus* chloroplast genomes (Fig. 3B; Table S3). In the *A. ilicifolius*, *A. leucostachyus* and *A. mollis* chloroplast genomes, the majority of mononucleotide SSRs were A/T repeats accounting for 93.83%, 95.51% and 92.31% of all mononucleotide types, respectively (Fig. 3C; Table S4). For dinucleotide repeats, only AT/AT repeats were observed in three species (Fig. 3C; Table S4). Among the trinucleotide classes, AAT/ATT repeats were the most abundant type, accounting for 75%, 88.89% and 83.33% of the motifs in this class (Fig. 3C; Table S4).

3.3. Codon preference

Codon usage frequencies and the RSCU values were calculated based on the sequences of protein-coding genes in the chloroplast genomes of *Acanthus*. The *A. ilicifolius*, *A. leucostachyus* and *A. mollis* chloroplast genomes consisted of 26,467, 26,067, and 24,584 codons, respectively. Among the amino acids encoded, leucine (Leu) was the most abundant amino acid with a frequency of 10.70, 10.73 and 10.27%, followed by isoleucine (Ile) (8.42, 8.44 and 8.46%), while cysteine (Cys) had a minimum number with a frequency of 1.11, 1.15 and 1.12% (Fig. 4A; Table S5). This result is concordant with many previous observations of angiosperms (Raman and Park, 2016, Park et al., 2017, Gichira et al., 2017). Due to the value of RSCU > 1.00, 32, 33 and 31 codons showed codon usage bias in the *A. ilicifolius*, *A. leucostachyus* and *A. mollis*, respectively (Fig. 4B; Table S6). Notably, all preferred codons except UUG, UCC and AUG were A/U-ending codons (Fig. 4B, Table S6), which supported the opinion that this biased use of specific degenerate codons was probably the result of adaptive evolution of the chloroplast genomes. Conversely, most of the codons ending in C/G had RSCU < 1.00, indicating that these codons were less common in the chloroplast genomes of the *Acanthus*. Stop codons were biased towards UAA (RSCU > 1.00) (Fig. 4B; Table S6). In addition, we also observed that tryptophan

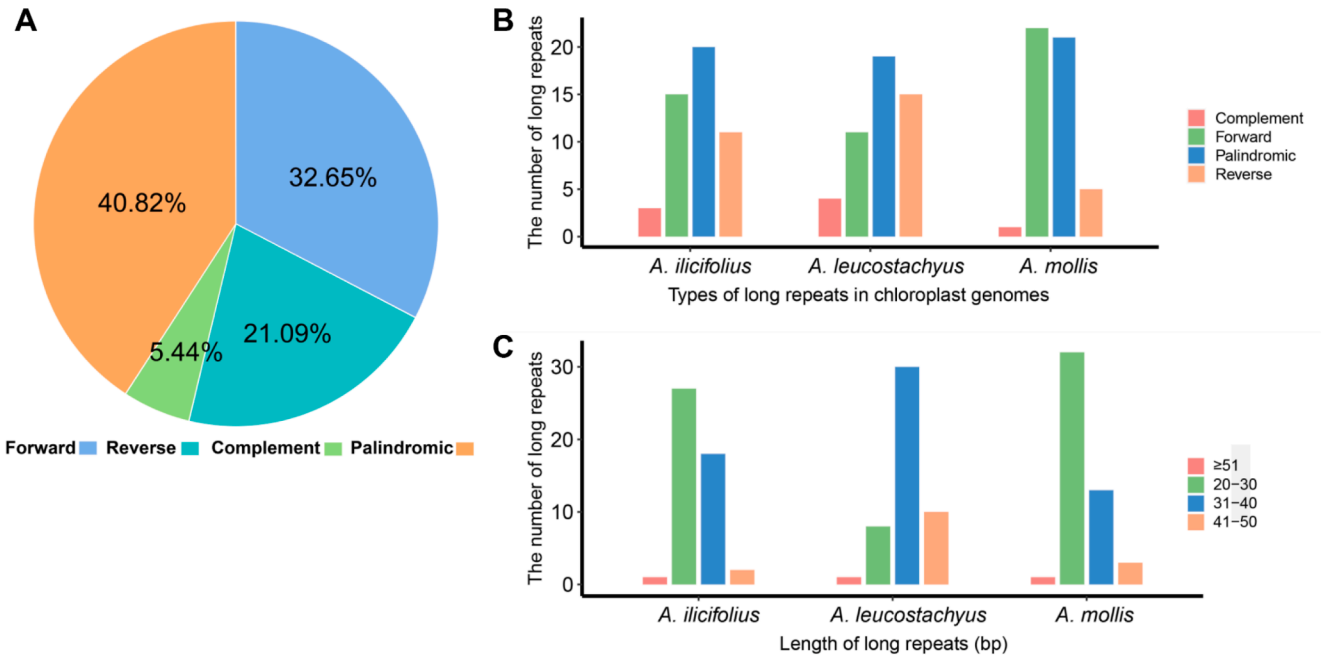

Fig. 2. Long repeat sequences in the three chloroplast genomes of *Acanthus*. **A,** Percentage of four repeat types in *Acanthus*. **B,** Total of four long repeat types in the three chloroplast genomes of *Acanthus*. **C,** Length distribution of long repeat sequences in the three chloroplast genomes of *Acanthus*.

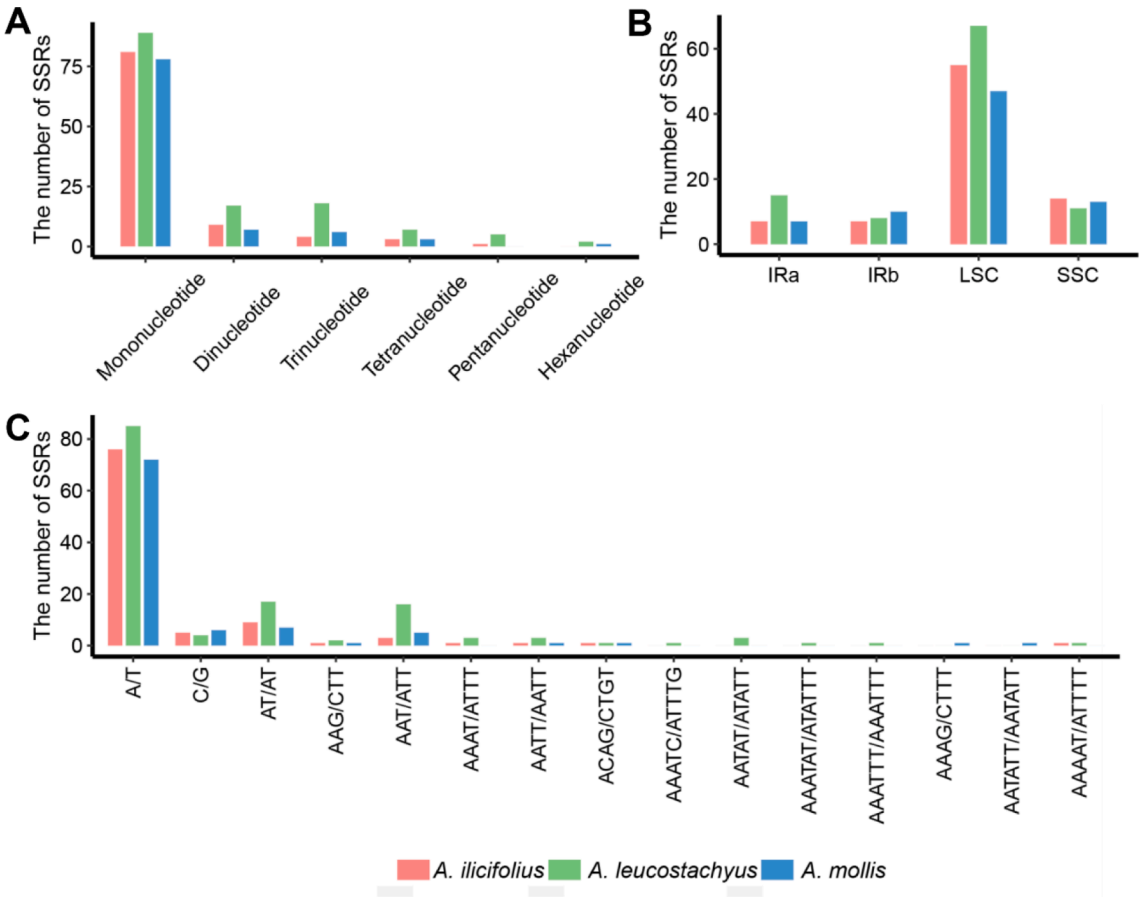

Fig. 3. Simple sequence repeats (SSRs) sequences in the three chloroplast genomes of *Acanthus*. **A,** Number of different SSR types detected in the three chloroplast genomes of *Acanthus*. **B,** Percentage of SSRs in the chloroplast distribution, LSC: Large single copy region, SSC: Small single copy region, IR: Inverted repeat region. **C,** The frequency of the identified SSRs in different repeat class types.

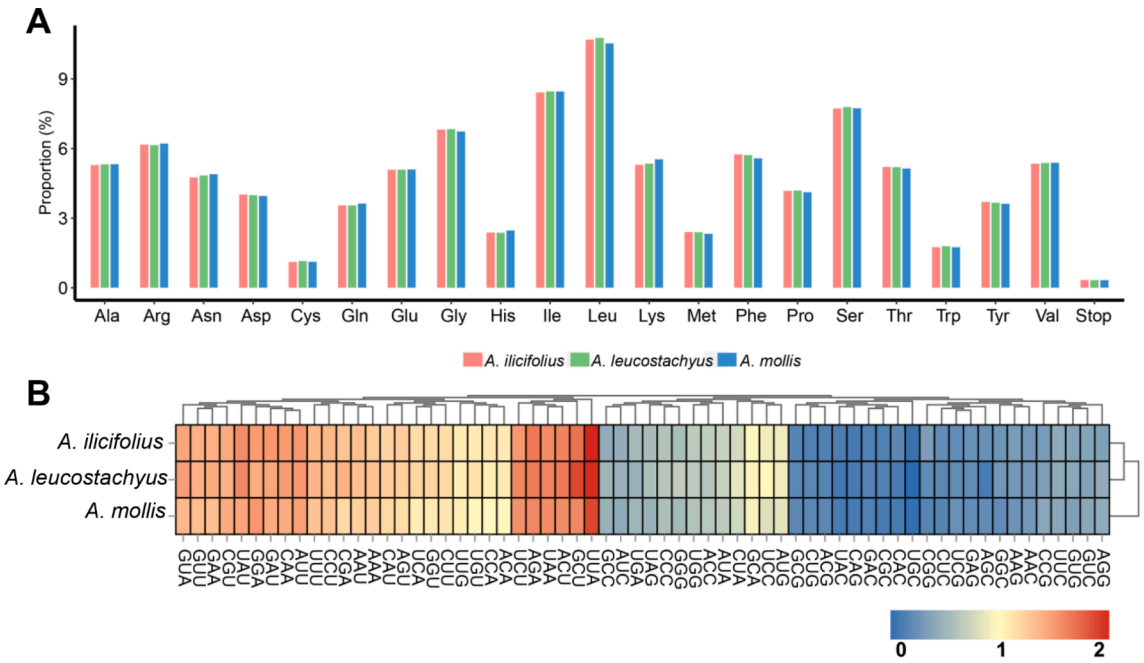

Fig. 4. Codon content in all protein coding genes in the three chloroplast genomes of *Acanthus*. **A,** Amino acids and stop codons proportion in protein coding genes of three chloroplast genomes of *Acanthus*. **B,** Heat map analysis for codon distribution of all protein coding genes of three chloroplast genomes of *Acanthus*. Red color indicates higher RSCU values and blue color indicates lower RSCU values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Trp) did not show codon preference in any of the three *Acanthus* species (Fig. 4B; Table S6).

3.4. IR expansion and contraction

We used the IRscope to compare the boundaries of the SSC/IRs/LSC regions among the three species in the *Acanthus* and found that the IRs regions were highly conserved, with sizes of 25,354, 25,375 and 25,302, respectively. The genes located in the boundary regions of LSC/IRb, SSC/IRa and IRa/LSC were also highly conserved. Within all three chloroplast genomes, the boundary of LSC/IRb was located between gene *rps19* and *rpl2*. There were 323, 18 and 302 bp between *rps19* and the LSC/IRb boundary, and the distances between *rpl2* and the LSC/IRb boundary were 36, 46 and 39 bp of the three species, respectively (Fig. 5). **ycf1-ndhF* genes were located at the boundary of the IRb/SSC region. However, the genes at the IRb/SSC boundary were inconsistent in three species. In *A. mollis*, the *ndhF* gene was non-existent in the boundary of the IRb/SSC region. And for *A. leucostachyus*, the **ycf1* gene was absent. However, in the *A. ilicifolius*, the boundary of IRb/SSC was located in the middle of **ycf1* and *ndhF* (Fig. 5). The spaces between the *ndhF* and IRb/SSC boundary were 31 bp and 38 bp in the *A. ilicifolius* and *A. leucostachyus*. The SSC/IRa boundary was situated in the internal region of *ycf1* gene which crossed into the boundary in these three species. However, the length of *ycf1* in the IRa region varied among the *Acanthus* species and were 817, 827 and 821 bp, respectively (Fig. 5). The *rpl2* and *trnH* genes of all three *Acanthus* species were located at the boundary of the IRa/LSC region, and the *trnH* gene was at a distance of 51, 28 and 73 bp from the IRa/LSC boundary, respectively (Fig. 5). In general, the boundaries of the three species were dynamically diverse.

3.5. Divergence sequence hotspots in Acanthus

To investigate the hotspots of differentiation within different regions among the three *Acanthus* species, we analyzed the nucleotide diversity (Pi) values via DnaSP. The results showed that the Pi values for each locus of protein-coding regions had a varied range of 0 to 0.09857 with

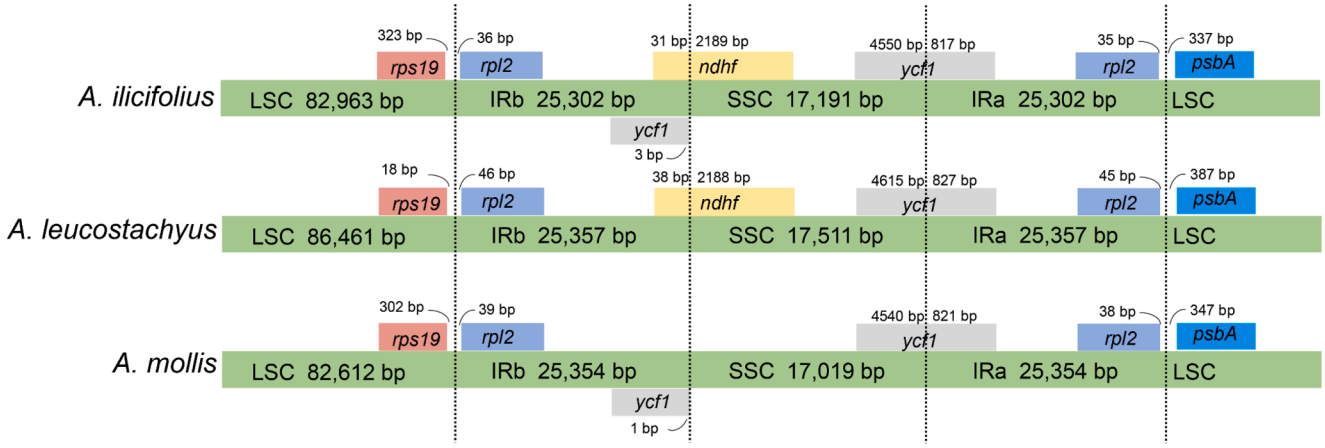

Fig. 5. Comparison of the border positions of LSC, SSC, and IR regions in the three chloroplast genomes of *Acanthus*. Gene names are indicated in boxes.

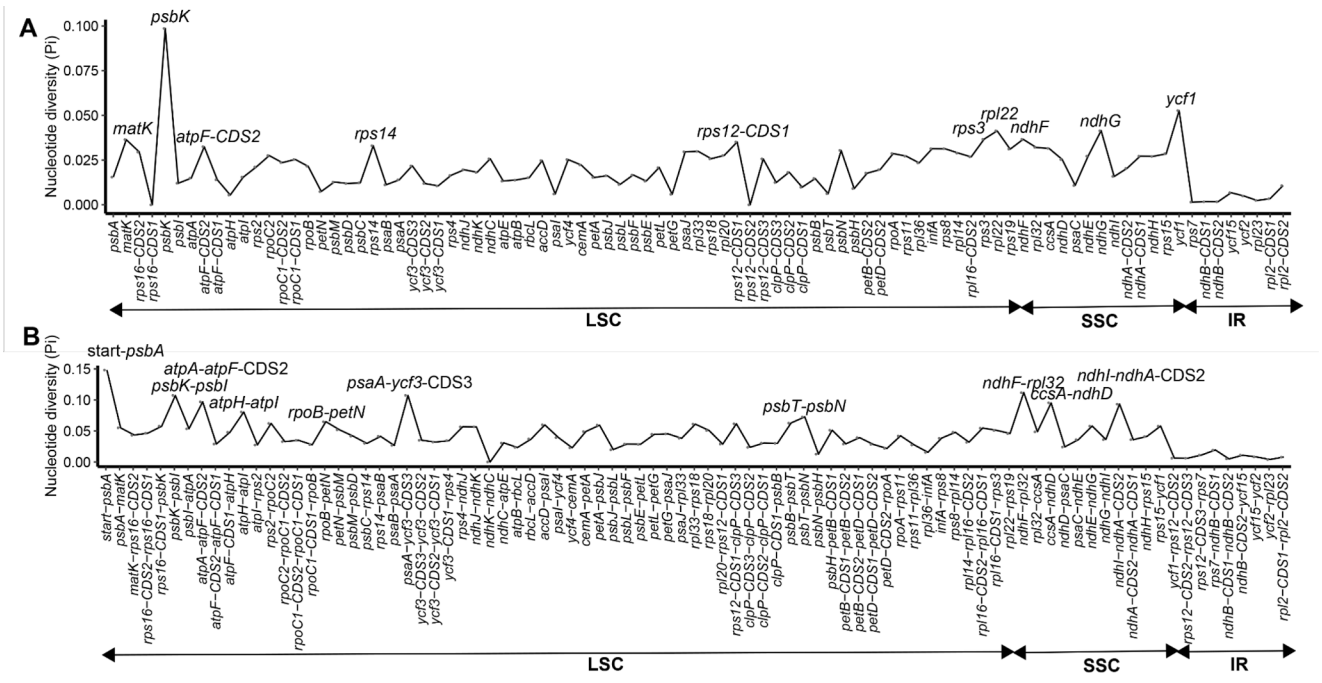

Fig. 6. Nucleotide diversity (Pi) values of various regions in the three chloroplast genomes of *Acanthus*. **A,** Protein coding region. Peak regions with a Pi value of > 0.032 were labeled with loci tags of genic names. **B,** Intergenic regions. Peak regions with a Pi value of > 0.065 were labeled with loci tags of intergenic region names.

an average value of 0.02051 (Table S8). For these protein-coding regions, ten regions (*psbK*, *ycf1*, *ndhG*, *rpl22*, *rps3*, *ndhF*, *matK*, *rps12*-CDS1, *rps14*, and *atpF*-CDS2) showed remarkably high values (Pi > 0.03244, Fig. 6a). Within the intergenic regions, the Pi values ranged from 0 to 0.14815 and had an average of 0.042971 (Table S8). Among intergenic regions, ten most divergent regions (start-*psbA*, *ndhF-rpl32*, *psaA-ycf3*-CDS3, *psbK-psbI*, *atpA-atpF*-CDS2, *ccsA-ndhD*, *ndhI-ndhA*-CDS2, *atpH-atpI*, *psbT-psbN*, and *rpoB-petN*) with Pi values ranging from 0.06502 to 0.14815 were identified (Fig. 6B).

3.6. Genetic selection pressure analysis

Ka/Ks is widely used to infer genomic evolutionary rates and selection pressure on individual genes (Gao et al., 2018), with Ka/Ks < 1, Ka/Ks = 1, and Ka/Ks > 1, indicating genes subjected to purifying selection, neutral selection, and positive selection, correspondingly. In the present study, Ka/Ks values were calculated for a total of 72 protein-coding genes in the chloroplast genomes of *Acanthus*. The results showed that a total of 4 genes (*atpF*, *rps11*, *psaJ*, and *rpl33*) in *A. mollis* and *A. leucostachyus* (Table S9), a total of 3 genes (*ccsA, atpF* and *rpl32*) in *A. leucostachyus* and *A. ilicifolius* (Table S9), and a total of 4 genes (*psaJ, ycf4, rps16*, and *psbK*) in *A. mollis* and *A. ilicifolius* (Table S9) were obtained with a Ka/Ks ratio that was significantly>1, respectively. By comparing the *A. ilicifolius* and *A. mollis*, we found that a total of 7 genes underwent positive selection in *A. ilicifolius* (Fig. 7). These genes may be of importance in the adaptation of mangrove plants to the harsh coastal intertidal environment. The results from Ka/Ks also revealed that the degree of differentiation between *A. ilicifolius* and *A. leucostachyus* was less than that between *A. ilicifolius* and *A. mollis*, illustrating a closer relationship between *A. ilicifolius* and *A. leucostachyus* (Table S9).

To further unravel the mechanism of the evolutionary process of these three *Acanthus* species from arid-jungle to intertidal, we calculated the Ka/Ks ratios in two group of genes (*A. leucostachyus* and *A. mollis*, *A. ilicifolius* and *A. leucostachyus*). There were 6 genes with Ka/Ks > 0.5

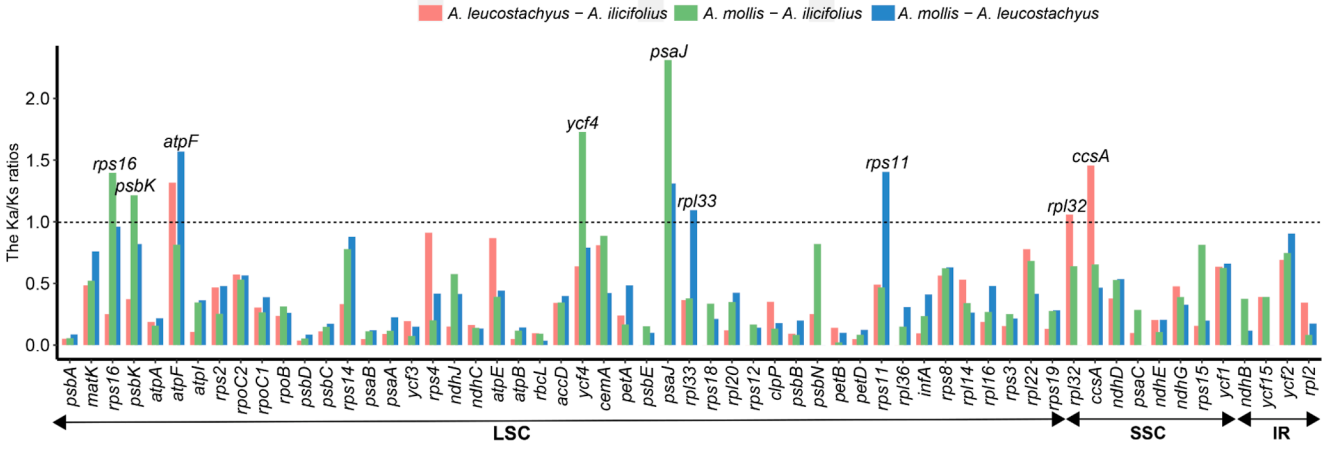

Fig. 7. Protein coding genes in the three chloroplast genomes of *Acanthus* were used to comparatively analyze non-synonymous (Ka) and synonymous (Ks) ratios. Genes with Ka/Ks > 1 were labeled with loci tags.

(Table S10), and this result showed that these 6 genes were particularly vital in the evolution of the three related species of *Acanthus* from land to coastal intertidal habitats. Among them, the *atpF* gene showed a strong positive selection during terrestrial to marine evolution (*A. mollis*-*A. leucostachyus*:1.57 and *A. leucostachyus*-*A. ilicifolius*:1.32).

3.7. Chloroplast phylogenetic analysis

Phylogenetic relationships of the Acanthaceae family and taxonomic status were systematically investigated through analysis of chloroplast sequences. In this study, we combined 15 published complete chloroplast genomes and 3 chloroplast genomes of *Acanthus* to construct the phylogenetic trees using MEGA7 software. We used five different types of data, including the complete chloroplast genome sequences, CDS, LSC, SSC, and IRs regions for constructing the phylogenetic trees (Fig. 8, Table S11). Consequently, the phylogenetic trees constructed with complete chloroplast genomes, CDS, LSC, and SSC data had the same topology, whereas the tree constructed from the IRs dataset had a minor difference. On one hand, the result showed that *A. ilicifolius*, *A. leucostachyus* and *A. mollis* form a single group. On the other hand, our results also showed that the genera *Echinacanthus* and *Justicia* form one branch. In addition, the *Acanthus*, *Barleria prionitis* and *Aphelandra knappiae* were inferred to be closely related to a clade comprising all the other genera in the family Acanthaceae. Going further, the phylogenetic trees constructed based on chloroplast genomes, CDS, LSC, and SSC data showed that *A. ilicifolius* and *A. leucostachyus* were more closely related to *A. mollis*. This also confirmed the evolutionary direction of *Acanthus* going from terrestrial to coastal intertidal habitats.

4. Discussion

Chloroplast genomes have been widely used in taxonomic and phylogenetic studies to assess interspecific evolutionary relationships and to scrutinize genomic structures, especially among closely related species (Moore et al., 2007, Jansen et al., 2007). Several studies have also shown that the analysis of the adaptive genetic evolution of chloroplast genomes is increasingly becoming a major part of the study of changes in gene function and gene structure (Nei and Kumar, 2000). The three *Acanthus* species grow in different ecological niches, ranging from arid terrestrial (*A. mollis*) to damp forest (*A. leucostachyus*), and to coastal intertidal habitats (*A. ilicifolius)*. Species with such a unique distribution may have evolved adaptive mechanisms to different light conditions. On this basis, we performed sequencing of three species of the *Acanthus* using Illumina sequencing technology. This is a new method that does not require extraction of chloroplast DNA, and it has been used in several chloroplast genomic studies, such as *B. prionitis* (Alzahrani et al., 2020), ten complete chloroplasts of Zingiberoideae (Li et al., 2021a), and five *Dicliptera* species (Huang et al., 2020). We assembled and annotated the chloroplast genomes by bioinformatics methods, which was the first comprehensive comparative chloroplast genomics and phylogenetic analysis of the genus *Acanthus*.

4.1. Chloroplast genomic characteristics and sequence variation in the three Acanthus species

The complete chloroplast genomes *A. ilicifolius*, *A. leucostachyus* and *A. mollis* exhibited typical quadripartite structures that are similar in composition and structure to other sequenced Acanthaceae chloroplast genomes (Gao et al., 2018, Zhuang and Tripp, 2017). The chloroplast genome lengths of *A. ilicifolius*, *A. leucostachyus* and *A. mollis* were 150,758, 154,686, and 150,339 bp, respectively. Among them, *A. leucostachyus* possessed the largest chloroplast length. Further analysis revealed that the length variation of the LSC region of chloroplast genomes of *Acanthus* was greater than that of the SSC and IRs regions, and it was primarily related to the difference of the LSC region, with that *A. leucostachyus* had the largest genome (Table 1). Previous studies on other species had also confirmed the existence of this phenomenon (Huang et al., 2020, Zhao et al., 2018, Meng et al., 2018). Moreover, (Alzahrani et al., 2020) also concluded that genome size in all the studied species was related to the variations in the LSC region.

The chloroplast genomes of most angiosperms encode 74 protein-coding genes, but there are also some genes, rearrangement, and loss

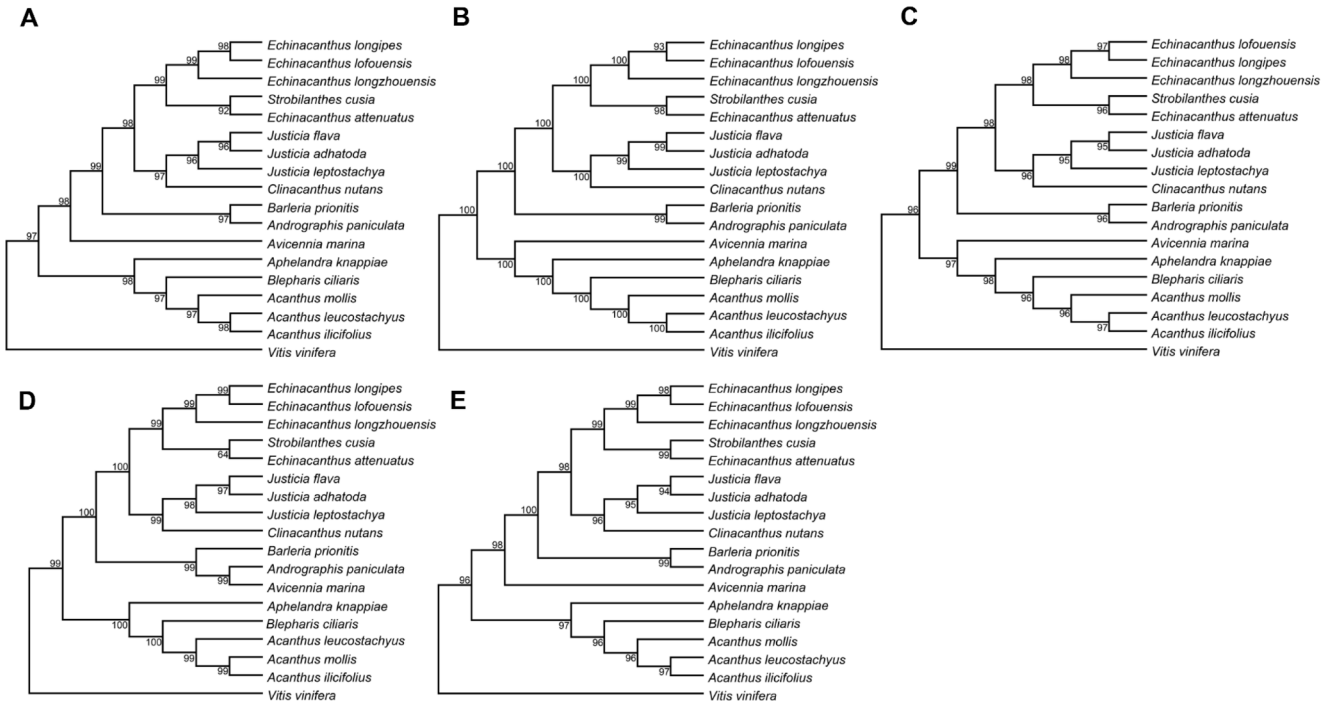

Fig. 8. Phylogenetic trees of the Acanthaceae species based on maximum likelihood (ML). **A,** Phylogenetic tree constructed using the complete chloroplast genome data. **B,** Phylogenetic tree constructed using coding region data. **C,** Phylogenetic tree constructed using LSC data. **D,** Phylogenetic tree constructed using SSC data. **E,** Phylogenetic tree constructed using IR data.

in different families and species (Kim et al., 2009; Millen et al., 2001). *A. ilicifolius* had the highest number of genes in the genus *Acanthus*. It lives in a coastal intertidal habitat with high salinity, hypoxia, and a strong UV (Xu et al., 2017). Therefore, it may have more copies of genes that allow it to better adapt and survive in such environments compared to its two terrestrial relatives. In addition, we found a large number of losses of the NDH (*ndhA*, *ndhF*, *ndhH*, *ndhI*, and *ndhK*) in the terrestrial plant *A. mollis* (Table 2). Similarly, *Cassytha* lost five NDH genes (*ndhB*, *ndhD*, *ndhE*, *ndhF*, and *ndhH*) in its plastomes (Wu et al., 2017). In the plastids of *Pedicularis*, a total of eleven NDH genes were pseudogenized or have been lost (Li et al., 2021b). In plants, the loss of NDH genes was known as a common but random thing (Bellot et al., 2016). In photosystem II of the chloroplast, NDH proteins played a role in adjusting the redox level of the cyclic electron transport mechanism, thus enabling the right operation of the photosynthesis (Peltier et al., 2016, Shikanai, 2016, Laughlin et al., 2019). The mechanisms underlying the loss of NDH genes are still unclear, but some studies suggest that some NDH loss of function may be replaced by the chloroplast/cytosol ATP exchange machinery (Suorsa et al., 2012, Strand et al., 2019). In summary, the loss of NDH genes in *A. mollis* may cause a carbon-dependent reduction in photosynthesis.

Expansion and contraction of the boundaries of the IRs/SC usually lead to size variations of the chloroplast genomes and play a crucial part in evolutionary processes (Henriquez et al., 2020). In this study, we conducted a deeper analysis of the expansion and contraction of the IRs/SC boundary and the variation hotspot differences in the three chloroplast genomes of *Acanthus*. We found that the boundaries between LSC/IRs regions were relatively conserved, and the distribution and location of diverse gene species were also highly consistent. Sequence comparison revealed that the SC region sequences of these chloroplast genomes had higher variance compared to the IRs region sequences, a phenomenon that is also seen in many other higher plants, which may be caused by the IRs region sequences being copy-corrected through gene conversion.

Meanwhile, we identified 10 genes with high variability in the genus *Acanthus*, including *psbK*, *ycf1*, *ndhG*, *rpl22*, *rps3*, *ndhF*, *matK*, *rps12*-CDS1, *rps14* and *atpF*-CDS2 genes, and 10 intergenic regions with high variability, including start-*psbA*, *ndhF-rpl32*, *psaA-ycf3*-CDS3, *psbK-psbI*, *atpA-atpF*-CDS2, *ccsA-ndhD*, *ndhI-ndhA*-CDS2, *atpH-atpI*, *psbT-psbN* and *rpoB-petN* genes. Among them, *ycf1* has been recognized as the core DNA barcode of plants (Douglas, 1998). *ndhF*, *matK*, and *ndhF-rpl32* region have been widely used in phylogenetic studies in other species (Yao et al., 2016, Song et al., 2015, Smith and Donoghue, 2008). It is evident that the LSC and SSC regions manifest a higher sequence variation than the IRs region, both in the protein-coding and intergenic regions (Fig. 6). This result is similar to the structure of most angiosperm chloroplast genomes (Wicke et al., 2011). The various regions we detected will provide an invaluable resource for species identification, breeding direction, and phylogenetic and population genetics studies of the genus *Acanthus*.

4.2. Molecular markers

SSRs, also known as microsatellites, are short repeats in the chloroplast genome that are inherited from the chloroplast genome of a single parent. They are highly polymorphic and therefore are often used as molecular markers for evolution researches such as identification of genetic diversity and species groups (Bryan et al., 1999, Provan, 2000, Ebert and Peakall, 2009). We identified 98, 138 and 95 SSRs in *A. ilicifolius*, *A. leucostachyus* and *A. mollis*, respectively. Among these SSRs, mononucleotide (A/T) was the predominant repeat type, which is consistent with the report that poly A and T are the most abundant repeats in the plant chloroplast genome (Dong et al., 2017, Ye et al., 2018, Yang et al., 2016). *A. leucostachyus* had the largest number of SSRs and was mainly located in the LSC region, which may explain the reason that its genome was the largest compared to *A. ilicifolius* and *A. mollis*. The SSRs identified in our study will be helpful for evolutionary studies, species identification and conservation of the genus *Acanthus*.

It has been reported that long repeat element is of great importance in chloroplast genomes recombination and sequence variation (Xue et al., 2019, Weng et al., 2014). In the present study, we found that *A. ilicifolius*, *A. leucostachyus* and *A. mollis* contained the same number of long repeats, but the types and length distribution of long repeats were uneven. Compared to other species of Acanthaceae, they contained a low number of repeats (Raubeson et al., 2007, Ding et al., 2016, Chen et al., 2018). These long repeats can be an important resource for studying differences in chloroplast genes.

4.3. Adaptive evolution in the three Acanthus species

Synonymous and nonsynonymous substitutions in the gene are important markers for studying the evolution of genetic adaptations in species. We identified nine genes (*atpF*, *rps11*, *rps16*, *rpl32*, *rpl33*, *ccsA*, *ycf4*, *psaJ* and *psbK*) under positive selection (Ka/Ks $>$ 1) in the *Acanthus*. These positively selected genes of the genus *Acanthus* probably played a critical role in adaptation to different environments. Among them, four genes encoding ribosome subunit proteins (*rps11*, *rps16*, *rpl32* and *rpl33*) are participated in the expression of chloroplast genes that are important for the biogenesis and function of chloroplasts (Lee et al., 2021). *psaJ* and *psbK* encode photosystem proteins, and the photosynthetic system is the main functional unit of the chloroplast organ. The gene *ccsA* is essential because the protein encoded by itself is a necessity for heme attachment to c-type cytochromes (Xie and Merchant, 1996). The *ycf4* gene was located in the LSC region, and its product was a thylakoid protein involved in regulating the assembly of the photosystem I complex (Wicke et al., 2011, Boudreau et al., 1997). Notably, the *atpF* gene showed a strong positive selection throughout terrestrial to marine evolution (*A. mollis-A. leucostachyus*: Ka/Ks = 1.57 and *A. leucostachyus-A. ilicifolius*: Ka/Ks = 1.32). *atpF* is an ATP subunit gene, which plays a crucial role in ATP synthase, and the plant's genetic system produces ATP synthase to help the plant body perform photosynthesis better (Westhoff et al., 1985). Thus, a positive selection of energy and photosynthesis-related genes in *A. ilicifolius* would benefit plant growth and development in high-salt and hypoxic environments. In general, our study implies that positive selection genes are likely to function as a driving force in the genus *Acanthus* for adaptation from land to coastal intertidal habitats.

4.4. Phylogenetic analyses in genus Acanthus

Chloroplast genomes can provide valuable phylogenetic analyses between plant genera as well as between species (Guo et al., 2020). *A. ilicifolius* is the major mangrove plant, and although mangroves evolved from terrestrial to marine, no studies have reported on the evolutionary direction of the *A. ilicifolius*, *A. leucostachyus* and *A. mollis*. Based on this, we reconstructed the phylogenetic tree of the Acanthaceae family using five sets of data (complete chloroplast genomes, CDS, LSC, SSC, and IRs). Phylogenetic analysis showed that the phylogenetic trees constructed from four datasets had the same topology with bootstrap values, except for the IRs region, which may be due to the more conserved sequences and the lowest differentiation in the IRs region. The phylogenetic tree showed that species of the same genus clustered on one branch, with Justicieae and Ruellieae strongly supported as monophyletic groups, and the sister relationship between Andrographidae and Barrierida, these results are consistent with previous reports (Alzahrani et al., 2020, Huang et al., 2020, McDade et al., 2008). Phylogenetic analyses suggested a relatively recent split between *A. ilicifolius* and *A. leucostachyus* after their ancestor divergence with *A. mollis*. In conclusion, this is the first time we reconstructed the phylogenetic relationships within the three *Acanthus* species using complete chloroplast genomes. However, the small number of samples within the genera did not allow deeper phylogenetic studies, but these

data provide a research foundation for the genome-scale phylogenetic of the *Acanthus*. With the advancement of sequencing technology, an increasing number of chloroplast genomes of *Acanthus* species will be resolved, which will help us understand deeper for phylogenetic relationships.

Author contributions

HZ designed and coordinated the entire project. DM and QD performed the genome assembly, annotation, data analysis and wrote the manuscript. XH collection of samples for sequencing and submitted data to the database. DM and ZZ participated in manuscript revision. All authors read and approved the final manuscript.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was financially supported by the Natural Science Foundation of China (NSFC) (32171740, 31870581, 31570586) and the National Key Research and Development Program of China (2017YFC0506102). We appreciate anonymous reviewers and editor for the insightful comments and valuable suggestions.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2023.147479.

References

Alzahrani, D.A., Yaradua, S.S., Albokhari, E.J., Abba, A., 2020. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. BMC Genom. 21, 393.

Ball, M.C.J.T., 1988. Ecophysiology of mangroves. 2, 129–142.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Beier, S., Thiel, T., Munch, T., Scholz, U., Mascher, M., 2017. MISA-web: a web server for microsatellite prediction. Bioinformatics 33, 2583–2585.

Bellot, S., Cusimano, N., Luo, S., Sun, G., Zarre, S., Gröger, A., Temsch, E., Renner, S.S., 2016. Assembled plastid and mitochondrial genomes, as well as nuclear genes, place the parasite family cynomoriaceae in the Saxifragales. Genome Biol. Evol. 8, 2214–2230.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., Pirovano, W., 2011. Scaffolding preassembled contigs using SSPACE. Bioinformatics 27, 578–579.

Boetzer, M., Pirovano, W., 2012. Toward almost closed genomes with GapFiller. Genome Biol. 13, R56.

Boudreau, E., Takahashi, Y., Lemieux, C., Turmel, M., Rochaix, J.D., 1997. The chloroplast ycf3 and ycf4 open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. Embo J. 16, 6095–6104.

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

Chen, H., Shao, J., Zhang, H., Jiang, M., Huang, L., Zhang, Z., Yang, D., He, M., Ronaghi, M., Luo, X., Sun, B., Wu, W., Liu, C., 2018. 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. 9, 324.

Daniel, T.F., 2003. *Familia Acanthaceae*.

Ding, P., Shao, Y., Li, Q., Gao, J., Zhang, R., Lai, X., Wang, D., Zhang, H., 2016. The complete chloroplast genome sequence of the medicinal plant *Andrographis paniculata*. Mitochondrial DNA A DNA Mapp Seq. Anal. 27, 2347–2348.

Dong, W., Xu, C., Li, W., Xie, X., Lu, Y., Liu, Y., Jin, X., Suo, Z., 2017. Phylogenetic resolution in Juglans based on complete chloroplast genomes and nuclear DNA sequences. Front. Plant. Sci. 8, 1148.

Douglas, S.E., 1998. Plastid evolution: origins, diversity, trends. Curr Opin Genet Dev 8, 655–661.

Ebert, D., Peakall, R., 2009. 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 9, 673–690.

Gao, C., Deng, Y., Wang, J., 2018. The complete chloroplast genomes of echinacanthus species (Acanthaceae): phylogenetic relationships, adaptive evolution, and screening of molecular markers. Front Plant Sci 9, 1989.

Gichira, A.W., Li, Z., Saina, J.K., Long, Z., Hu, G., Gituru, R.W., Wang, Q., Chen, J., 2017. The complete chloroplast genome sequence of an endemic monotypic genus Hagenia (Rosaceae): structural comparative analysis, gene content and microsatellite detection. PeerJ 5, e2846.

Guo, W., Banerjee, A.K., Ng, W.L., Yuan, Y., Li, W., Huang, Y., 2020. Chloroplast DNA phylogeography of the Holly mangrove *Acanthus ilicifolius* in the Indo-West Pacific. Hydrobiologia 847, 3591–3608.

Henriquez, C.L., Abdullah, I., Ahmed, M.M., Carlsen, A., Zuluaga, T.B., Croat & M. R. McKain, 2020. Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). Genomics 112, 2349–2360.

Huang, S., Ge, X., Cano, A., Salazar, B.G.M., Deng, Y., 2020. Comparative analysis of chloroplast genomes for five Dicliptera species (Acanthaceae): molecular structure, phylogenetic relationships, and adaptive evolution. PeerJ 8, e8450.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci U S A 104, 19369–19374.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kim, Y.K., Park, C.W., Kim, K.J., 2009. Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthis saniculifolia*, and its evolutionary implications. Mol Cells 27, 365–381.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33, 1870–1874.

Laslett, D., Canback, B., 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32, 11–16.

Laughlin, T.G., Bayne, A.N., Trempe, J.F., Savage, D.F., Davies, K.M., 2019. Structure of the complex I-like molecule NDH of oxygenic photosynthesis. Nature 566, 411–414.

Lee, K., Leister, D., Kleine, T., 2021. Arabidopsis mitochondrial transcription termination factor mTERF2 promotes splicing of group IIB introns. Cells 10.

Lei, W., Ni, D., Wang, Y., Shao, J., Wang, X., Yang, D., Wang, J., Chen, H., Liu, C., 2016. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus*. Sci Rep 6, 21669.

Li, D.M., Li, J., Wang, D.R., Xu, Y.C., Zhu, G.F., 2021a. Molecular evolution of chloroplast genomes in subfamily Zingiberoideae (Zingiberaceae). BMC Plant Biol 21, 558.

Li, X., Yang, J.-B., Wang, H., Song, Y., Corlett, R.T., Yao, X., Li, D.-Z., Yu, W.-B., 2021b. Plastid NDH pseudogenization and gene loss in a recently derived lineage from the largest hemiparasitic plant genus Pedicularis (Orobanchaceae). Plant Cell Physiol. 62, 971–984.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Lohse, M., Drechsel, O., Kahlau, S., Bock, R., 2013. OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucl. Acids Res. 41, W575–W581.

McDade, L.A., Daniel, T.F., Kiel, C.A., 2008. Toward a comprehensive understanding of phylogenetic relationships among lineages of Acanthaceae s.l. (Lamiales). Am. J. Bot. 95, 1136–1152.

Meng, X.X., Xian, Y.F., Xiang, L., Zhang, D., Shi, Y.H., Wu, M.L., Dong, G.Q., Ip, S.P., Lin, Z.X., Wu, L., Sun, W., 2018. Complete chloroplast genomes from Sanguisorba: identity and variation among four species. Molecules 23.

Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, N.T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Morden, C.W., Calie, P.J., Jermiin, L.S., Wolfe, K.H., 2001. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13, 645–658.

Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., Punta, M., 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucl. Acids Res. 41, e121–e.

Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci U S A 104, 19363–19368.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3, 418–426.

Nei, M., Kumar, S., 2000. Molecular evolution and phylogenetics. Oxford University Press.

Palmer, J.D., Shields, C.R., Cohen, D.B., Orton, T.J., 1983. Chloroplast DNA evolution and the origin of amphidiploid Brassica species. Theor. Appl. Genet. 65, 181–189.
Park, I., Kim, W.J., Yeo, S.M., Choi, G., Kang, Y.M., Piao, R., Moon, B.C., 2017. The complete chloroplast genome sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and comparative analysis with other Fritillaria species. Molecules 22.
Patel, R.K., Jain, M., 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7, e30619.
Peltier, G., Aro, E.M., Shikanai, T., 2016. NDH-1 and NDH-2 plastoquinone reductases in oxygenic photosynthesis. Annu Rev Plant Biol 67, 55–80.
Pfannschmidt, T., Nilsson, A., Allen, J.F., 1999. Photosynthetic control of chloroplast gene expression. Nature 397, 625–628.
Provan, J., 2000. Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana*. Mol Ecol 9, 2183–2185.
Qu, X.J., Moore, M.J., Li, D.Z., Yi, T.S., 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods 15, 50.
Raman, G., Park, S., 2016. The complete chloroplast genome sequence of Ampelopsis: gene organization, comparative analysis, and phylogenetic relationships to other Angiosperms. Front Plant Sci 7, 341.
Raubeson, L.A., Peery, R., Chumley, T.W., Dziubek, C., Fourcade, H.M., Boore, J.L., Jansen, R.K., 2007. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics 8, 174.
Shikanai, T., 2016. Chloroplast NDH: A different enzyme with a structure similar to that of respiratory NADH dehydrogenase. Biochim Biophys Acta 1857, 1015–1022.
Singh, D., Aeri, V., 2013. Phytochemical and pharmacological potential of *Acanthus ilicifolius*. J. Pharm. Bioallied Sci. 5, 17–20.
Smith, S.A., Donoghue, M.J., 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322, 86–89.
Song, Y., Dong, W., Liu, B., Xu, C., Yao, X., Gao, J., Corlett, R.T., 2015. Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. Front Plant Sci 6, 662.
Strand, D.D., D'Andrea, L., Bock, R., 2019. The plastid NAD(P)H dehydrogenase-like complex: structure, function and evolutionary dynamics. Biochem J 476, 2743–2756.
Sun, C., Chen, F., Teng, N., Xu, Y., Dai, Z., 2021. Comparative analysis of the complete chloroplast genome of seven Nymphaea species. Aquat. Bot. 170.
Suorsa, M., Järvi, S., Grieco, M., Nurmi, M., Pietrzykowska, M., Rantala, M., Kangasjärvi, S., Paakkarinen, V., Tikkanen, M., Jansson, S., Aro, E.M., 2012. PROTON GRADIENT REGULATION5 is essential for proper acclimation of Arabidopsis photosystem I to naturally and artificially fluctuating light conditions. Plant Cell 24, 2934–2948.
Weng, M.L., Blazier, J.C., Govindu, M., Jansen, R.K., 2014. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol 31, 645–659.
Westhoff, P., Alt, J., Nelson, N., Herrmann, R.G., 1985. Genes and transcripts for the ATP synthase CF0 subunits I and II from spinach thylakoid membranes. Mol Gen Genet 199, 290–299.
Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F., Quandt, D., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol 76, 273–297.
Wu, C.S., Wang, T.J., Wu, C.W., Wang, Y.N., Chaw, S.M., 2017. Plastome evolution in the sole hemiparasitic genus laurel dodder (Cassytha) and insights into the plastid phylogenomics of Lauraceae. Genome Biol Evol 9, 2604–2614.
Xia, X., 2018. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. Mol. Biol. Evol. 35, 1550–1552.
Xie, Z., Merchant, S., 1996. The plastid-encoded ccsA gene is required for heme attachment to chloroplast c-type cytochromes. J Biol Chem 271, 4632–4639.
Xu, S., He, Z., Zhang, Z., Guo, Z., Guo, W., Lyu, H., Li, J., Yang, M., Du, Z., Huang, Y., Zhou, R., Zhong, C., Boufford, D.E., Lerdau, M., Wu, C.I., Duke, N.C., Shi, S., 2017. The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. Natl Sci Rev 4, 721–734.
Xue, S., T. Shi, W. Luo, X. Ni, S. Iqbal, Z. Ni, X. Huang, D. Yao, Z. Shen & Z. Gao (2019) Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. *Hortic Res,* 6, 89.
Yang, L., Feng, C., Cai, M.-M., Chen, J.-H., Ding, P., 2020. Complete chloroplast genome sequence of *Amomum villosum* and comparative analysis with other Zingiberaceae plants. Chinese Herbal Medicines 12, 375–383.
Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., Zhao, G., 2016. Comparative analysis of the complete chloroplast genomes of five Quercus species. Front Plant Sci 7, 959.
Yao, X., Tan, Y.H., Liu, Y.Y., Song, Y., Yang, J.B., Corlett, R.T., 2016. Chloroplast genome structure in Ilex (Aquifoliaceae). Sci Rep 6, 28559.
Ye, W.Q., Yap, Z.Y., Li, P., Comes, H.P., Qiu, Y.X., 2018. Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). Mol Phylogenet Evol 127, 978–987.
Zhao, Z., Wang, X., Yu, Y., Yuan, S., Jiang, D., Zhang, Y., Zhang, T., Zhong, W., Yuan, Q., Huang, L., 2018. Complete chloroplast genome sequences of Dioscorea: characterization, genomic resources, and phylogenetic analyses. PeerJ 6, e6032.
Zhuang, Y., Tripp, E.A., 2017. The draft genome of Ruellia speciosa (Beautiful Wild Petunia: Acanthaceae). DNA Res. 24, 179–192.