

RESEARCH

Open Access



Assembly and comparative analysis of the complete mitochondrial genome of *Echinacanthus longipes* (Acanthaceae), endemic to the Sino-Vietnamese karst flora

Chunming Gao¹, Shu Wang^{2,3,4}, Yusong Huang^{2,3,4,5} and Yunfei Deng^{2,3,6*}

Abstract

Background *Echinacanthus longipes* is an endemic species in the Sino-Vietnamese karst flora in the family Acanthaceae. It displays distinctive environmental adaptation characteristics in karst regions. Although it provides an important model for understanding the role of limestone karst in speciation and endemism, the mitochondrial genome (mtDNA) of *E. longipes* has not been fully characterized.

Results Here, the mtDNA of *E. longipes* was successfully assembled as a complex structure in the form of two small circular and three linear molecules with a total length of 810,200 bp. The annotated results revealed 36 protein-coding genes (PCGs), 22 tRNA genes, and three rRNA genes in this mtDNA. Notably, substantial sequence repeats and more tRNAs translocations from the chloroplast to the mtDNA were identified. Among the PCGs of *E. longipes*, the majority of 401 RNA editing sites were involved in amino acid transitions to hydrophobic sites. The current phylogenetic analysis based on PCGs revealed the evolution of Lamiales and a close relationship between *E. longipes* and *Avicennia marina*. However, comparative analyses, including size, structure, GC contents, and genes, reflected the variation in the mitogenomes within Acanthaceae, and the collinearity analysis confirmed the low level of conservation in the genomes of related species in Lamiales. Moreover, the Ka/Ks analysis revealed that negative selection occurred on most PCGs, with the notable exception of *ccmB*, which underwent positive selection. Interestingly, the *ccmB* gene had the most protein editing sites.

Conclusions This study will be invaluable for the mitochondrial study of Acanthaceae. It also provides extensive information for functional genetic and adaptive studies of *Echinacanthus* in karst regions in the future.

Keywords *Echinacanthus longipes*, Acanthaceae, Mitochondrial genome, Sino-Vietnamese karst flora

*Correspondence:

Yunfei Deng
yfdeng@scib.ac.cn

¹College of Biological and Environmental Engineering, Shandong University of Aeronautics, Binzhou 256600, China

²State Key Laboratory of Plant Diversity and Specialty Crops, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong 510650, China

³Key Laboratory of National Forestry and Grassland Administration on Plant Conservation and Utilization in Southern China, Guangzhou 510650, China

⁴University of the Chinese Academy of Sciences, Beijing 100049, China

⁵Guangxi Key Laboratory of Functional Phytochemicals Research and Utilization, Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and the Chinese Academy of Sciences, Guilin, Guangxi 541006, China

⁶Southeast Asia Bioaffiliation Research Institute, Chinese Academy of Sciences, Yezin, Nay Pyi Taw 05282, Myanmar



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Acanthaceae comprise approximately 191 genera and approximately 4900 species distributed in tropical and subtropical regions [1, 2]. *Echinacanthus* Nees is a small genus characterized by its axillary or terminal thyrse inflorescence and the anthers with spurred thecae. It consists of four species, *E. attenuatus* Nees, *E. longipes* H. S. Lo et D. Fang, *E. longzhouensis* H. S. Lo, and *E. lofouensis* (H. Lév.) J. R. I. Wood [3, 4]. Amongst, *E. attenuatus* is restricted to Bhutan, India, and Nepal in the western Himalayas, and, the other three species are endemic to southern China and northern Vietnam in the Sino-Vietnamese karst flora [5, 6, 7]. Among them, *E. longipes* is distributed in western Guangxi and southern Yunnan of China, and northern Vietnam, and moreover, it is acaulescent rosette or caulescent herb with purple corollas growing in the damp spaces of limestone hills (Fig. 1). Additionally, it strongly adapts to shallow karst soil with high pH, lower water storage capacity, and high concentrations of magnesium (Mg) and calcium (Ca) [8]. Therefore, *E. longipes* is a typical limestone species and displays distinctive environmental adaptation characteristics in karst regions. Although it provides an important model for understanding the role of limestone karst in speciation and endemism, the mitochondrial genome of *E. longipes* remains poorly characterized.

Mitochondria, as cellular organelles within eukaryotic cells, play important roles in energy provision and diverse physiological activities [9, 10]. Using Margulis's

endosymbiosis theory, mitochondria originated from archaea when eukaryotes engulfed bacteria and then evolved into organelles with independent genomes [11, 12]. Unlike nuclear genomes, which are derived from biparental contributions, mitochondrial genomes exhibit an exclusive maternal inheritance pattern [13]. With the development of high-throughput sequencing technologies, the complexity and the variability of plant mitochondrial genomes were revealed [14]. Furthermore, it has been widely utilized for reconstructing phylogenetic relationships and studies of evolutionary biology [15]. Although Acanthaceae is a large family consisting of 4900 species, the complete mtDNA has been sequenced and submitted to National Center for Biotechnology Information (NCBI) for only one species, *Avicennia marina* (Forssk.) Vierh. (PP908999.1). Thus, the mitochondrial genomes of species within Acanthaceae need to be further studied.

In the present study, the initial complete mtDNA of *E. longipes* was reported. This is the first comprehensive study of the mtDNA of an *Echinacanthus* species. The results contribute to the growth of a mitochondrial DNA database specific to the family Acanthaceae, providing crucial genetic data for *Echinacanthus* species and identifying the genes underlying positive selection in *E. longipes*.

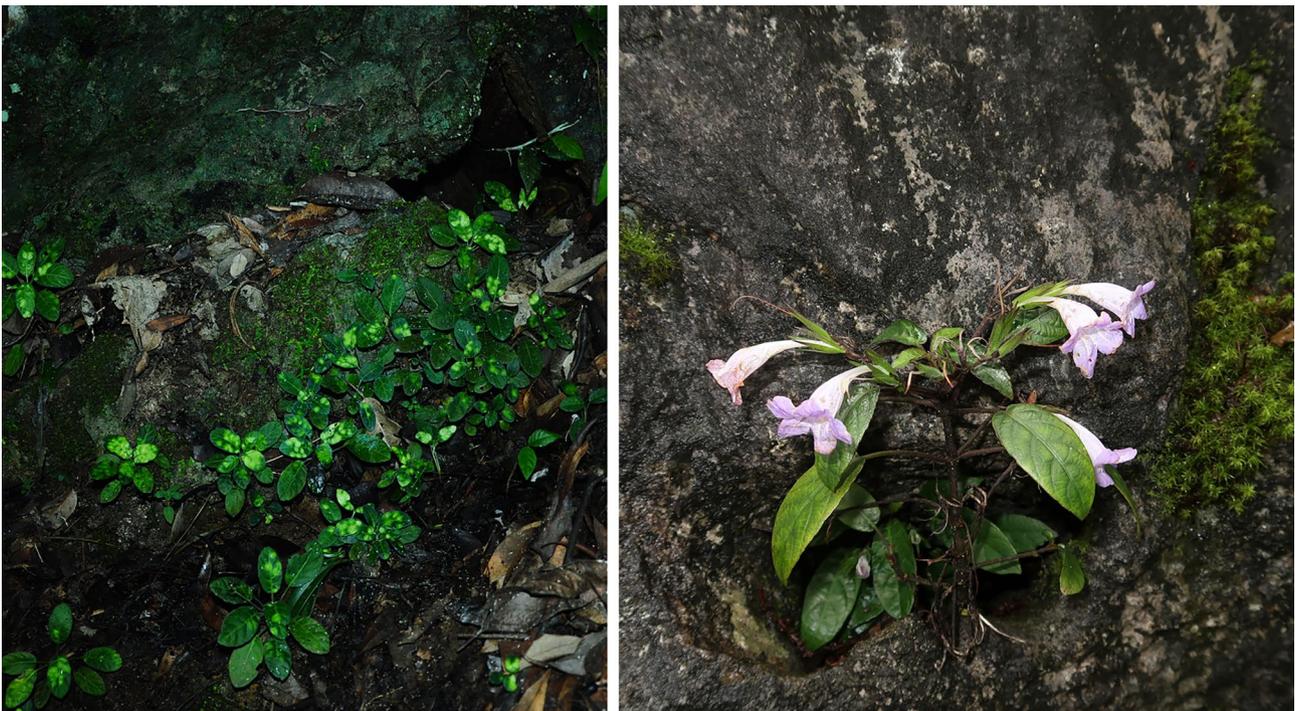


Fig. 1 The plant of *Echinacanthus longipes* and its habitat

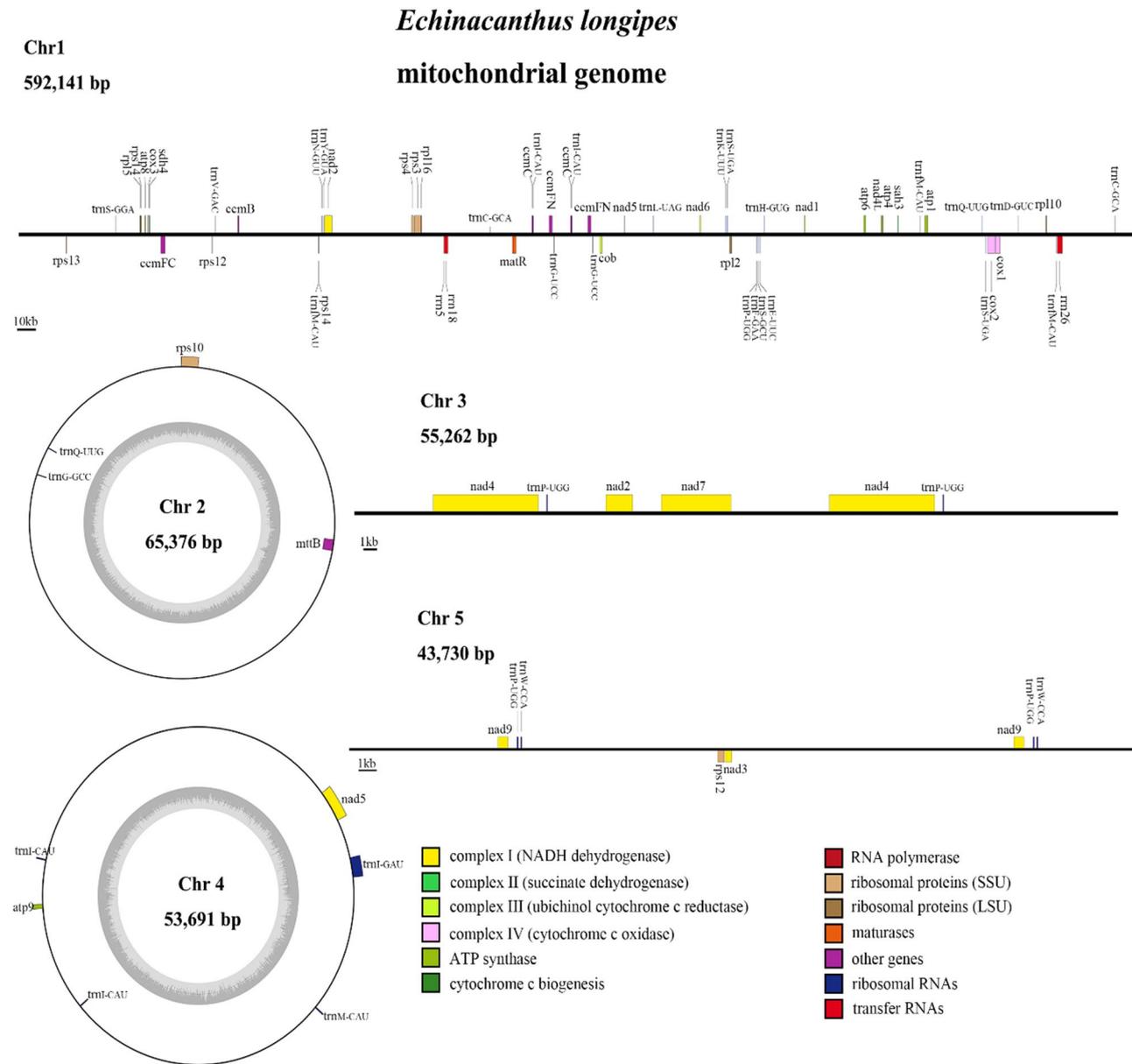


Fig. 2 The maps of *Echinacanthus longipes* mitochondrial genome

Table 1 The information of *E. longipes* mitochondrial genome

Mitogenome	Lengths (bp)	GC contents (%)	Structure	Accession number (NCBI)
Chromosome 1	592,141	42.3	linear	PQ164709
Chromosome 2	65,376	43.0	circular	PQ164710
Chromosome 3	55,262	44.4	linear	PQ164711
Chromosome 4	53,691	42.6	circular	PQ164712
Chromosome 5	43,730	40.3	linear	PQ164713

Results

Characteristics of the mtDNA of *E. longipes*

Here, the entire mtDNA of *E. longipes* was obtained. It was assembled as a complex structure in the form of small circular and linear molecules (Fig. 2). Notably, it consists of five contigs named chromosome 1–5 with descending lengths as follows: 592,141 bp, 65,376 bp, 55,262 bp, 53,691 bp and 43,730 bp (Table 1). In total, the length of the mtDNA of *E. longipes* is 810,200 bp, and the GC content is 42.4%. Sixty-one unique genes are annotated in the mtDNA, including 22 tRNA genes (*trnG-UCC*, *trnS-UGA*, *trnC-GCA*, *trnQ-UUG* and *trnW-CCA* with two copies, *trnM-CAU* with three copies, *trnI-CAU* with four copies, and *trnP-UGG* with five

Table 2 The mitochondrial genome encoding genes of *E. longipes*

Group of genes	Name of genes
ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i>
NADH dehydrogenase	<i>nad1*, nad2*, nad3, nad4*(x2), nad4L, nad5*, nad6, nad7*, nad9 (x2)</i>
Cytochrome b	<i>cob</i>
Cytochrome c biogenesis	<i>ccmB, ccmC (x2), ccmFc*, ccmFn (x2)</i>
Cytochrome c oxidase	<i>cox1*, cox2*, cox3</i>
Maturases	<i>matR</i>
Transport membrane protein	<i>mttB</i>
Large subunit of ribosome	<i>rpl2, rpl5, rpl10, rpl16</i>
Small subunit of ribosome	<i>rps3*, rps4, rps10*, rps12 (x2), rps13, rps14 (x2)</i>
Succinate dehydrogenase	<i>sdh3, sdh4</i>
Ribosome RNA	<i>rrn5, rrn18, rrn26</i>
Transfer RNAs	<i>trnG-UCC (x2), trnI-GAU, trnP-UGG(x5), trnS-UGA(x2), trnC-GCA(x2), trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnH-GUG, trnK-UUU, trnL-UAG, trnI-CAU(x4), trnM-CAU(x3), trnM-CAU, trnN-GUU, trnQ-UUG(x2), trnS-GCU, trnS-GGA, trnV-GAC, trnW-CCA(x2), trnY-GUA</i>

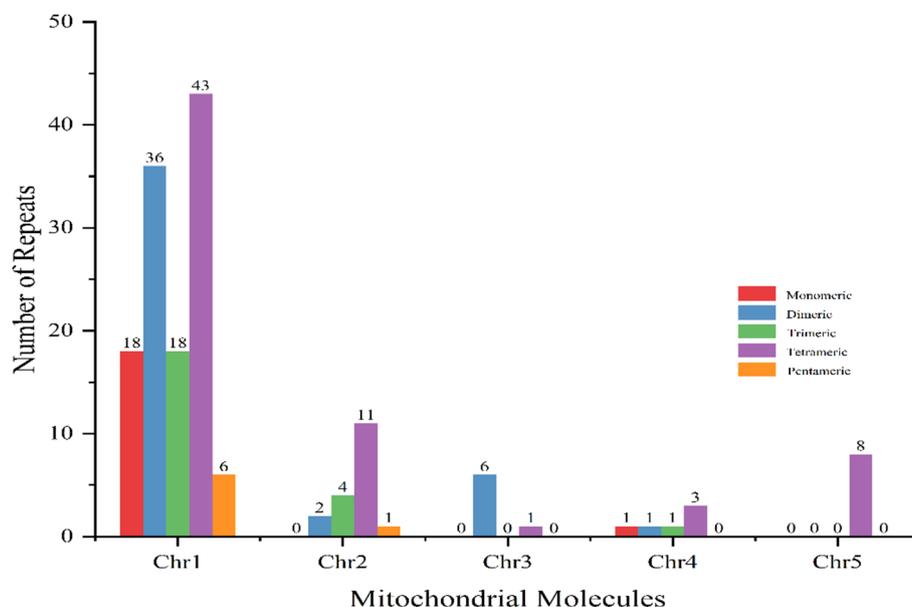
Notes: the value in parentheses indicates the number of the gene's copies. * Labeled the genes that contain introns

copies), 36 protein-coding genes (*nad4, nad9, ccmC, ccmFn, rps12, and rps14* with two copies), and three rRNA genes (Table 2). Annotation of the mtDNA of *E. longipes* reveals 24 mitochondrial core genes and 12 non-core genes. The core genes include nine NADH dehydrogenase genes (*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, and nad9*), five ATP synthase genes (*atp1, atp4, atp6, atp8, and atp9*), four cytochrome C biogenesis genes (*ccmB, ccmC, ccmFc, and ccmFn*), three

cytochrome C oxidase genes (*cox1, cox2, and cox3*), one transport membrane protein-encoding gene (*mttB*), one maturase gene (*matR*), and one apocytochrome b gene (*cob*). The noncore genes include six ribosomal small subunit genes (*rps3, rps4, rps10, rps12, rps13, and rps14*), four ribosomal large subunit genes (*rpl2, rpl5, rpl10, and rpl16*), and two succinate dehydrogenase genes (*sdh3 and sdh4*). Additionally, ten intron-containing genes (*nad1, nad2, nad4, nad5, nad7, ccmFc, cox1, cox2, rps3, and rps10*) are identified in the mitogenome of *E. longipes*.

Repeat sequence analysis

A total of 160 simple sequence repeat (SSR) loci, including 19 monomeric simple sequence repeats (SSRs) (11.88%), 45 dimeric SSRs (28.13%), 23 trimeric SSRs (14.38%), 66 tetrameric SSRs (41.25%), and seven pentameric SSRs (4.38%), were discovered in the mtDNA of *E. longipes*. However, no hexameric SSRs were found in this study. As a result, the tetrameric and pentameric SSRs contributed the most and the least, respectively. The total SSR loci were spread across the different chromosomes as follows: 121 were distributed on chromosome 1, 18 were distributed on chromosome 2, seven were distributed on chromosome 3, six were distributed on chromosome 4, and eight were distributed on chromosome 5 (Fig. 3). Tetrameric SSRs, which were the most abundant, were distributed on all chromosomes, with 43 on chromosome 1, 11 on chromosome 2, three on chromosome 4, and eight on chromosome 5. The next most abundant type was dimeric SSRs, which were distributed mainly on chromosome 1 (36). Additionally, as shown in Fig. 3 23 trimeric SSRs were identified on chromosome 1 (18), chromosome 2 (four),

**Fig. 3** The simple sequence repeats in *E. longipes* mitochondrial genome

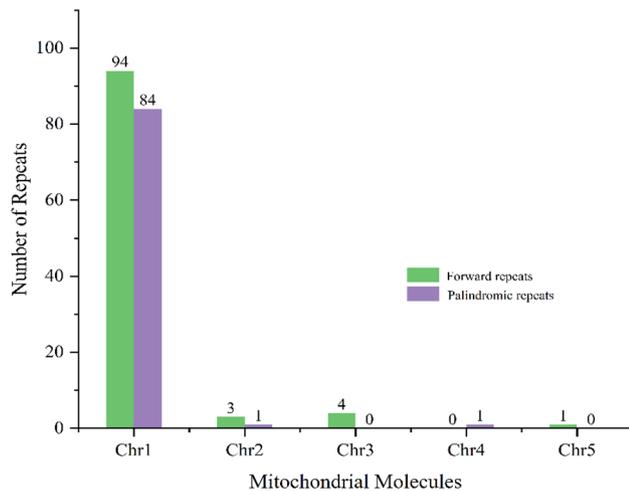


Fig. 4 The dispersed repeats in *E. longipes* mitochondrial genome

Table 3 Distribution of the tandem repeats in *E. longipes* mitochondrial genome

No.	Repeat sequence	Size	Copy	Percent matches	Location
1	CAAAGTTATATAAG	16	2	100	Chr1:169140–169,171
2	TATATAGACT	12	2	100	Chr1:359501–359,526

and chromosome 4 (one). When monomeric SSRs were analyzed, almost all were distributed on chromosome 1. Seven pentameric SSRs were split on chromosomes 1 and 2. Moreover, of the detected SSR regions, most of the SSRs (132) were identified in intergenic spaces, whereas 28 SSRs were located in the *nad1*, *nad2*, *nad4*, *ccmFC*, *cob*, *rps3*, *rps10*, and *rrn18* genetic spaces. Interestingly, the majority of the SSRs were especially rich in A or T bases (Supplementary Table S1).

Mitochondrial genome of *E. longipes* also contained numerous dispersed repeats, consisting of 102 pairs of forward repeats and 86 pairs of palindromic repeats (Supplementary Table S2). However, neither reverse repeats nor complement repeats were identified. The 40s bp repeats were the most abundant for both types, accounting for more than half of them (119). Moreover, the longest forward repeats and palindromic repeats were all found on chromosome 1, reaching 20,708 and 8,116 bp. The greatest number of dispersed repeats was detected on chromosome 1, with 94 pairs of forward repeats and 84 pairs of palindromic repeats (Fig. 4).

A total of two tandem repeats with matching degrees greater than 95% and lengths spanning 12–16 bp (Table 3) were present in the mtDNA of *E. longipes*. In addition, they were all detected on chromosome 1.

Protein-coding genes codon usage analysis

The length of the PCGs in *E. longipes* was 31,969 bp. For the majority of these genes, ATG was the typical start codon, with the exception of GTG for the *rpl16* gene and ACG for the *nad4L* gene. TGA, TAA, and TAG were detected as termination codons. In the complete mtDNA of *Echinacanthus*, 61 different types of amino acid codons encoding all 20 amino acids were identified. Among all 61 codons, excluding methionine (AUG) and tryptophan (UGG), which exhibited no codon preference (relative synonymous codon usage (RSCU) = 1), 29 codons were used more frequently than expected (RSCU > 1), and 30 codons had RSCU values less than 1 (Supplementary Table S3). The amino acids with the highest frequencies were serine (Ser), arginine (Arg), and leucine (Leu), whereas tryptophan (Trp) and methionine (Met) had the lowest frequencies (Fig. 5).

Prediction of RNA editing sites

A total of 401 potential RNA editing sites were identified across the 34 PCGs from *E. longipes* mtDNA, primarily involving the conversion of C to T (Supplementary Table S4). Notably, *ccmB* presented the highest frequency of RNA editing sites, totaling 36 edits. Both the *mttB* and *nad4* genes each presented 32 RNA editing sites. In addition, the *nad2* gene was associated with 24 RNA editing events. However, *atp1*, *atp8*, and *atp9* each had only a single RNA editing site (Fig. 6). Furthermore, 68.33% (274) of these RNA editing sites were located at the second site, whereas 28.68% (115) occurred at the first position of the triplet codes. Twelve particular cases were edited at the first and second sites simultaneously, which led to an amino acid change from proline (CCC or CCT) to phenylalanine (TTC or TTT). As a result of all editing, 194 (48.38%) amino acids underwent a transition from the hydrophilic to the hydrophobic state. Meanwhile, 30 (7.48%) amino acids were anticipated to shift from the hydrophobic to the hydrophilic state. It was also found that two hydrophilic amino acids, arginine and glutamine, were converted to stop codons (Table 4).

Homology analysis of genome sequences

The complete chloroplast genome (cpDNA) of *E. longipes* was 152,644 bp in length, encompassing 113 unique genes [7]. The migration of sequences between the chloroplast and the mitochondria in *E. longipes* were analyzed. Cumulatively, 72 sequences, varying in length from 71 to 10,303 bp, were identified between the mtDNA and cpDNA (Supplementary Table S5). The total length of these migrant sequences was 118,235 bp, constituting 14.59% of the overall mtDNA. Furthermore, the most frequent migration occurred between chromosome 1 and cpDNA. However, only four sequences migrated between chromosome 2 and cpDNA (Fig. 7). Additionally,

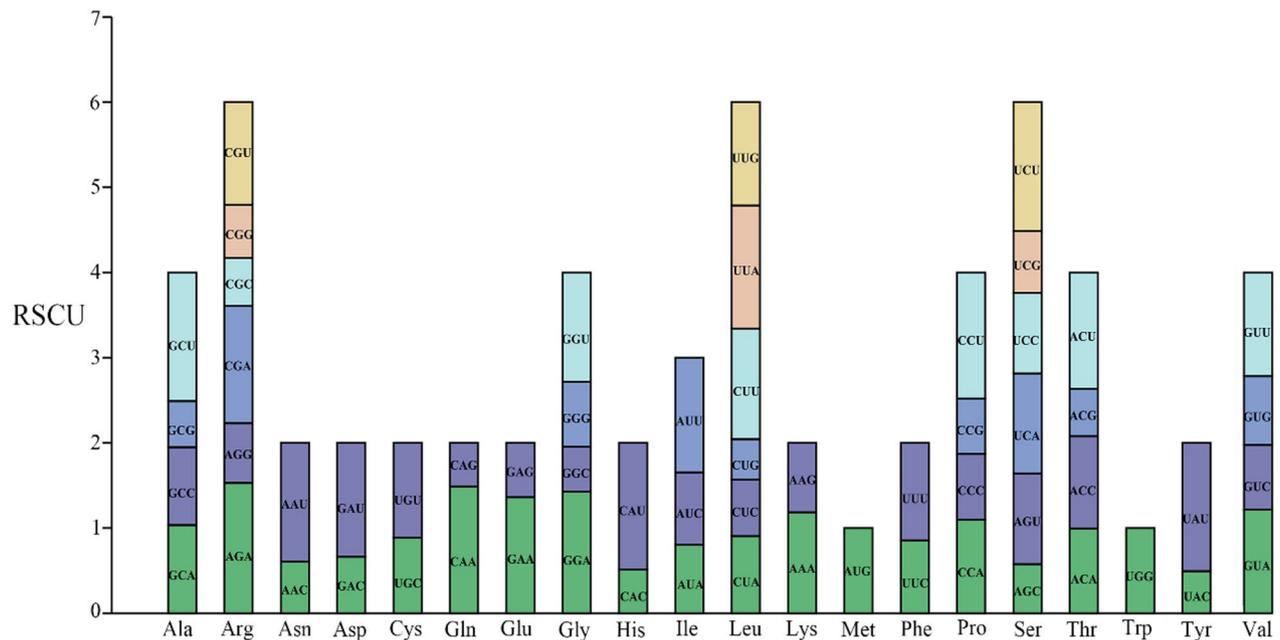


Fig. 5 The relative synonymous codons usage of *E. longipes* mitochondrial genome. The x axis shows the amino acid, and the bars represent the value of each amino acid codons

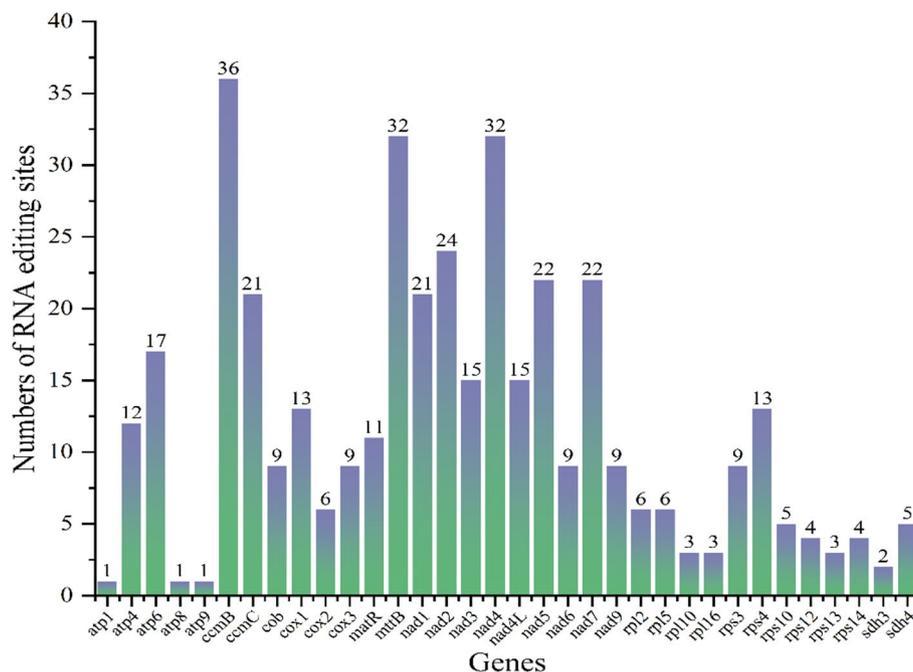


Fig. 6 The distribution of RNA editing sites of the protein-coding genes in *E. longipes* mitochondrial genome

annotation of these sequences in mtDNA resulted in the identification of 13 tRNAs, namely, *trnD*-GUC, *trnH*-GUG, *trnI*-GAU, *trnL*-UAG, *trnM*-CAU, *trnJ*-CAU, *trnN*-GUU, *trnP*-UGG, *trnQ*-UUG, *trnW*-CCA, *trnS*-GGA, *trnS*-UGA and *trnV*-GAC, two PCGs including *rps14* and *rps12*, and one rRNA *rrn18* (Supplementary Table S5).

Phylogenetic and Ka/Ks analysis

Thirty-one conserved PCGs, including *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *cox1*, *cox2*, *cox3*, *ccmC*, *ccmB*, *ccmFC*, *ccmFN*, *matR*, *mttB*, *rps3*, *rps4*, *rps12*, *rps13*, *rps14*, *rpl5*, *rpl10*, and *sdh4*, which were retrieved from 25 Lamiales species and two outgroup species, were used to construct

Table 4 Prediction of RNA editing sites in *E. longipes*

Type	RNA-editing	Position	Number	
Hydrophobic-hydrophobic	CCA (P) => CTA (L)	2	41	123
	CCG (P) => CTG (L)	2	27	(30.67%)
	CCT (P) => CTT (L)	2	15	
	CCC (P) => TTC (F)	1+2	6	
	CTT (L) => TTT (F)	1	6	
	CTC (L) => TTC (F)	1	6	
	GCG (A) => GTG (V)	2	5	
	GCA (A) => GTA (V)	2	2	
	CCC (P) => CTC (L)	2	5	
	CCT (P) => TTT (F)	1+2	6	
	GCT (A) => GTT (V)	2	3	
	GCC (A) => GTC (V)	2	1	
	Hydrophilic-hydrophilic	CGT (R) => TGT (C)	1	23
CAT (H) => TAT (Y)		1	17	(12.97%)
CGC (R) => TGC (C)		1	5	
CAC (H) => TAC (Y)		1	7	
Hydrophobic-hydrophilic	CCT (P) => TCT (S)	1	17	30
	CCG (P) => TCG (S)	1	5	(7.48%)
	CCC (P) => TCC (S)	1	4	
	CCA(P) => TCA (S)	1	4	
Hydrophilic-hydrophobic	TCA(S) => TTA (L)	2	63	194
	TCT(S) => TTT (F)	2	42	(48.38%)
	TCG(S) => TTG(L)	2	34	
	TCC(S) => TTC(F)	2	25	
	CGG(R) => TGG (W)	1	19	
	ACT(T) => ATT (I)	2	4	
	ACC(T) => ATC (I)	2	3	
	ACG(T) => ATG (M)	2	2	
	ACA(T) => ATA (I)	2	2	
Hydrophilic-stop	CGA(R) => TGA (X)	1	1	2
	CAA (Q) => TAA (X)	1	1	(0.50%)

a phylogenetic tree. As shown in Fig. 8A, all the sampled species from the eight families in the order Lamiales were clustered into one clade with 1.00 posterior probability (PP). The results of this study show that, within Lamiales, Lamiaceae and Orobanchaceae are formed a clade, which was identified as a sister to Lentibulariaceae with a high support value (PP = 1.00). Two species of Acanthaceae, *E. longipes* and *A. marina*, are clustered as a group with 1.00 posterior probability value. Therefore, *E. longipes* was determined to be closely related to *A. marina* in the present study. Acanthaceae are formed a sister group to the branch containing Lamiaceae, Orobanchaceae, Lentibulariaceae and Bignoniaceae, and the analyses obtained high support for the node (PP = 1.00). At the same time, the present phylogenetic tree is favored Plantaginaceae and the branch of the above families including Lamiaceae, Orobanchaceae, Lentibulariaceae, Bignoniaceae and Acanthaceae as a sister group. Additionally, Oleaceae is the earliest diverging lineage in Lamiales in this study.

The rates of nonsynonymous substitutions (Ka) and synonymous substitutions (Ks) were calculated to

evaluate whether selective pressure existed on PCGs during evolution within Lamiales (Supplementary Table S6). When the Ka/Ks ratio is > 1, genes are undergoing positive selection, conversely, Ka/Ks < 1 denotes negative selection, and Ka/Ks = 1 represents neutral selection. Thus, 16 PCGs from *E. longipes* mtDNA were compared with those from the selected species in Lamiales. The Ka/Ks values of most PCGs in *E. longipes* were determined to be less than 1. The genes with Ka/Ks values higher than 1 included *ccmB*, *ccmC*, *cox3*, *matR* and so on. However, *ccmB* was the only gene with Ka/Ks > 1 compared with all the other species (Fig. 8B).

Comparison of the mtDNA of *E. longipes* and other species

Firstly, the comparative mtDNA analysis will be performed among species within Acanthaceae. To date, the available mitochondrial information in Acanthaceae is limited to *A. marina*. So, the comparisons will be focus on *E. longipes* and *A. marina*. As shown in Table 5, the differences between *A. marina* and *E. longipes* are reflected mainly in the size, structures, GC contents and

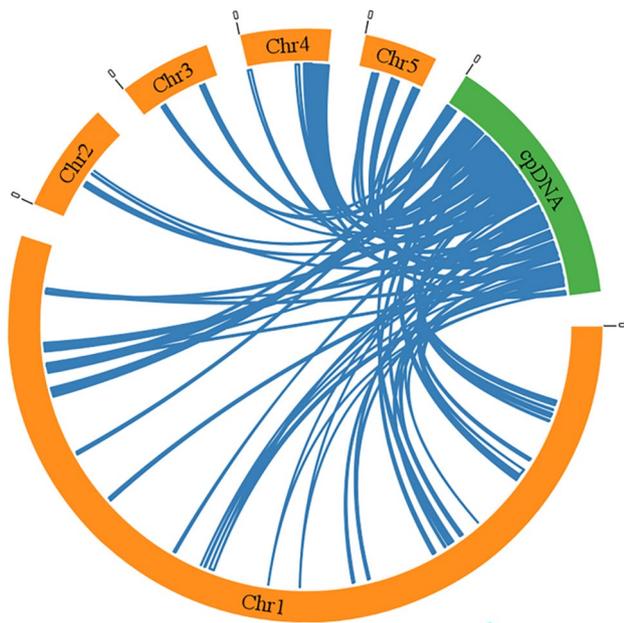


Fig. 7 Homology analysis between mitochondrial genome and chloroplast genome of *E. longipes*. The orange and the green arcs denote the mtDNA and cpDNA, respectively. The blue lines signify the homologous sequences

genes. Combination the data of *E. longipes* in this study and the data of *A. marina* in NCBI (<https://www.ncbi.nlm.nih.gov/nucleotide/PP908999.1?report=genbank>), a deeply comparison of these mitochondrial genomes was conducted. The mtDNA of *A. marina* is typically single circular, in contrast, *E. longipes* possesses complex structures with three linear and two circular chromosomes. The gene annotations of these two genomes were compared with each other. Notably, the mtDNA of *E. longipes* lost *rps7*, but it had more tRNA genetic duplications which may lead to their differences in length.

In order to illustrate the homologous of mitochondrial genomes between *E. longipes* and the closely related species in Lamiales, *E. longipes* was further compared with *A. marina* (Acanthaceae), *Plantago ovata* (Plantaginaceae), *Markhamia cauda-felina* (Bignoniaceae) and *Utricularia reniformis* (Lentibulariaceae) using the Blastn program to analyze the homologous collinear blocks, excluding blocks less than 0.5 kb in length. As a result, some fine homologous regions and numerous inversion regions were identified between *E. longipes* and the related species. However, the arrangement order of collinear blocks was inconsistent across these species. Additionally, compared with other species, unique regions with substantial genomic rearrangements were identified in the mtDNA of *E. longipes* (Fig. 9).

Discussion

Mitochondria are the “energy factories” within eukaryotic cells. In addition, they also play critical roles in physiological activities such as information transmission, division, cell differentiation, and apoptosis [9, 10]. In 1992, the first mitochondrial genome from terrestrial plants was reported [16]. An increasing number of plants mitochondrial genome have since been sequenced and reported. Acanthaceae are a large family of angiosperms comprising approximately 191 genera and 4900 species [2]. However, until now, only one mtDNA within Acanthaceae has been reported (submitted to NCBI). In the present study, the complete mtDNA of *E. longipes*, an endemic species in the Sino-Vietnamese karst region, was sequenced and assembled. It was the first mtDNA sequenced from the genus *Echinacanthus*. According to previous studies, the length of angiosperm mitochondria genomes usually ranges from 221 kb to 11.3 mb [10]. The mtDNA of *E. longipes* spans 810,200 bp, with 61 genes encompassing 36 PCGs, 22 tRNA genes and three rRNA genes. In addition to their typical circular shape, the mitochondrial genomes of terrestrial plants present complex structures, featuring linear and branched molecules. In contrast to previous studies, in which the predominant structures of plant mitochondrial genomes including the single circular (e.g., *Suaeda glauca* and *Rothea serrata*), multiple circulars (e.g., *Angelica biserrate* and *Ajuga ciliata*), linear (e.g., *Primulina hunanensis*) and branched (e.g., *Ventilago leiocarpa* and *Quercus acutissima*) configurations [17–22], the *E. longipes* mtDNA exhibits a complex structure consisting of three linear and two circular chromosomes.

Repeat sequences are considered to play a vital role in shaping the mitochondrial genomes because of inter-molecular recombination [17]. They are widely dispersed in the mitochondrial genome and have been extensively used in biological evolution and population genetics studies [23]. In the present study, three types of repeat sequences, SSRs, dispersed repeats and tandem repeats, were identified across the mtDNA of *E. longipes*. The majority of SSRs were tetrameric repeats. These tetrameric repeats could serve as candidate molecular markers in *Echinacanthus* for future research on species identification, speciation mechanisms, and interspecific polymorphisms. Moreover, the SSRs in *E. longipes* were mainly located in intergenic spacer regions and contributed to A/T richness. Consistent with observations in other angiosperms such as *Primulina hunanensis*, *Gleditsia sinensis* and *Acer truncatum*, the elevated AT-content in SSR regions further supports the genome-wide high AT composition observed in plant mitochondrial genomes [18, 24, 25]. Notably, compared with two tandem repeats, dispersed repeats including forward and palindromic repeats, accounted for 188 in the *E. longipes*

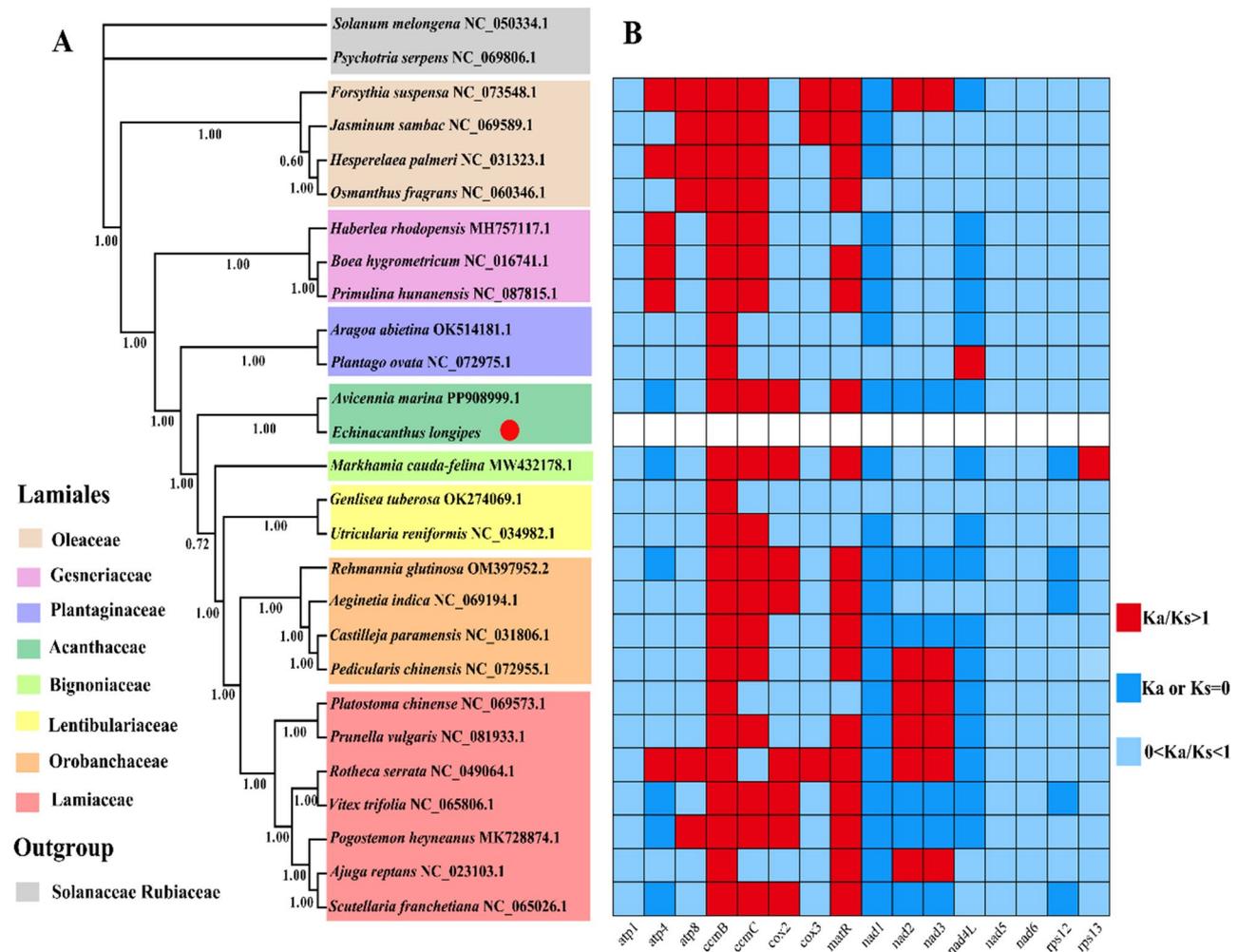


Fig. 8 Phylogeny and Ka/Ks analysis of eight families in Lamiales. **(A)** Phylogenetic tree of eight families in Lamiales based on 31 PCGs. The numbers of the branches represent posterior probability (PP). **(B)** The Ka/Ks values of 16 protein-coding genes of *E. longipes* versus 24 species of Lamiales

Table 5 General features of the two mitochondrial genomes in Acanthaceae

Characters	<i>Avicennia marina</i>	<i>Echinacanthus longipes</i>
Size	574,037 bp	810,200 bp
Structure	circular	two circular, three linear
GC content	44.70%	42.40%
Gene numbers	62	61
Protein coding	37	36
rRNA	3	3
tRNA	22	22

mtDNA. These highly repetitive sequences may not only increase the frequency of intermolecular recombination but also change the conformations and sizes of plant mitochondrial genomes [17, 18, 26–29].

Thirty-six PCGs with a total length of 31,969 bp were aligned in the *E. longipes* mtDNA. The analysis of codon usage in *E. longipes* presented herein indicates that mtDNA predominantly utilizes ATG as the initiation codon. Additionally, GTG serves as translation initiation

codon for the *rpl16* gene in plant mtDNA [30]. Furthermore, RNA editing, an RNA nucleotide posttranscriptional modification process, widely occurs in the PCGS of mitochondrial genomes and significantly affects gene expression [17, 31]. Previous reports have shown that the number of editing sites and the genes with the maximum number of editing sites vary across species [32]. Here, a total of 401 RNA editing sites within 34 PCGs were identified, and *ccmB* presented the greatest number of RNA editing sites. Most RNA editing events involved the conversion to hydrophobic amino acids in the mtDNA of *E. longipes*, which is helpful for altering physicochemical properties and increasing protein folding [31, 33, 34]. Another remarkable variation among these RNA editing sites featured hydrophilic amino acids changing to stop codons in the *rps10* and *atp6* genes, which may provide essential clues for nonfunctional editing and the consideration of pseudogenes in *E. longipes* [22, 35].

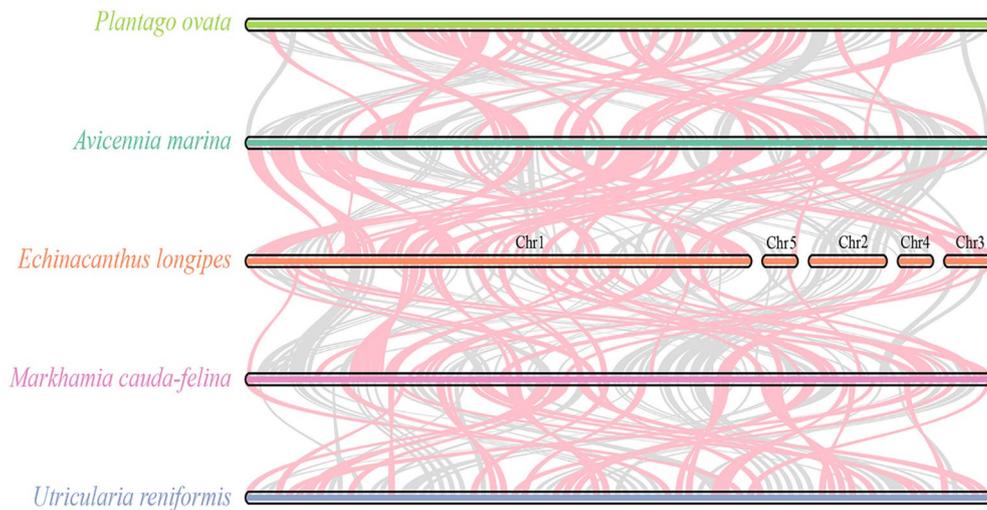


Fig. 9 Collinear analysis across the mitochondrial genomes of *E. longipes*, *A. marina* and three other closed species in Lamiales. The gray ribbons indicate fine homologous regions and the pink ribbons indicate the inversion regions

The phylogenetic relationships of *E. longipes* with other related species in Lamiales based on PCGs of the mitochondrial genomes were analyzed. Phylogenetic analyses of eight families confirmed the monophyly of Lamiales. Notably, the current phylogenetic relationships within Lamiales based on PCGs of the mitochondrial genomes were the same as those reported by APGIV [36]. Furthermore, *E. longipes* was determined to be closely related to *A. marina* in Acanthaceae. To further explore the evolutionary characteristics of the mitochondrial genomes within Acanthaceae, we exhaustively compared *E. longipes* with *A. marina*. Notably, the present study revealed relatively conserved GC contents but considerable structural variation in their mitochondrial genomes. And the different genome components exhibited different lengths ranging from 574,037 bp to 810,200 bp. Moreover, the core gene contents in the mitochondrial genomes of these two Acanthaceae species remained relatively consistent, but the numbers of noncore genes and tRNA genes copies varied significantly. Colinear analysis plays an important role in revealing species evolution and diversity [18, 37]. Therefore, further investigations of collinearity among the mitochondrial genomes of *E. longipes*, *A. marina* and three related species in Lamiales were conducted to determine their organization. The results revealed that these mitochondrial genomes exhibited substantial rearrangement and limited collinearity, which was considered a driving force for the mtDNA evolution of *E. longipes* in Lamiales. Thus, the results of the current phylogenetic and comparative analyses suggest that mitochondrial genomes are useful for studying the phylogeny, evolution and speciation.

Acanthaceae are a large family of angiosperms with high species, geographic, and ecological diversity [2, 38]. *Echinacanthus longipes* is a species endemic to karst

regions. Therefore, it is necessary to analyze the adaptive evolution of genes in *E. longipes*. The nonsynonymous substitution and synonymous substitution ratio (Ka/Ks) are very useful for measuring selective pressure at the protein level in mtDNA [39]. Genes with positive selection play key roles in adaptation to diverse environments [7, 40]. Ka/Ks analysis of PCGs between *E. longipes* and other Lamiales species was carried out. As a result, most of the PCG genes with Ka/Ks values < 1 underwent negative selection. However, only one gene, *ccmB*, with Ka/Ks > 1 had undergone positive selection compared with all the other selected species. Therefore, the *ccmB* gene, which plays an important role in cytochrome c biogenesis, was determined to be under positive pressure in *E. longipes*. Interestingly, this gene is also subjected to positive selection in *Suaeda glauca*, which is a prominent salt-tolerant species [17]. Notably, the *ccmB* gene had the most protein editing sites in mtDNA of *E. longipes*. A further study of the gene *ccmB* was necessary in the future.

Mitochondrial plastid DNAs (MTPTs), which are sequences transferred from plastid (e.g., chloroplast) to mtDNA, are common in angiosperms and originated at least 300 million years ago (Mya) [41, 42]. It was originally hypothesized that majority of the MTPTs usually became nonfunctional genes [43]. The lengths of MTPTs vary significantly across different species, ranging from less than 1 kb to more than 130 kb, and constitute 1–12% of mtDNA [18, 22, 44]. In the present analysis, *E. longipes* presented 118,235 bp MTPTs, accounting for 14.59% of the mtDNA. These results indicate that there were more transfer events in *E. longipes* than that in most reported plants. Moreover, the cpDNA of *E. longipes* contributed numerous sequences to its mtDNA, which may explain the complicated structure and diversity of the mtDNA of *E. longipes*. These MTPTs included tRNA genes,

protein-coding genes, intergenic regions, and even part of the rRNA gene. Notably, tRNA genes exhibited the highest frequency of MTPTs in *E. longipes*, a pattern consistent with prior observations in other species (e.g., *Capsicum pubescens* and *Astragalus membranaceus*) [45, 46]. In addition to the common tRNA genes in angiosperms, namely, *trnH*-GUG, *trnM*-CAU, *trnN*-GUU, *trnW*-CCA, *trnP*-UGG, *trnS*-GGA, and *trnD*-GUC (in dicots), five genes, *trnI*-GAU, *trnL*-UAG, *trnM*-CAU, *trnS*-UGA, and *trnV*-GAC, were identified in the MTPTs of *E. longipes*. Previous studies demonstrate that the transfer of a greater number of tRNA genes from chloroplasts to mitochondria is more conducive to meeting the demands of protein synthesis, thereby facilitating plant evolution and adaptation [34, 47]. Thus, in *E. longipes*, the increased number of tRNA genes associated with MTPTs may increase adaptation to karst environments. However, the underlying mechanism and role of MTPTs deserve in-depth study in the future.

Conclusions

This study successfully assembled and annotated the complete mtDNA of *E. longipes*, an endemic species in the Sino-Vietnamese karst region, for the first time. It spans 810,200 bp, with 61 unique genes encompassing 36 PCGs, 22 tRNA genes and three rRNA genes, and displays a complex structure in the form of two small circular chromosomes and three linear chromosomes. Two tandem repeats, 188 dispersed repeats and 160 simple sequence repeats were identified in the *E. longipes* mtDNA. Moreover, 401 RNA editing sites indicated that *ccmB* had the most protein editing sites and the most amino acids were converted to hydrophobic residues. Furthermore, this research confirmed that more MTPTs occurred in the tRNA genes of *E. longipes*. The results of the current phylogenetic and comparative analyses raise the possibility that mitochondrial genomes are useful for studying the phylogeny and speciation of Lamiales. Ka/Ks analysis revealed that most of the genes underwent negative selection. However, *ccmB* is the only gene that has undergone positive selection when compared with all the other selected species. This study will be invaluable for the mitochondrial study of Acanthaceae. It also provides extensive information for functional genetic and adaptive studies of *Echinacanthus* in karst regions in the future.

Materials and methods

Plant materials, DNA extraction and sequencing

Plant sample was investigated and collected by Yusong Huang, Yunfei Deng, Yi Tong et al. in its native habitats from Jingxi County, Guangxi, China. This area is a public land and permitted to collect *E. longipes*. The sample was identified by Yunfei Deng and the voucher specimen

(No.16101704) was deposited in Herbarium of South China Botanical Garden, Chinese Academy of Sciences (IBSC). Total genomic DNA was extracted from 100 mg silica gel-dried leaves following the method of CTAB [48]. After removing RNA contaminants and checking the quality of the DNA, an excellent integrity of DNA molecules was observed. The library construction and sequencing procedures were performed at Biomarker Technologies CO., LTD (Beijing, China). For PacBio HiFi sequencing, genomic DNA was fragmented to 15 kb to construct a long-read library (SMRT bell library) according to the manufacturer's instructions (Pacific Biosciences, CA, USA), and then the library was sequenced on a PacBio Sequel II platform. After filtering out the low-quality reads and sequence adapters, 23.69Gb of clean reads with N50 = 13.51Kb were obtained.

Assembly and annotation of mtDNA

The PacBio HiFi sequencing data was used to assemble the mtDNA. Mitochondrial genomes of three species, including *Liriodendron tulipifera* L. (NC_021152.1), *Ajuga reptans* L. (NC_023103.1), and *Rothea serrata* (L.) Steane & Mabb. (NC_049064.1) were downloaded from GenBank database as reference sequences. The Seqkit software was used to select 10X contigs as mitochondrial sequences [49]. Then, the assembly was performed using Flye software with the default parameters [50]. Bandage software was used to check the result of the assembly and screen the mitochondrial sequences again [51]. Finally, the complete mtDNA was obtained by the comparison results of blast. The mitogenomes were annotated by the online tool IPMGA (<http://www.1kmpg.cn/ipmga/>) using *L. tulipifera*, *A. reptans*, and *R. serrata* as reference genomes and visualized in the online tool OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) [52, 53].

Repeat sequence analysis

The simple sequence repeats of *E. longipes* were identified with MISA by setting the minimum number of repeats to 10, 5, 4, 3, 3 and 3 for mono-, di-, tri-, tetra-, penta- and hexanucleotides, respectively [54]. The online tool REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) was used to analyze dispersed repeats with settings as follows: 3 for hamming distance, 40 for minimal repeat size, and 5000 for maximum computed repeats. Tandem repeats were obtained by TRF network server (<https://tandem.bu.edu/trf/trf.html>).

Codon usage analysis

The PCGs were extracted using Phylosuite software with default settings [55]. MEGA v7.0 software was employed to analyze codon usage bias through calculation of RSCU values based on the PCGs of the mtDNA [56].

Prediction of RNA editing sites

The RNA editing sites from all PCGs encoded within the mtDNA of *E. longipes* were predicted on the Plant Predictive RNA Editor (PREP) suite (<http://prep.unl.edu/>) with a cut off value of 0.2.

Homology analysis

The cpDNA of *E. longipes* (NC_039761) was downloaded from NCBI Organelle Genome Resources Database. Identification of genes transferred from chloroplasts to mitochondria were detected using BLASTN software on NCBI with parameters “e-value $\leq 1e^{-10}$, length ≥ 50 , matching rate $\geq 90\%$ ”, and the outcomes were visually represented utilizing the advanced circos package in TBtools [57].

Phylogenomic and Ka/Ks analysis

Twenty-five mitochondria genomes of eight families in Lamiales and two outgroups were downloaded from GenBank. Thirty-one conserved protein-coding genes of all mitochondria genomes were extracted in Phylosuite software and aligned with MAFFT v.7 (auto-strategy) [58]. TVM+I+G model was selected as the best model using program Modeltest 3.7 [59]. Then, the concatenated data matrix of the PCGs was subjected to Bayesian analysis using MrBayes v.3.2 [60]. All parameters were set according to the TVM+I+G model as follows: statefreqpr = fixed (0.2604, 0.2130, 0.2118, 0.3149) revmat = fixed (1.4764, 1.5241, 0.5723, 0.9977, 1.5241, 1.000), shapepr = fixed (0.7854), pinvar = fixed (0.4345). The analysis implemented Markov chain Monte Carlo (MCMC) algorithm and ran for 1,000,000 generations. The first 1,000 trees were considered as the “burn-in” period and discarded. The remaining trees were used to construct the majority-rule consensus tree. Posterior probabilities > 0.95 were considered significant support for a clade. Meanwhile, the Ka and Ks rates of the 16 PCGs in *E. longipes* with the 24 species of Lamiales were analyzed in DnaSP v.6 [61]. And then ratios of Ka/Ks were calculated.

Colinear analysis

The comparisons between the mtDNA of *E. longipes* and four closely related species were conducted in Blastn 2.14.0+ software with setting as “-evaluate 1e-5 -outfmt 6”. Then the MCScanX software and Mauve 2.4.0 were used to generate the multiple synteny plot and analysis of mitogenome collinearity [62, 63].

Abbreviations

NCBI	National Center for Biotechnology Information.
PCGs	Protein-coding genes.
RSCU	Relative synonymous codon usage.
SSRs	Simple sequence repeats.
cpDNA	Chloroplast genome.
mtDNA	Mitochondrial genome.
MTPTs	Mitochondrial plastid DNAs.

Ka	Nonsynonymous substitutions.
Ks	Synonymous substitutions.
PP	Posterior probability.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11448-6>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6

Acknowledgements

We are grateful to Mr. Yi Tong and Ms. Chunyu Zou for their help in taking photos of *Echinacanthus longipes*.

Author contributions

CMG designed the project, analyzed the data and draft this paper. SW assembled and annotated the mitogenome. YSH investigated and collected the samples. YFD conceived and revised this paper. All authors read the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32070363, 31400184, 32470219), and Southeast Asia Biodiversity Research Institute, the Chinese Academy of Sciences (Y4ZK111B01).

Data availability

The accession number of *Echinacanthus longipes* mitogenome in Gene Bank (NCBI) is PQ164709- PQ164713. All data generated or analyzed during this study were provided as supplementary information files.

Declarations

Ethics approval and consents to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 September 2024 / Accepted: 5 March 2025

Published online: 14 March 2025

References

- McDade LA, Daniel TF, Kiel CA. Toward a comprehensive Understanding of phylogenetic relationships among lineages of Acanthaceae S.I. (Lamiales). *Am J Bot.* 2008;95(9):1136–52.
- Manzitto-Tripp EA, Darbyshire I, Daniel TF, Kiel CA, McDade LA. Revised classification of Acanthaceae and worldwide dichotomous keys. *Taxon.* 2021;71(1):103–53.
- Hu CC, Deng YF, Wood JRI. *Echinacanthus* Nees. In: Wu ZY, Raven PH, Hong DY, editors. *Flora of China* 19. St. Louis, MO: Sciences & Missouri Botanical Garden; 2011. pp. 380–430.
- Deng YF, Hai DV, Xuyen DT. *Echinacanthus* Nees (Acanthaceae), a newly recorded genus from Vietnam. *J Trop Subtrop Bot.* 2010;18:40–2.
- Wood JRI. Notes relating to the flora of Bhutan XXIX: acanthaceae, with special reference to *Strobilanthes*. *Edinb J Bot.* 1994;51:175–274.

6. Tripp EA, Daniel TF, Fatimah S, McDade LA. Phylogenetic relationships within Ruellieae (Acanthaceae) and a revised classification. *Int J Plant Sci.* 2013;174:97–137.
7. Gao CM, Deng YF, Wang J. The Complete Chloroplast Genomes of Echinacanthus Species (Acanthaceae): Phylogenetic Relationships, Adaptive Evolution, and Screening of Molecular Markers. *Front. Plant Sci.* 2019;9:1989.
8. Hao Z, Kuang YW, Kang M. Untangling the influence of phylogeny, soil and climate on leaf element concentrations in biodiversity hotspots. *Funct Ecol.* 2015;59:165–76.
9. Bullerwell CE. Organelle genetics: evolution of organelle genomes and gene expression. Dordrecht: Springer Science & Business Media; 2011.
10. Qu K, Chen Y, Liu D, Guo HL, Xu T, Jing Q, Ge L, Shu XG, Xin XW, Xie XM, Tong BQ. Comprehensive analysis of the complete mitochondrial genome of *Lilium Tsingtauense* reveals a novel multichromosome structure. *Plant Cell Rep.* 2024;43:150.
11. Cavalier-Smith T. The origin of nuclei and of eukaryotic cells. *Nature.* 1975;256:463–8.
12. Zimorski V, Ku C, Martin WF, Gould SB. Endosymbiotic theory for organelle origins. *Curr Opin Microbiol.* 2014;22:38–48.
13. Birky CW. The inheritance of genes in mitochondria and chloroplasts: law, mechanisms, and models. *Annu Res Genet.* 2001;35:125–48.
14. Wang J, Zou Y, Mower JP, Reeve W, Wu ZQ. Rethinking the mutation hypotheses of plant organellar DNA. *Genomics Commun.* 2024;1:e003.
15. Wei L, Liu TJ, Hao G, Ge XJ, Yan HF. Comparative analyses of three complete *primula* mitogenomes with insights into mitogenome size variation in Ericales. *BMC Genomics.* 2022;23(1):770.
16. Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J Mol Biol.* 1992;223(1):1–7.
17. Cheng Y, He XX, Priyadarshani SVGN, Wang Y, Ye L, Shi C, Ye KZ, Zhou Q, Luo ZQ, Deng F, Cao L, Zheng P, Aslam M, Qin Y. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. *BMC Genomics.* 2021;22:167.
18. Chen LL, Dong X, Huang H, Xu HX, Rono PC, Cai XZ, Hu GW. Assembly and comparative analysis of the initial complete mitochondrial genome of *Primulina hunanensis* (Gesneriaceae): a cave-dwelling endangered plant. *BMC Genomics.* 2024;25:322.
19. Wang L, Liu X, Xu YJ, Zhang ZW, Wei YS, Hu Y, Zheng CB, Qu XY. Assembly and comparative analysis of the first complete mitochondrial genome of a traditional Chinese medicine *Angelica biserrate* (Shan et Yuan) Yuan et Shan. *Int J Biol Macromol.* 2024;257(1):128571.
20. Liu F, Fan W, Yang JB, Xiang CL, Mower JP, Li DZ, Zhu A. Episodic and guanine-cytosine-biased bursts of intragenomic and interspecific synonymous divergence in Ajugoideae (Lamiaceae) mitogenomes. *New Phytol.* 2020;228:1107–14.
21. Li D, Guo HL, Zhu JL, Qu K, Chen Y, Guo YT, Ding P, Yang HP, Xu T, Jing Q, Han SJ, Li W, Tong BQ. Complex physical structure of complete mitochondrial genome of *Quercus acutissima* (Fagaceae): A significant energy plant. *Genes.* 2022;13(8):1321.
22. Guo S, Li ZY, Li CL, Liu Y, Liang XL, Qin YM. Assembly and characterization of the complete mitochondrial genome of *Ventilago Leiocarpa*. *Plant Cell Rep.* 2024;43:77.
23. Morley SA, Nielsen BL. Plant mitochondrial DNA. *Front Biosci.* 2017;22(6):1023–32.
24. Yang H, Li W, Yu X, Zhang X, Zhang Z, Liu Y, Wang W, Tian X. Insights into molecular structural, genome evolution and phylogenetic implication through mitochondrial genome sequence of *Gleditsia sinensis*. *Sci Rep.* 2021;11(1):14850.
25. Ma QY, Wang YX, Li SS, Wen J, Zhu L, Yan KY, Du YM, Ren J, Li SX, Chen Z, Bi CW, Li QZ. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: a Woody oil-tree species producing nervonic acid. *BMC Plant Biol.* 2022;22:29.
26. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. *G3-Genes, Genomes Genet.* 2019;9(2):549–59.
27. Jiang M, Ni Y, Li JL, Liu C. Characterization of the complete mitochondrial genome of *Taraxacum Mongolicum* revealed five repeat-mediated recombination. *Plant Cell Rep.* 2023;42:775–89.
28. Yang HY, Ni Y, Zhang XY, Li JL, Chen HM, Liu C. The mitochondrial genomes of *Panax Notoginseng* reveal recombination mediated by repeats associated with DNA replication. *Int J Biol Macromol.* 2023;252:126359.
29. Hao ZG, Zhang ZP, Jiang J, Pan L, Zhang JN, Cui XF, Li YB, Li JQ, Luo LX. Complete mitochondrial genome of *Melia azedarach* L., reveals two conformations generated by the repeat sequence mediated recombination. *BMC Plant Biol.* 2024;24:645.
30. Bock H, Brennicke A, Schuster W. rps3 and rpl16 genes do not overlap in *Oenothera* mitochondria: GTG as a potential translation initiation codon in plant mitochondria? *Plant Mol Biol.* 1994;24(5):811–8.
31. Samuel CE. Transcription RNA editing. In: Jez J, editor. *Encyclopedia of biological chemistry III.* 3rd ed. Oxford: Elsevier; 2019. pp. 449–54.
32. Salmana ML, Chaw SM, Lin CP, Shin AC, Wu YW, Mulligan RM. Editing site analysis in a gymnosperm mitochondrial genome reveals similarities with angiosperm mitochondrial genomes. *Curr Genet.* 2010;56:439–46.
33. Yura K, Go M. Correlation between amino acid residues converted by RNA editing and functional residues in protein three-dimensional structures in plant organelles. *BMC Plant Biol.* 2008;8:79.
34. Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, Jiang A, Ye Q, Ye N. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by cooperative genomics approaches. *BioMed Res Int.* 2016. <https://doi.org/10.1155/2016/5040598>.
35. Schuster W, Brennicke A. RNA editing makes mistakes in plant mitochondria: editing loses sense in transcripts of a rps19 pseudogene and in creating stop codons in coxi and rps3 mRNAs of *Oenothera*. *Nucleic Acids Res.* 1991;19:6923–8.
36. Angiosperm Phylogeny Group. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc.* 2016;181(1):1–20.
37. Liu G, Cao D, Li S, Su A, Geng J, Grover CE, et al. The complete mitochondrial genome of *Gossypium hirsutum* and evolutionary analysis of higher plant mitochondrial genomes. *PLoS ONE.* 2013;8(8):e69476.
38. Tripp EA, Siti F. Comparative anatomy, morphology and molecular phylogenetics of the African genus *Satanocrater* (Acanthaceae). *Am J Bot.* 2012;99:967–82.
39. Xie P, Wu JR, Lu MY, Tian TX, Wang DM, Luo ZW, Yang DH, Li LL, Yang XW, Liu DC, Cheng HT, Tan JX, Yang HS, Zhu DQ. Assembly and comparative analysis of the complete mitochondrial genome of *Fritillaria ussuriensis* Maxim. (Liliaceae), an endangered medicinal plant. *BMC Genomics.* 2024;25:773.
40. Fan WB, Wu Y, Yang J, Shahzad K, Li ZH. Comparative Chloroplast genomics of *Dipsacales* species: insights into sequence variation, adaptive evolution, and phylogenetic relationships. *Front. Plant Sci.* 2018;9:689.
41. Wang D, Wu YW, Shih ACC, Wu CS, Wang YN, Chaw SM. Transfer of Chloroplast genomic DNA to mitochondrial genome occurred at least 300 Mya. *Mol Biol Evol.* 2007;24(9):2040–8.
42. Wang XC, Chen H, Yang D, Liu C. Diversity of mitochondrial plastid DNAs (MTPTs) in seed plants. *Mitochondrial DNA DNA Mapp Seq Anal.* 2018;29:635–42.
43. Warren JM, Sloan DB. Interchangeable parts: the evolutionarily dynamic tRNA population in plant mitochondria. *Mitochondrion.* 2020;52:144–56.
44. Sloan DB, Wu Z. History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. *Genome Biol Evol.* 2014;6:3210–21.
45. Zhang K, Qu GY, Zhang Y, Liu JX. Assembly and comparative analysis of the first complete mitochondrial genome of *Astragalus Membranaceus* (Fisch.) Bunge: an invaluable traditional Chinese medicine. *BMC Plant Biol.* 2024;24:1055.
46. Li L, Fu HZ, Altaf MA, Wang ZW, Lu X. The complete mitochondrial genome assembly of *Capsicum pubescens* reveals key evolutionary characteristics of mitochondrial genes of two *Capsicum* subspecies. *BMC Genomics.* 2024;25:1064.
47. Qu Y, Zhou P, Tong C, Bi C, Xu L. Assemble and analysis of the *Populus deltoides* mitochondrial genome: the first report of a multicircular mitochondrial conformation for the genus *Populus*. *J Res.* 2023;34(3):717–33.
48. Doyle JJ, Doyle JL. A rapid total DNA Preparation procedure for fresh plant tissue. *Focus.* 1990;12:13–5.
49. Shen W, Le S, Li Y, Hu FQ, SeqKit. A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE.* 2016;11(10). <https://doi.org/10.1371/journal.pone.0163962>.
50. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37(5):540–6.
51. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de Novo genome assemblies. *Bioinformatics.* 2015;31(20):3350–52.

52. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017;45:W6–11.
53. Greiner S, Lehwark P, Bock R. Organellar genomedraw (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 2019;47(W1):W59–64.
54. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33:2583–5.
55. Zhang D, Gao F, Li WX, Jakovlić I, Zou H, Zhang J, Wang GT. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 2019;20:348–55.
56. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis Version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
57. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13(8):1194–202.
58. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
59. Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics.* 1998;14:817–8.
60. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61:539–42.
61. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. Mol Biol Evol. 2017;34(12):3299–302. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets.
62. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40:e49.
63. Darling AE, Mau B, Perna NT. Progressive Mauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE.* 2010;5:e111147.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.