

RESEARCH

Open Access



# Complete mitogenome assembly of *Selenicereus monacanthus* revealed its molecular features, genome evolution, and phylogenetic implications

Guilong Lu<sup>1</sup>, Wenhua Wang<sup>2</sup>, Juan Mao<sup>2</sup>, Qing Li<sup>2\*</sup> and Youxiong Que<sup>3\*</sup>

## Abstract

**Background** Mitochondria are the powerhouse of the cell and are critical for plant growth and development. Pitaya (*Selenicereus* or *Hylocereus*) is the most important economic crop in the family Cactaceae and is grown worldwide, however its mitogenome is unreported.

**Results** This study assembled the complete mitogenome of the red skin and flesh of pitaya (*Selenicereus monacanthus*). It is a full-length, 2,290,019 bp circular molecule encoding 59 unique genes that only occupy 2.17% of the entire length. In addition, 4,459 pairs of dispersed repeats ( $\geq 50$  bp) were identified, accounting for 84.78% of the total length, and three repeats (394,588, 124,827, and 13,437 bp) mediating genomic recombination were identified by long read mapping and Sanger sequencing. RNA editing events were identified in all 32 protein-coding genes (PCGs), among which four sites (*nad1-2*, *nad4L-2*, *atp9-copy3-223*, and *ccmFC-1309*) were associated with the initiation or termination of PCGs. Seventy-eight homologous fragments of the chloroplast genome were identified in the mitogenome, the longest having 4,523 bp. In addition, evolutionary analyses suggest that *S. monacanthus* may have undergone multiple genomic reorganization events during evolution, with the loss of at least nine PCGs (*rpl2*, *rpl10*, *rps2*, *rps3*, *rps10*, *rps11*, *rps14*, *rps19*, and *sdh3*).

**Conclusions** This study revealed the genetic basis of the *S. monacanthus* mitogenome, and provided a scientific basis for further research on phenotypic traits and germplasm resource development.

**Keywords** *Selenicereus Monacanthus*, Mitogenome, RNA editing, Gene loss, Evolution analysis

\*Correspondence:

Qing Li

liqing130811@126.com

Youxiong Que

queyouxiong@126.com

<sup>1</sup>College of Horticulture and Landscape Architecture, Henan Institute of Science and Technology, Xinxiang 453003, China

<sup>2</sup>Institute of Vegetables, Tibet Academy of Agricultural and Animal Husbandry Sciences, Lhasa 890032, China

<sup>3</sup>College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

Mitochondria are important organelles within eukaryotic cells that are central to cellular respiration and energy metabolism [1]. They originated from endosymbiotic Alphaproteobacteria and became semi-autonomous organelles by gradually reducing their autonomy through gene transfer to the host cell nucleus [2, 3]. Mitochondria are maternally inherited in plants [4], except in some plants such as *Chlorophytum* [5] and *Cucumis* [6], and the green alga, *Chlamydomonas* [7]. There are various types of plant mitogenomes (including circular, linear, and reticulate) [8], with genome sizes ranging from 66 Kbp (*Viscum scurruloideum*) [9] to 11.7 Mbp (*Larix sibirica* Ledeb.) [10], and large interspecies variation within the same genus [11, 12]. The sequence and structure of plant mitogenomes are highly variable owing to widespread horizontal gene transfer and genome rearrangement [13]. However, the mitogenomes of higher plants have a smaller gene density, with their gene coding regions typically accounting for approximately 10% of their mitogenome, together with many repetitive sequences [14] and RNA editing sites [15]. Simple sequence repeats (SSRs) in plant mitogenomes are often used as genetic markers [16]. Therefore, plant mitogenomes have become important tools for species identification, phylogenetic analysis, and inheritance patterns [17, 18].

Mitochondria play an important role in plant development, ecological adaptation, and reproduction [1, 19]. In the mitochondrial genome, due to the frequent insertion/loss of genes, gene fragments or non-coding sequences, and repetitive sequence recombination, the normal functional exercise of mitochondrial genes are largely affected and thus the agronomic traits in plants are altered [8, 13, 20]. Numerous studies show that plant mitochondria are closely associated with traits including stress tolerance, plant growth vigor [21, 22], and cytoplasmic male sterility [23, 24]. Dispersed repeats (also known as transposable elements) are a class of DNA sequences that can move their position on the genome, regulate gene expression, and influence plant phenotypic traits [25] such as fruit shape in tomatoes [26], fruit color in apples [27], and plant and ear height in maize [28]. However, the abundance of repetitive sequences and complex physical structures make the assembly of complete plant mitogenome sequences particularly difficult, with complete mitogenomes reported for only 602 species to date. This is much lower than the number of chloroplast (No. 10,479) and plastid (No. 1,301) genomes (April 5, 2023, <https://www.ncbi.nlm.nih.gov/genome/browse/#/organelles/>). Therefore, assembling and deciphering the mitogenome of a species is important for a deeper understanding of its genetic characteristics and for breeding research.

Pitaya (also known as pitahaya or dragon fruit) belongs to the genus *Selenicereus* or *Hylocereus* of the Cactaceae family [29, 30]. It originated in Costa Rica, Mexico, Colombia, and other Central American regions and is now widely grown in tropical and subtropical regions [31]. It is nutritionally rich and unique in function; it contains plant albumin, betaine, and water-soluble dietary fibers that are rarely found in general plants, and has high ornamental and medicinal value [32, 33]. The pitaya industry has rapidly developed in recent years, and there is an urgent need for high yielding quality varieties that are resistant to the biotic and abiotic stresses associated with production [34]. Basic genetic research on pitaya is important to further promote the use of superior germplasm resources, improve agronomic traits, and ensure industrial safety [35–37]; however, studies on its mitochondrial genome are lacking.

This study chose red skin and red flesh pitaya (*Selenicereus monacanthus*) to conduct the following research: (1) assemble the mitochondrial genome and describe its features, (2) identify repetitive sequences and predict recombination, (3) predict and validate the presence of RNA editing events, (4) assemble its chloroplast genome and identify homologous fragments with the mitogenome, and (5) perform phylogenetic and synteny analysis of closely related species. We expect this study to provide a scientific and theoretical basis for an in-depth understanding of the genetic characteristics and evolutionary history of *S. monacanthus*.

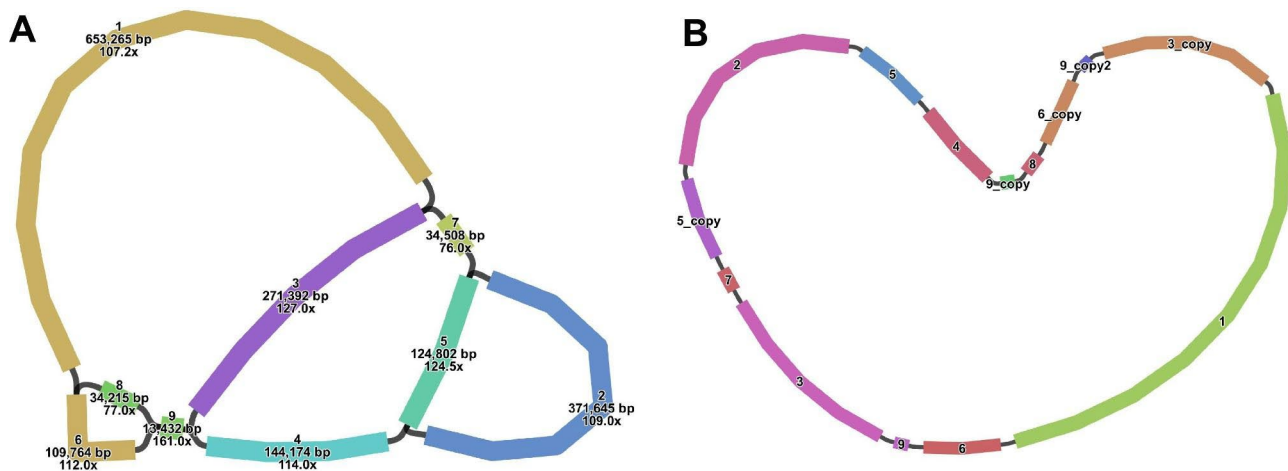
## Results

### *S. monacanthus* mitogenome assembly

The *S. monacanthus* mitogenome was assembled using 10.50 Gb short-reads and 9.62 Gb long-reads using a hybrid assembly strategy. The genome sketch contains nine contigs (Fig. 1A). Contig1 and contig9 are the longest and shortest with lengths of 653,265 bp and 13,432 bp, respectively, and they also include three double bifurcating structures. We obtained a simplified main circular structure (Fig. 1B) with a total length of 2,290,019 bp after excluding the repetitive regions using the Nanopore data. This genome size was significantly larger than that of *Pereskia aculeata* (515,187 bp, NC\_067638.1) of the same family [38].

### Molecular features

The GC content of the *S. monacanthus* mitogenome was 43.37%, with adenine (A), thymine (T), cytosine (C), and guanine (G) representing 28.25%, 28.38%, 21.74%, and 21.63%, respectively. Thirty-two unique PCGs were annotated, including 24 core genes and eight non-core genes, as well as 24 tRNA genes (of which 14 tRNAs were multiple copies), and three rRNA genes (all multiple copies) (Fig. 2; Table 1). Nine genes were lost in the



**Fig. 1** *S. monacanthus* mitogenome sketch (A) and master circle structure (B)

*S. monacanthus* mitogenome compared to the PCGs of “fossilized” *Liriodendron tulipifera* (*rpl2*, *rpl10*, *rps2*, *rps3*, *rps10*, *rps11*, *rps14*, *rps19*, and *sdh3*) [39], while *rps3* was present in *P. aculeata* of the same family [38]. In addition, the total length of the PCGs (35,235 bp), tRNA (4,427 bp), and rRNA (10,066 bp) coding sequences was 49,728 bp. This accounted for 2.17% of the whole genome, while over 97% of the regions were intergenic.

The region of the *S. monacanthus* mitogenome encoding amino acids developed a unique codon usage preference during the evolution of plant adaptation (Fig. S1 and Table S1). There was a general codon preference for leucine (Leu: UUA, RSCU=1.59), alanine (Ala: GCU, RSCU=1.58), and a non-preference for glutamine (Gln: CAG, RSCU=0.46) and tyrosine (Tyr: UAC, RSCU=0.50), while the universal start codon was AUG and that for tryptophan was only UGG (both RSCUs=1.00). The termination codon (End) preferred UAA (RSCU=1.55), and not UAG (RSCU=0.41).

#### Repeat elements and repeat-mediated recombination

In the *S. monacanthus* mitogenome, several repetitive sequences were observed (Fig. S2). A total of 616 SSRs were identified (Fig. 3A and Table S2), with the monomeric and dimeric forms accounting for 45.78% of the total SSRs. Adenine (A) monomeric repeats accounted for 54.04% (107/198) of the monomeric SSRs. In addition, 94 tandem repeats with  $\geq 74\%$  matches and lengths between 10 and 45 bp were identified in this genome (Table S3). There were 4,459 pairs of dispersed repeats with a length of  $\geq 50$  bp (Fig. 3B and Table S4), including 2,345 pairs of palindromic repeats and 2,114 pairs of forward repeats. The longest forward and backward repeats were 394,588 bp and 13,437 bp, respectively. However, no reverse repeat or complementary repeat was detected. In addition, the total lengths of the SSRs, tandem repeats, and dispersed repeats were 7,130 bp, 4,557 bp, and

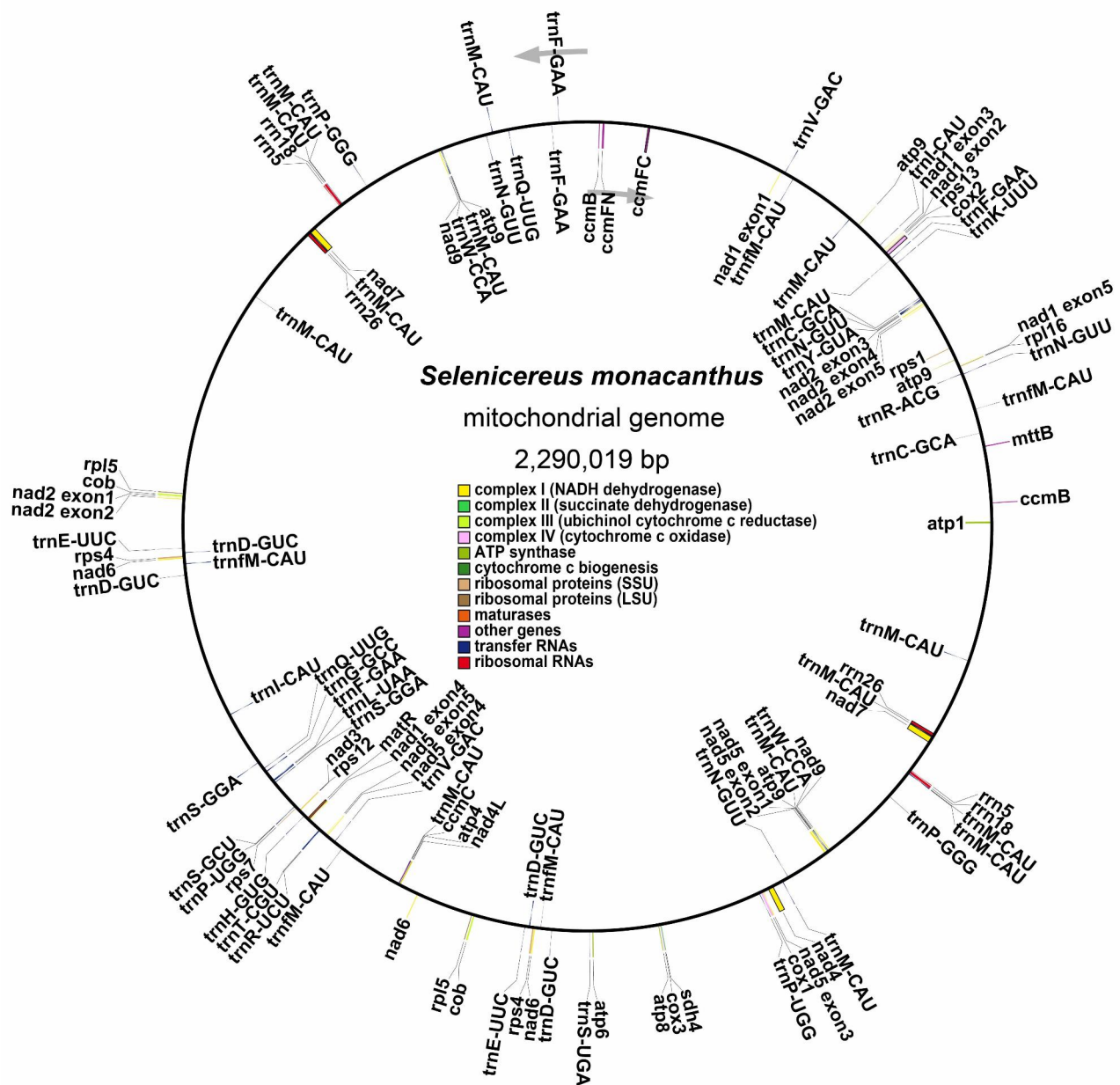
1,941,444 bp accounting for 0.31%, 0.20%, and 84.78% of the mitogenome length, respectively.

Repetitive sequences that mediate genomic recombination may lead to multiple conformations in plant mitogenomes [40]. Specifically, there may be a secondary genomic structure mediated by repeat fragments R1 (394,588 bp, contig3+contig9+contig6), R2 (124,827 bp, contig5), and R3 (13,437 bp, contig9) (Table 2) in the *S. monacanthus* mitogenome. The recombination mediated by these three repeats was identified using a validated junction approach; the primer design and electrophoresis results are shown in Fig. S3, and detailed sequencing comparison results are shown in Fig. S4. In short, there are multiple potential recombination conformations in the *S. monacanthus* mitogenome.

#### RNA editing events in the PCGs

There were a total of 398 RNA editing sites, and they were spread throughout the 32 PCGs of the *S. monacanthus* mitogenome. Each event was a C to U conversion (Fig. 4A and Table S5), and their editing frequencies were mostly above 0.80 (Fig. 4B). Among these, the highest number of RNA editing sites was in the *ccmB-copy2* gene (No. 32), followed by that in the *ccmC* gene (No. 31). In addition, non-synonymous codon changes in the 356 RNA editing events mainly involved the following five amino acid changes: Ser to Leu (No. 87), Pro to Leu (No. 78), Ser to Phe (No. 44), Pro to Ser (No. 35), and Arg to Trp (No. 32).

In particular, we predicted that four editing sites were associated with the creation of start and stop codons in PCGs, that is, ACG (Thr) to AUG (Met) changes found in *nad1-2* and *nad4L-2*, and CGA (Arg) to UGA (End) changes found in *atp9-copy3-223* and *ccmFC-1309*. These four editing sites were validated by PCR products and Sanger sequencing comparison results (Fig. 4C and Supplementary File 1), where *atp9-copy3-223* and



**Fig. 2** Schematic mitogenome diagram of *S. monacanthus*. Genes belonging to different functional groups are color-coded

*ccmFC*-1309 were edited with low frequency (the red line indicates base T in the transcription). However, its specific effects on the function and metabolism of mitochondria in plants are unknown.

#### DNA transfer

Mitochondrial plastid DNAs (MTPTs) are plastid-derived DNA fragments found in the mitochondrial genome. In the present study, sequencing data were used to assemble the *S. monacanthus* chloroplast (cp.) genome, which was 133,408 bp in size (Fig. 5A). A total of 78 MTPTs were identified in the *S. monacanthus* cp. genome (Fig. 5B

and Table S6), with a total length of 46,496 bp, accounting for 2.03% of the mitogenome length. There were 16 fragments with lengths greater than 1,000 bp, of which MTPT18 was the longest at 4,523 bp. Twenty-five complete genes were identified, including 14 PCGs (*atpA*, *atpB*, *atpE*, *psbA*, *psbD*, *psbE*, *psbF*, *psbJ*, *psbL*, *rpoC1*, *rps2*, *rps4*, *rps7*, and *ycf15*), and 11 tRNA genes (*trnD-GUC*, *trnF-GAA*, *trnH-GUG*, *trnN-GUU*, *trnM-CAU*, *trnR-ACG*, *trnR-UCU*, *trnS-GGA*, *trnT-CGU*, *trnV-GAC*, and *trnW-CCA*). In addition, 30 plastid gene fragments were identified among the homologous fragments. Detailed information on the DNA transfer fragments

**Table 1** Gene composition in the *S. monacanthus* mitogenome

Group of genes		Name of genes
Core genes	ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i> (×4)
	NADH dehydrogenase	<i>nad1, nad2, nad3, nad4, nad4L, nad5, nad6</i> (×3), <i>nad7</i> (×2), <i>nad9</i> (×2)
	Cytochrome <i>b</i>	<i>cob</i> (×2)
	Cytochrome <i>c</i> biogenesis	<i>ccmB</i> (×2), <i>ccmC, ccmFC, ccmFN</i>
	Maturases	<i>matR</i>
	Protein transport subunit	<i>mttB</i>
	Variable genes	Ribosomal protein large subunit
Ribosomal protein small subunit		<i>rps1, rps4</i> (×2), <i>rps7, rps12, rps13</i>
Succinate dehydrogenase		<i>sdh4</i>
rRNA genes	Ribosome RNA	<i>rrn5</i> (×2), <i>rrn18</i> (×2), <i>rrn26</i> (×2)
tRNA genes	Transfer RNA	<i>trnC-GCA</i> (×2), <i>trnD-GUC</i> (×4), <i>trnE-UUC</i> (×2), <i>trnF-GAA</i> (×4), <i>trnI-M-CAU</i> (×5), <i>trnG-GCC, trnH-GUG, trnI-CAU</i> (×2), <i>trnK-UUU, trnL-UAA, trnM-CAU</i> (×14), <i>trnN-GUU</i> (×4), <i>trnP-GGG</i> (×2), <i>trnP-UGG</i> (×2), <i>trnQ-UUG</i> (×2), <i>trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA</i> (×2), <i>trnS-UGA, trnT-CGU, trnV-GAC</i> (×2), <i>trnW-CCA</i> (×2), <i>trnY-GUA</i>

Note: The number in brackets represents the copy number of the gene

and gene annotations is presented in Table S6. However, these homologous genes are pseudogenized in mitochondria and do not exercise their normal functions [41], and their specific roles remain to be studied in depth in *S. monacanthus*.

### Genome evolution

Plant mitochondrial genomes commonly undergo a loss or gain of PCGs during evolution [13, 42]. Only 19 PCGs were common among *S. monacanthus* and 24 related genera used for phylogenetic analysis: *atp1, atp4, atp6, atp8, ccmB, ccmC, ccmFC, ccmFN, cox2, cox3, matR, nad1, nad2, nad3, nad5, nad6, nad7, nad9, and sdh4*. The phylogenetic tree showed that *S. monacanthus* was closely related to *P. aculeata* (Fig. 6A). Moreover, the topology based on the phylogeny of mitochondrial DNA coincided with the latest classification of the angiosperm phylogenetic group.

Collinear relationships between *S. monacanthus* and seven related species in Caryophyllales showed that 227 colinear blocks were detected between *S. monacanthus* and *P. aculeata* of the Cactaceae family; this accounted for 99.68% (513,522 bp) of the entire *P. aculeata* mitogenome but only 22.42% of the *S. monacanthus* genome (Fig. 6B and Table S7). Many homologous syntenic regions were detected between *S. monacanthus* and closely related species, however the length of these

colinear blocks was short. Among the mitogenomes of these species, the inconsistent order of the colinear block arrangement implies that *S. monacanthus* may have undergone multiple genomic rearrangement events with closely related species.

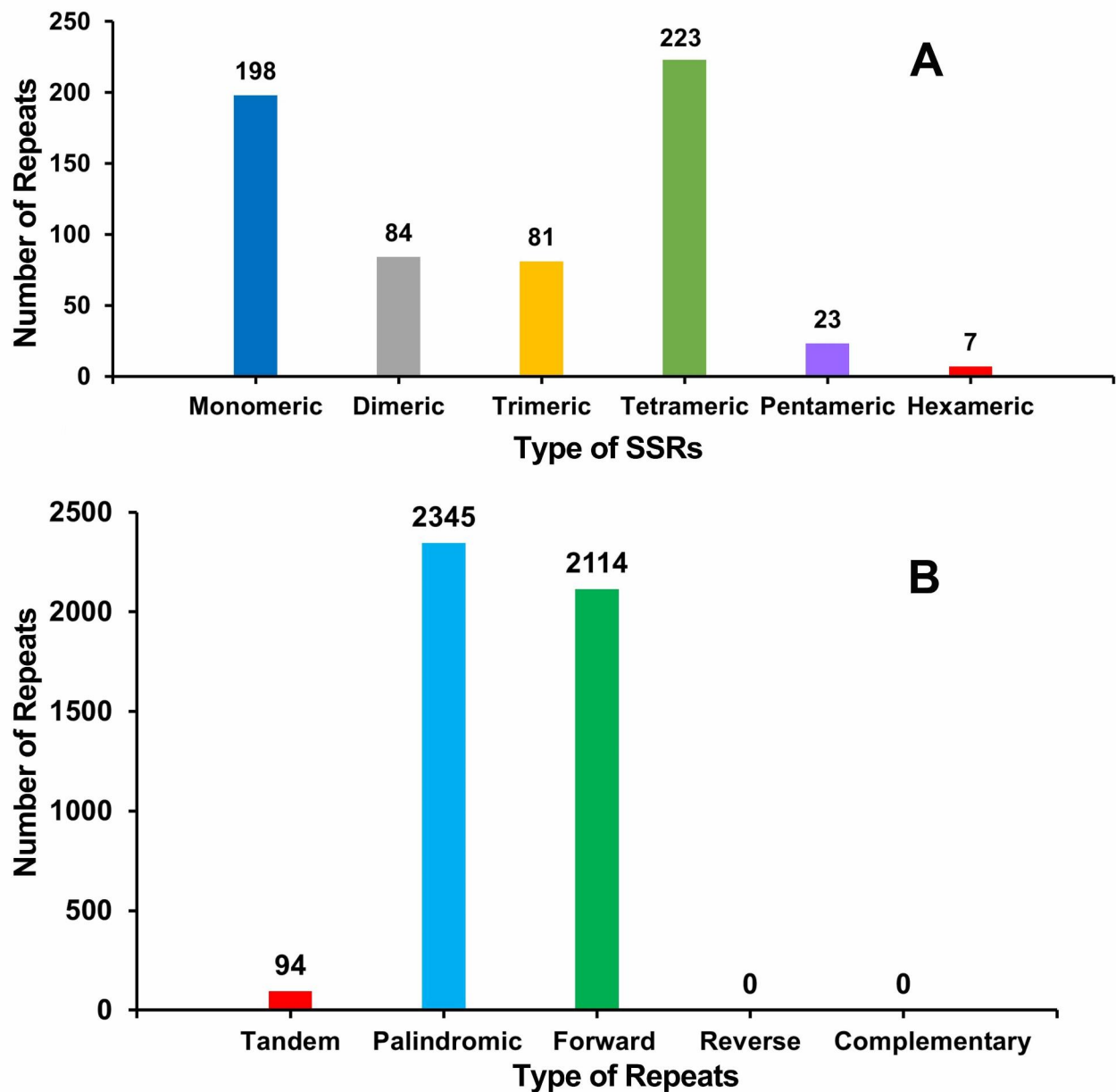
## Discussion

### Size and genetic composition properties of the *S. monacanthus* mitogenome

This study combined short- and long-reads using a hybrid assembly strategy to assemble a high-quality *S. monacanthus* mitogenome, which is a full-length 2,290,019 bp ring-like molecule and one of the larger genomes, significantly larger than that of *P. aculeata* (515.2 kb) in the same family [38]. The diversity of GC content in plant mitogenomes reflects their adaptive consequences [11, 43]. The GC content in the mitogenome of *S. monacanthus* was 43.37%. This was relatively less different from that of other terrestrial plants (23.9–50.5%). Gene transfer is the main pathway for the loss of mitogenome PCGs in plants, most of the transferred genes are ribosomal protein-encoding genes, with only some respiratory chain-related genes transferred to the nucleus during evolution (*rps2, rps11, and rps19*) [42]. *L. tulipifera* is a “fossilized” plant whose mitogenome evolved very slowly, retaining 41 PCGs from the ancestral angiosperms [39]. It was hypothesized that *S. monacanthus* lost at least nine PCGs through gene loss or transfer events during evolution (*rpl2, rpl10, rps2, rps3, rps10, rps11, rps14, rps19, and sdh3*). In addition, gene distribution density of the *S. monacanthus* mitogenome was very low. The coding sequence was highly conserved and its coding region accounted for only 2.17% of the full genome length; this was significantly lower than that of *L. tulipifera* (excluding *cis*-spliced introns, 7.9%) [39] and *Populus simonii* (8.25%) [44].

### Repeated sequences are exceptionally abundant in the *S. monacanthus* mitogenome

Repetitive sequences are abundant in the mitogenome and play important roles in the evolution of plant adaptation, regulation of gene expression, and variation in epistatic traits [45]. This study obtained 616 SSRs from the *S. monacanthus* mitogenome. This provides a large number of reference loci for further species identification and genetic evolution. Moreover, mitogenome-coding sequences have a slower evolutionary rate than chloroplast and nuclear genes [11]. Therefore, the development of mitogenome molecular markers is more accurate for species identification. Numerous studies show that dispersed repeats can affect plant phenotypic traits through the expression of regulatory genes [26–28, 46]. A total of 4,459 pairs of dispersed repeats ≥ 50 bp in length were detected in the *S. monacanthus* mitogenome; they mainly



**Fig. 3** Type and number of SSRs and repeats in the *S. monacanthus* mitogenome. **(A)** Type and number of SSRs in the *S. monacanthus* mitogenome. The dark blue, gray, orange, light green, purple, and red legend indicates monomeric, dimeric, trimeric, tetrameric, pentameric, and hexameric SSRs, respectively. **(B)** Type and number of repeats in the *S. monacanthus* mitogenome. The red, blue, and green legend indicates tandem repeats, palindromic repeats, and forward repeats, respectively

existed in the intergenic region and accounted for 84.78% of the mitogenome. This is one of the largest numbers of repeats identified in all other published mitogenic sequences. This implies that dispersed repeats may play an important role in genome expansion and gene regulation, and also provides scientific reference information for further study of their effects on agronomic traits in *S. monacanthus*.

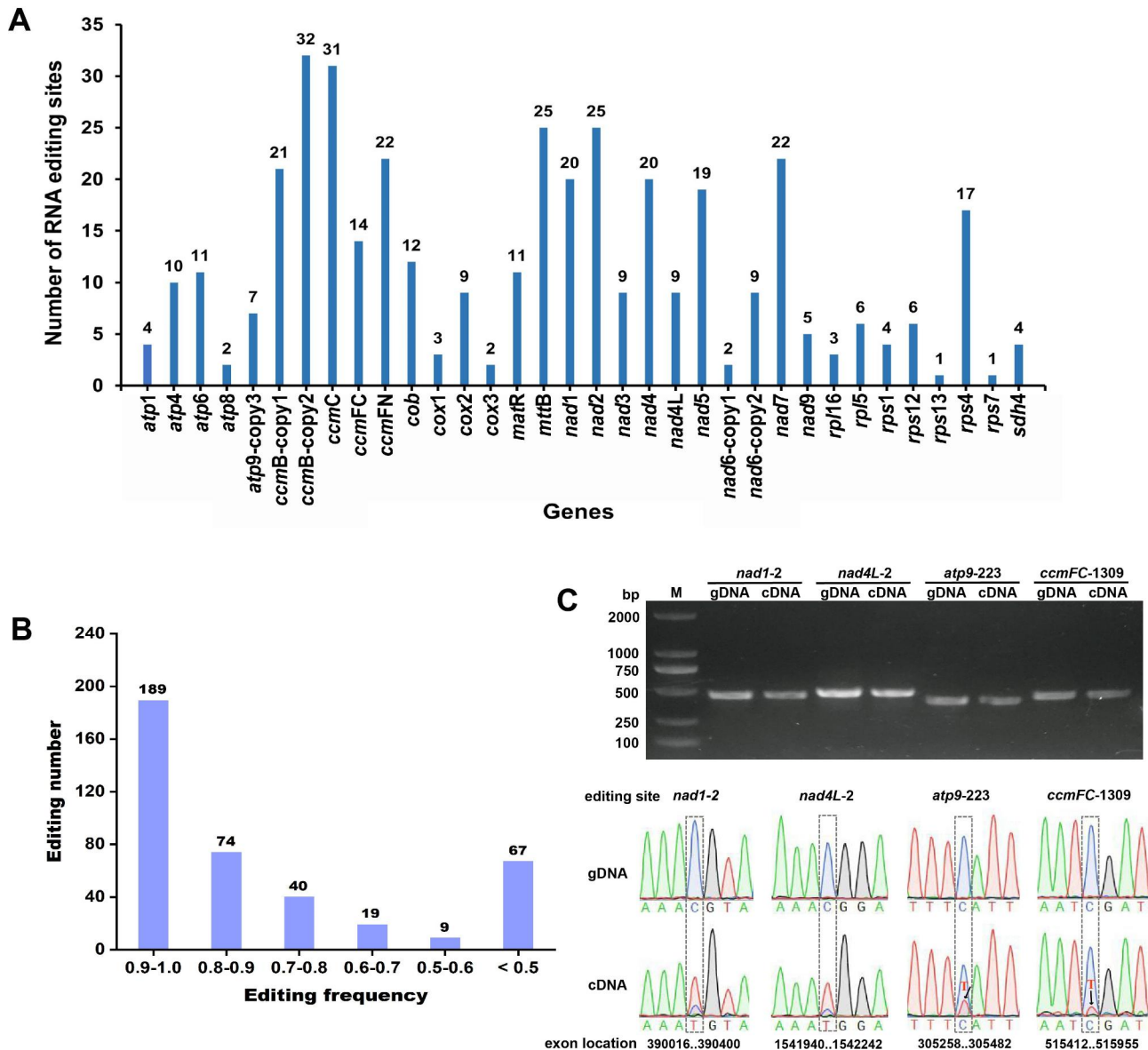
#### ***S. monacanthus* has multiple conformations**

Plant mitogenomes are commonly characterized by rearrangements that are important in promoting genome evolution and enriching genetic diversity [8, 40]. Moreover, the frequency of repeated recombination is related to the length of the repeat sequence and the characteristics of the species. In general, long repetitive sequences (>1000 bp) with high similarity are more likely to recombine [14]. We predicted and confirmed the existence of

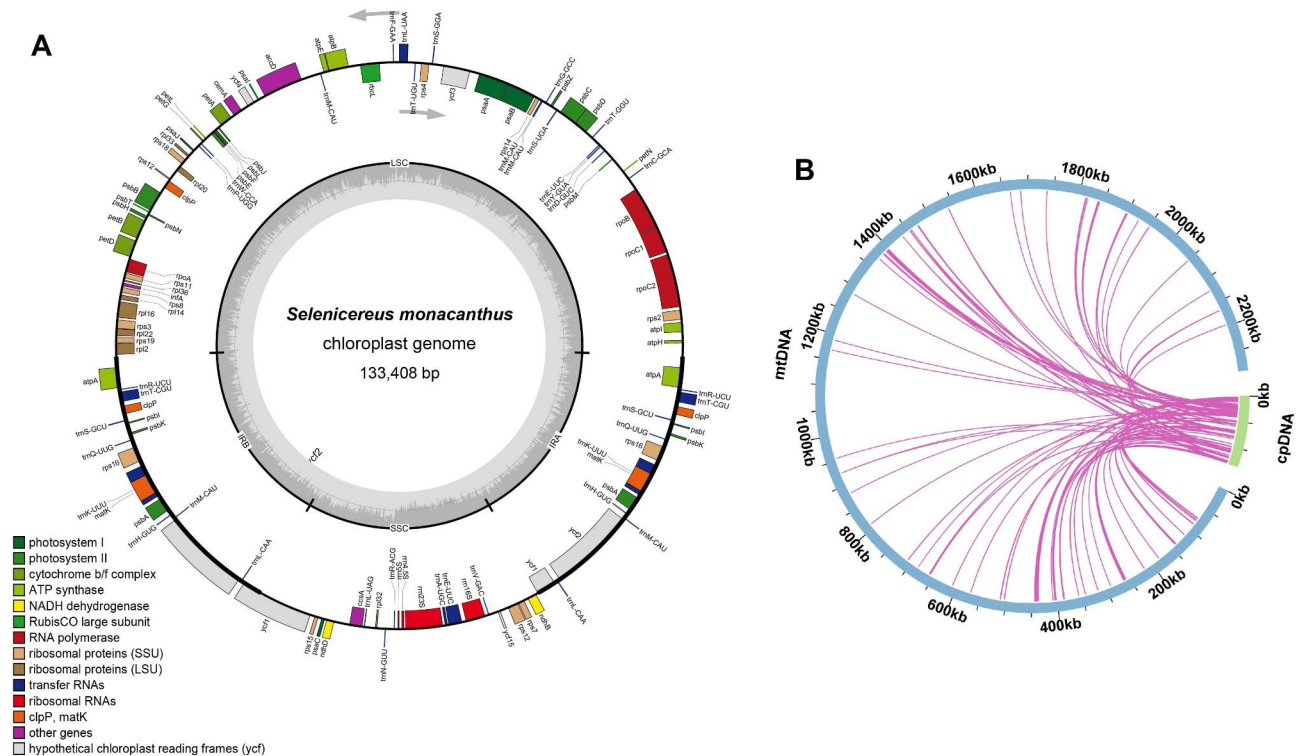
genomic recombination mediated by three repeats in the *S. monacanthus* mitogenome based on long-read mapping results: R1 (394,588 bp), R2 (124,827 bp), and R3 (13,437 bp). However, the presence of short-repeat sequence-mediated recombination could not be determined, and this needs to be supported by high-sequencing depth data. This implies possible differentiation within the mitogenome of the genus *Selenicereus* and provides useful information to gain insight into the evolution of genomes in *S. monacanthus* and closely related genera.

**Table 2** List of three repeated sequences mediating genomic recombination in the *S. monacanthus* mitogenome

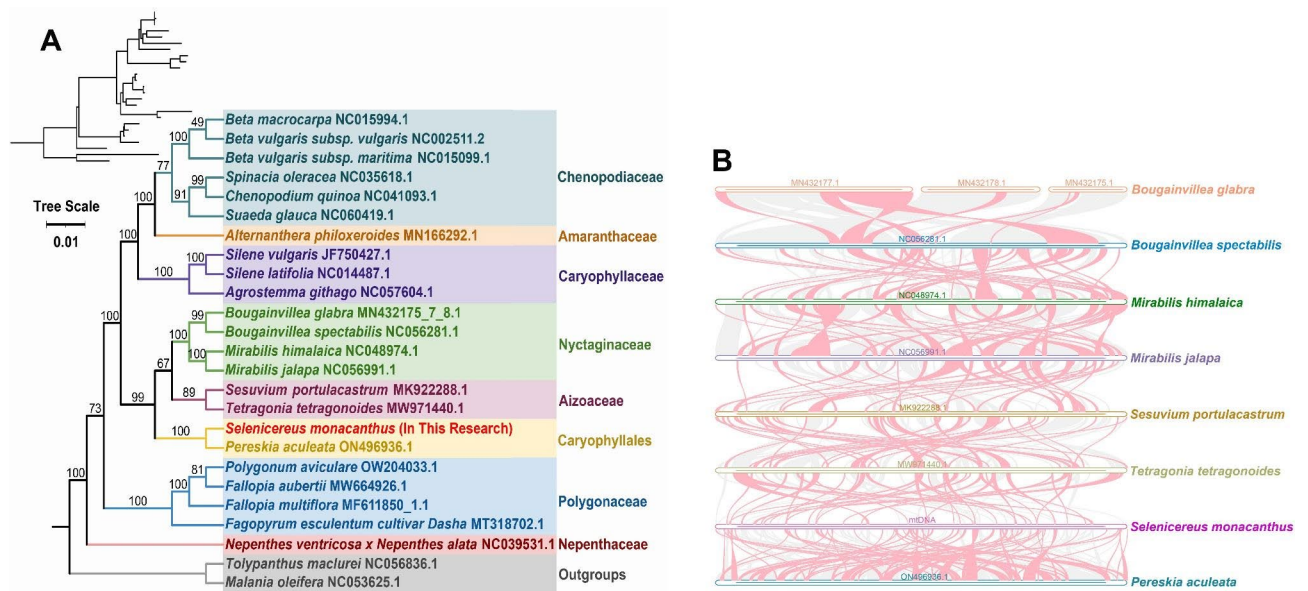
Repeat Name	Repeat 1	Repeat 2	Repeat 3
Identities (%)	100	100	100
Length (bp)	394,588	124,827	13,437
Position-1	653,265-1,047,853	1,082,361-1,207,188	763,029-776,466
Position-2	1,895,431-2,290,019	1,578,808-1,703,635	1,847,779-1,861,216
E-value	0	0	0
Type	Forward	Forward	Palindromic



**Fig. 4** Editing number in PCGs (A), frequency distribution (B), and site-specific validation (C) of RNA-editing sites in *S. monacanthus*



**Fig. 5** Chloroplast genome (A) and DNA transfer analysis (B) of *S. monacanthus*. The light blue and light green arcs represent the mitogenome and cp. genome, respectively. The purple lines between the arcs correspond to the homologous genomic fragments



**Fig. 6** Phylogenetic analysis (A) and collinearity analysis (B) between *S. monacanthus* and related genera. (A) Bootstrap support values with different colors indicate different orders. (B) The pink, blue, dark green, dark purple, orange, gray, purple, and light blue lines represent *Bougainvillea glabra*, *Bougainvillea spectabilis*, *Mirabilis himalaica*, *Mirabilis jalapa*, *Sesuvium portulacastrum*, *Tetragonia tetragonoides*, *Selenicereus monacanthus*, and *Pereskia aculeata*, respectively. The red curved areas indicate regions where inversions occur and the gray areas indicate regions of good homology

**RNA editing events are prevalent in the PCGs of the *S. monacanthus* mitogenome**

RNA editing is an important post-transcriptional regulatory mechanism and a biological process prevalent in

higher plant mitochondria where single-base conversion is the most prevalent RNA editing event [47, 48]. Our study showed that all 32 PCGs of the *S. monacanthus* mitogenome underwent RNA editing events. Each



event was a single-base edit (C to U) that mostly resulted in amino acid conversion. This may allow the genes to acquire new structures and functions. RNA editing is closely related to cytoplasmic male sterility. Stop codon editing shortens the *orf77* chimeric open reading frame associated with male sterility in maize and eventually leads to pollen abortion [49]. Meanwhile, the change from CGA (Arg) to UGA (End) at position 223 of *atp9* ensured normal synthesis of this polypeptide in the Yunnan purple rice maintenance line (YingxiangB) [50], whereas no RNA editing occurred at this site in the sterile line (YingxiangA). Plant mitochondrial RNA editing can introduce new start codons. For example, the conversion of ACG (Thr) to AUG (Met) is the starting point for the transcription of the *nad1* gene in wheat and the *cox1* gene in tomato and potato [51, 52]. Usually after generating new start and stop codons, it encodes proteins that are more conserved, and higher homology with corresponding proteins from other species allows for better expression of genes in mitochondria [47]. This study further revealed that the start or stop codons of four genes were generated by RNA editing events in the *S. monacanthus* mitogenome, i.e., new start codons by loci *nad1-2* and *nad4L-2*, while new stop codons by *atp9-copy3-223* and *ccmFC-1309*. However, their effects on mitochondria and plants require further investigation.

#### Gene transfer and gene loss are common during *S. monacanthus* evolution

Plant mitogenomes can integrate exogenous or migratory DNA sequences by intracellular or horizontal transfer [13, 20]. This leads to the accumulation of large amounts of repetitive sequences and the gain/loss of large DNA fragments in the genome [8, 53]. The introduced genes usually degenerate into pseudogenes [54]. This study found 78 homologous fragments of the chloroplast genome with a total length of 46,496 bp in the *S. monacanthus* mitogenome. The same phenomenon was observed in *Mangifera indica* [55], *Taraxacum mongolicum* [56], and *P. aculeata* [38]. However, no opposite sequence migration was observed in *S. monacanthus*. Fourteen PCGs and eleven tRNAs were identified among the homologous sequences of *S. monacanthus*, and most of the remaining gene sequences lost their integrity. Evolutionary analysis and comparison revealed that only 19 PCGs were identical among 25 closely related species, and the genome sequences of *S. monacanthus* and seven closely related species of the same Order were highly inconsistent in terms of genome sequence, even for the more closely related *P. aculeata* of the same family. This suggested that the species may have undergone frequent genome recombination events during evolution. This study further confirms the idea of mitogenomic gene transfer or loss and provides an effective way to deeply

explore the evolutionary history of *S. monacanthus* and closely related species.

#### Conclusions

This is the first published assembly of the *S. monacanthus* mitogenome, which is 2,290,019 bp in length. It encoded 59 unique genes that accounted for only 2.17% of the total length. Several dispersed repeats, plastid DNA fragments, and RNA editing events were identified in this genome, and multiple potential conformations may exist since the three repeats mediate recombination. Evolutionary analysis suggested that multiple genomic recombination and gene loss events may have occurred in *S. monacanthus* during its evolution. This study provided important information for an in-depth study of the evolutionary history and molecular breeding of *S. monacanthus*. Further, the genome of *S. monacanthus* can also be used as a reference genome for other *Selenicereus* species.

#### Materials and methods

##### Plant material and sequencing

The pitaya plant (Hong long 1) was cultivated at the National Agricultural Science and Technology Park in Lhasa, Tibet Autonomous Region, China (location: 91°2'8"E, 29°38'15"N; altitude: 3650 m). The young shoots were harvested, immediately frozen in liquid nitrogen, and stored at -80 °C in an ultra-low temperature refrigerator (Qingdao Aucma Co., Ltd, Qingdao, China). DNA and RNA were extracted from the epidermal tissue of pitaya shoots using the TianGen Super Plant Genomic DNA Kit and the RNAprep Pure Plant Kit (Polysaccharides & Polyphenolics-rich) (Beijing, China), respectively. The quality of the DNA and RNA was checked using a NanoDrop One Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA) and sent to Wuhan Benagen Tech Solutions Co., Ltd. (Wuhan, China) for sequencing. Short-reads, long-reads, and long non-coding RNA (lncRNAs) were sequenced using a DNBSEQ-T7 Genetic Sequencer (Shenzhen Huada Intelligent Technology Co., Ltd., Shenzhen, China), Nanopore PromethION sequencer (Oxford, UK), and MGISEQ-2000 sequencing platform (Shenzhen, China), respectively. Fastp v0.21.0 [57], NanoFilt v2.8.0 [58], and SOAPnuke v2.0 [59] were used to filter short-, long-, and lncRNA raw reads, respectively.

##### Mitogenome assembly

The assembly of long reads from the sequencing data was performed using Flye software [60] to obtain graphical results in GFA format [61]. Subsequently, the BLASTN program was used to identify contig fragments containing the mitogenome with the parameter “-evaluate 1e-5 -outfmt 6 -max\_hsps 10 -word\_size 7 -task blastn-short,” using the *Arabidopsis thaliana* genome as a query

sequence. The short- and long-read data were then compared to the mitogenome contigs using BWA v0.7.17 [62], and the well-matched reads were filtered and exported for subsequent assembly. Finally, the hybrid assembly was implemented to obtain the complete mitogenome of *S. monacanthus* using Unicycler v0.4.7 (The University of Melbourne, Victoria, Australia) with the parameter “--kmers 57,67” [63].

#### Gene annotation and codon preference analysis

The protein-coding genes (PCGs) of the *S. monacanthus* mitogenome were annotated using Geseq v2.03 (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) [64] with the mitogenomes of *A. thaliana* (NC\_037304) and *L. tulipifera* (NC\_021152.1) used as references. tRNAscan-SE v2.0.11 was used to annotate tRNA genes [65] and BLASTN v2.13.0 was used for rRNA gene annotation [66]. The errors were manually corrected using Apollo v1.11.8 [67]. The PCGs were extracted using PhyloSuite v1.2.2 [68] and used for codon preference analysis using Mega v7.0.26, with relative synonymous codon usage (RSCU) values calculated [69]. An RSCU value > 1 indicates that the codon is preferentially used by amino acids, whereas an RSCU value < 1 indicates the opposite trend.

#### Repeat element identification

The SSRs in the *S. monacanthus* mitogenome were identified using MISA v2.1 (<https://webblast.ipk-gatersleben.de/misa/>) [70] with the parameter “1–10 2–5 3–4 4–3 5–3 6–3”. Tandem repeats were recognized using TRF v4.09 (<https://tandem.bu.edu/trf/trf.unix.help.html>) with the parameter “2 7 7 80 10 50 500 -f -d -m” [71]. Dispersed repeats were detected using REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) [72] with the repeat size  $\geq 50$  bp. The results were visualized using Excel 2021 and Circos 0.69-9 [73].

#### Repeat-mediated recombination validation

Unicycler was used to derive the sequences at the branching nodes and map them to long reads; those supported by longer reads were prioritized. The correctness of the assembly was verified by extracting each pair of repetitive sequences and using the 500 bps upstream and downstream of the sequence as a reference. We then designed primers for the four paths of the double bifurcating structure using Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>) (Table S8), and the authenticity of the interface sequences was verified by PCR amplification and Sanger sequencing [12, 56]. The amplification was performed using an Applied Biosystems real-time PCR instrument (Thermo Fisher Scientific, Massachusetts, USA) in a total volume of 50  $\mu$ L, including 2  $\mu$ L of DNA template, 2  $\mu$ L each of upstream and downstream primer (10  $\mu$ mol/L), 25  $\mu$ L of 2 $\times$  Rapid Taq Master Mix

(Vazyme Biotech Co., Ltd., Nanjing, China), and 19  $\mu$ L of ddH<sub>2</sub>O. The cycling procedure included pre-denaturation at 95 °C for 3 min, followed by 35 cycles of 95 °C for 15 s (denaturation), 55 °C for 15 s (annealing), and 72 °C for 30 s (extension), with a final extension at 72 °C for 15 min.

#### RNA editing site prediction and validation

The transcripts from the *S. monacanthus* mitogenome were obtained from transcriptomic data by filtering, mapping to mitochondrial DNA sequences using TopHat2 with mismatches of 7 [74], and further comparison of DNA and RNA sequences using REDtools v2.0 [75] to identify the potential RNA editing events in mitogenome PCGs, with a coverage depth  $\geq 100\times$  and editing frequency  $\geq 0.10$ . Primers for specific editing sites were designed using Primer-BLAST software (Table S9). RNA was reverse transcribed into cDNA using a HiScript III 1st Strand cDNA Synthesis Kit (Vazyme, Nanjing, China). PCR amplification was performed using gDNA and cDNA as templates, and the validation method is the same as in the above section. The amplified products were compared by Sanger sequencing.

#### Homologous DNA analysis

The GetOrganelle v1.7.7.0 software [76] was used to extend the short reads of *S. monacanthus* chloroplast genome, the SPAdes software in Unicycler was used to assemble the extended reads to form a unitig map with the parameters “-R 15 -k 21,45,65,85,105 -F embplant\_pt”; and the long reads were utilized to solve the bifurcation structure in the unitig graph using Unicycler. Annotation was performed using CPGAVAS2 (<http://47.96.249.172:16019/analyzer/annotate>) [77] and the results were corrected using CPGView [78]. Homologous fragments of the chloroplast and mitochondrial genomes of *S. monacanthus* were analyzed using BLASTN [66] with an e-value of  $1e-6$  and a word size of 7.

#### Evolution analysis

The mitogenomes of twenty-four species closely related to *S. monacanthus* were downloaded from the NCBI (Table S10), with *Malania oleifera* (NC\_053625.1) and *Tolypanthus maclurei* (NC\_056836.1) (MK431827.1) set as outgroups. PhyloSuite software was used to extract the common genes [68], with MAFFT v7.505 used for multiple sequence alignment [79]. Phylogenetic analysis was performed using IQ-TREE v1.6.12 with the “GTR+I+R2” model [80], and the maximum likelihood tree was visualized using iTOL v6 (<https://itol.embl.de/>). The mitogenomes of *S. monacanthus* and seven closely related species in the same Order (Caryophyllales) were compared and analyzed using the BLAST program. Homologous sequences  $\geq 500$  bp in length were

retained as conserved co-linear blocks, and the Multiple Synteny Plot was plotted using the source program of MCscanX [81].

#### Abbreviations

PCG	protein coding gene
RSCU	relative synonymous codon usage values
SSR	simple sequence repeat
MTPT	mitochondrial plastid sequence
lncRNA	long non-coding RNA

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04529-9>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5  
Supplementary Material 6  
Supplementary Material 7  
Supplementary Material 8  
Supplementary Material 9  
Supplementary Material 10  
Supplementary Material 11

#### Authors' contributions

G.L. and Y.Q. conceived and designed the research. G.L. performed the bioinformatic analysis. G.L. and J.M. performed the experiments. G.L. wrote the initial draft. Q.L. and Y.Q. reviewed and edited the manuscript. W.W. provided the materials. All authors read and approved the final manuscript.

#### Funding

This research was supported by the Science and Technology Program of Tibet Autonomous Region, China (XZ202301ZY0020N) and the Scientific Research Start-up Fund for High-level Introduced Talents of Henan Institute of Science and Technology.

#### Data Availability

The mitogenome sequence data of *S. monacanthus* are available in NCBI Nucleotide Database under the GenBank accessions: OQ835513. The BGI and Nanopore sequencing data of *S. monacanthus* have been deposited in the Figshare platform: doi:<https://doi.org/10.6084/m9.figshare.22350940> and doi:<https://doi.org/10.6084/m9.figshare.22350505>. The raw transcriptome sequencing data of *S. monacanthus* have been submitted to the Sequence Read Archive (SRA) repository under SRR24044980. The mapping results (BAM files) were uploaded to the figshare platform: doi:<https://doi.org/10.6084/m9.figshare.22650259>.

#### Declarations

##### Ethical approval and consent to participate

This study's material collections and experimental research complied with relevant institutional, national, and international guidelines and legislation. No specific permissions or licenses were required.

##### Consent for publication

Not applicable.

#### Competing interests

The authors declare no conflict of interest.

Received: 13 August 2023 / Accepted: 16 October 2023

Published online: 04 November 2023

#### References

- Møller IM, Rasmusson AG, Van Aken O. Plant mitochondria - past, present and future. *Plant J*. 2021;108:912–59.
- Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. Archaea and the origin of eukaryotes. *Nat Rev Microbiol*. 2018;16:120.
- Fan L, Wu D, Goremykin V, Xiao J, Xu Y, Garg S, et al. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat Ecol Evol*. 2020;4:1213–9.
- Mccauley DE. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytol*. 2013;200:966–77.
- Sodmergen Q. Cytological evidence for preservation of mitochondrial and plastid DNA in the mature generative cells of *Chlorophytum* spp. (Liliaceae). *Protoplasma*. 2003;221:211–6.
- Calderon CI, Yandell BS, Havey MJ. Genetic mapping of paternal sorting of mitochondria in cucumber. *Theor Appl Genet*. 2012;125:11–8.
- Nakamura S. Paternal inheritance of mitochondria in *Chlamydomonas*. *J Plant Res*. 2010;123:163–70.
- Wu Z, Liao X, Zhang X, Tembrock LR, Broz A. Genomic architectural variation of plant mitochondria—A review of multichromosomal structuring. *J Syst Evol*. 2022;60:160–8.
- Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum Scurruoloideum* is extremely divergent and dynamic and has lost all *nad* genes. *Proc Natl Acad Sci USA*. 2015;112:E3515–3524.
- Putintseva YA, Bondar EI, Simonov EP, Sharov VV, Oreshkova NV, Kuzmin DA, et al. Siberian larch (*Larix sibirica* Ledeb) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genomics*. 2020;21:654.
- Fan W, Liu F, Jia Q, Du H, Chen W, Ruan J, et al. *Fragaria* mitogenomes evolve rapidly in structure but slowly in sequence and incur frequent multinucleotide mutations mediated by microinversions. *New Phytol*. 2022;236:745–59.
- Zhang S, Wang J, He W, Kan S, Liao X, Jordan DR, et al. Variation in mitogenome structural conformation in wild and cultivated lineages of sorghum corresponds with domestication history and plastome evolution. *BMC Plant Biol*. 2023;23:91.
- Garcia LE, Edera AA, Palmer JD, Sato H, Sanchez-Puerta MV. Horizontal gene transfers dominate the functional mitochondrial gene space of a holoparasitic plant. *New Phytol*. 2021;229:1701–14.
- Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. *G3 (Bethesda)*. 2019;9:549–59.
- Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A. RNA editing in plants and its evolution. *Annu Rev Genet*. 2013;47:335–52.
- Li Y, Zhang M, Shahzad K, Han Y, Zhang X, Guo L, et al. Development of mitochondrial simple sequence repeat markers to simultaneously distinguish cytoplasmic male sterile sources in cotton. *Funct Integr Genomics*. 2022;23:25.
- Zhao J, Shen J, Li H, Lou Q, Li J, Chen J-F. Detection of the paternal mitochondrial inheritance in *Cucumis*. *Acta Horticult Sin*. 2014;41:2250–8.
- Nong Q, Zhang M, Jiao Z, Cheng H, Zhang M, Jian S, et al. Cloning and potential application analysis of mitochondrial cytochrome *b* gene in dragon fruit. *Mol Plant Breeding*. 2019;17:4194–203.
- Barreto P, Koltun A, Nonato J, Yassitepe J, Maia IG, Arruda P. Metabolism and signaling of plant mitochondria in adaptation to environmental stresses. *Int J Mol Sci*. 2022;23:11176.
- Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet*. 2008;9:605–18.
- Liberatore KL, Dukowic-Schulze S, Miller ME, Chen C, Kianian SF. The role of mitochondria in plant development and stress tolerance. *Free Radical Bio Med*. 2016;100:238–56.
- Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. DNA repair and the stability of the plant mitochondrial genome. *Int J Mol Sci*. 2020;21:328.

23. Sloan DB, Müller K, McCauley DE, Taylor DR, Štorchová H. Intraspecific variation in mitochondrial genome sequence, structure, and gene content in *Silene vulgaris*, an angiosperm with pervasive cytoplasmic male sterility. *New Phytol.* 2012;196:1228–39.
24. Wang N, Li C, Kuang L, Wu X, Xie K, Zhu A, et al. Pan-mitogenomics reveals the genetic basis of cytonuclear conflicts in citrus hybridization, domestication, and diversification. *Proc Natl Acad Sci USA.* 2022;119:e2206076119.
25. Lisch D. How important are transposons for plant evolution? *Nat Rev Genet.* 2013;14:49–61.
26. Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science.* 2008;319:1527–30.
27. Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, et al. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat Commun.* 2019;10:1494.
28. Li C, Tang J, Hu Z, Wang J, Yu T, Yi H, Cao M. A novel maize dwarf mutant generated by *Ty1-copia* LTR-retrotransposon insertion in *Brachytic2* after spaceflight. *Plant Cell Rep.* 2020;39:393–408.
29. Hunt DR. *Selenicereus Monacanthus* (Lem) DR Hunt. *Cactaceae Syst Init.* 2017;36:33.
30. Korotkova N, Borsch T, Arias S. A phylogenetic framework for the Hylocereeae (Cactaceae) and implications for the circumscription of the genera. *Phyto-taxa.* 2017;327:1–46.
31. Ibrahim SRM, Mohamed GA, Khedr AIM, Zayed MF, El-Kholy AAS. Genus *Hylocereus*: beneficial phytochemicals, nutritional importance, and biological relevance—A review. *J Food Biochem.* 2018;42:e12491.
32. De Araújo FF, de Paulo Farias D, Neri-Numa IA, Pastore GM. Underutilized plants of the Cactaceae family: nutritional aspects and technological applications. *Food Chem.* 2021;362:130196.
33. Carreón-Hidalgo JP, Franco-Vásquez DC, Gómez-Linton DR, Pérez-Flores LJ. Betalain plant sources, biosynthesis, extraction, stability enhancement methods, bioactivity, and applications. *Food Res Int.* 2022;151:110821.
34. Trivellini A, Lucchesini M, Ferrante A, Massa D, Orlando M, Incrocci L, Mensuali-Sodi A. Pitaya, an attractive alternative crop for Mediterranean region. *Agronomy.* 2020;10:1065.
35. Pan L, Fu J, Zhang R, Qin Y, Lu F, Jia L, et al. Genetic diversity among germplasms of pitaya based on SSR markers. *Sci Hortic-amsterdam.* 2017;225:171–6.
36. Chen J, Xie F, Cui Y, Chen C, Lu W, Hu X, et al. A chromosome-scale genome sequence of pitaya (*Hylocereus undatus*) provides novel insights into the genome evolution and regulation of betalain biosynthesis. *Hortic Res.* 2021;8:164.
37. Zhang Z, Xing Y, Ramakrishnan M, Chen C, Xie F, Hua Q, et al. Transcriptomics-based identification and characterization of genes related to sugar metabolism in 'Hongshuijing' pitaya. *Hortic Plant J.* 2022;8:450–60.
38. Zhang X, Shan Y, Li J, Qin Q, Yu J, Deng H. Assembly of the complete mitochondrial genome of *Pereskia aculeata* revealed that two pairs of repetitive elements mediated the recombination of the genome. *Int J Mol Sci.* 2023;24:8366.
39. Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. The fossilized mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* 2013;11:29.
40. Cole LW, Guo W, Mower JP, Palmer JD. High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. *Mol Biol Evol.* 2018;35:2773–85.
41. Mower JP, Sloan DB, Alverson AJ. Plant Mitochondrial Genome Diversity: The Genomics Revolution. In: *Plant Genome Diversity Volume 1*. Edited by Wendel J, Greilhuber J, Dolezel J, Leitch I. Publisher: Springer, Vienna. 2012. [https://doi.org/10.1007/978-3-7091-1130-7\\_9](https://doi.org/10.1007/978-3-7091-1130-7_9).
42. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol.* 2003;29:380–95.
43. Trávníček P, Čertner M, Ponert J, Chumová Z, Jersáková J, Suda J. Diversity in genome size and GC content shows adaptive potential in orchids and is closely linked to partial endoreplication, plant life-history traits and climatic conditions. *New Phytol.* 2019;224:1642–56.
44. Bi C, Qu Y, Hou J, Wu K, Ye N, Yin T. Deciphering the multi-chromosomal mitochondrial genome of *Populus simonii*. *Front Plant Sci.* 2022;13:914635.
45. Shcherban AB. Repetitive DNA sequences in plant genomes. *Russ J Genet Appl Res.* 2015;5:159–67.
46. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol.* 2014;65:505–30.
47. Edera AA, Gandini CL, Sanchez-Puerta MV. Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. *Plant Mol Biol.* 2018;97:215–31.
48. Hao W, Liu G, Wang W, Shen W, Zhao Y, Sun J, et al. RNA editing and its roles in plant organelles. *Front Genet.* 2021;12:757109.
49. Gallagher LJ, Betz SK, Chase CD. Mitochondrial RNA editing truncates a chimeric open reading frame associated with S male-sterility in maize. *Curr Genet.* 2002;42:179–84.
50. Wei L, Yan Z, Yu J, Ding Y. Mitochondrial RNA editing of ATPase *atp9* gene transcripts of Yunnan purple rice (*Oryza sativa* L.). *J Wuhan Bot Res.* 2008;26:567–72.
51. Kadowaki K, Ozawa K, Kazama S, Kubo N, Akihama T. Creation of an initiation codon by RNA editing in the *coxI* transcript from tomato mitochondria. *Curr Genet.* 1995;28:415–22.
52. Quiñones V, Zanlungo S, Holuigue L, Litvak S, Jordana X. The *coxI* initiation codon is created by RNA editing in potato mitochondria. *Plant Physiol.* 1995;108:1327–8.
53. Kubo T, Newton KJ. Angiosperm mitochondrial genomes and mutations. *Mitochondrion.* 2008;8:5–14.
54. Mower JP, Stefanović S, Hao W, Gummow JS, Jain K, Ahmed D, Palmer JD. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol.* 2010;8:150.
55. Niu Y, Gao C, Liu J. Complete mitochondrial genomes of three *Mangifera* species, their genomic structure and gene transfer from chloroplast genomes. *BMC Genomics.* 2022;23:147.
56. Jiang M, Ni Y, Li J, Liu C. Characterisation of the complete mitochondrial genome of *Taraxacum mongolicum* revealed five repeat-mediated recombinations. *Plant Cell Rep.* 2023;42:775–89.
57. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–90.
58. De Coster W, D'hert S, Schultz DT, Cruets M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* 2018;34:2666–9.
59. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience.* 2018;7:gix120.
60. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37:540–6.
61. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics.* 2015;31:3350–2.
62. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2010;26:589–95.
63. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* 2017;13:e1005595.
64. 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–W11.
65. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 20: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 2021;49:9077–96.
66. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res.* 2015;43:7762–8.
67. Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, Richter J, et al. Apollo: a sequence annotation editor. *Genome Biol.* 2002;3:RESEARCH0082.
68. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li W, Wang G. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 2020;20:348–55.
69. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 70 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
70. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33:2583–5.
71. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27:573–80.
72. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 2001;29:4633–42.

73. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics*. 2013;14:244.
74. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.
75. Flati T, Gioiosa S, Spallanzani N, Tagliaferri I, Diroma MA, Pesole G, et al. HPC-REDtools: a novel HPC-aware tool for improved large scale RNA-editing analysis. *BMC Bioinformatics*. 2020;21:353.
76. Jin J, Yu W, Yang J, Song Y, dePamphilis CW, Yi T, Li D. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol*. 2020;21:241.
77. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res*. 2019;47:W65–W73.
78. Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. CPGView: a package for visualizing detailed chloroplast genome structures. *Mol Ecol Resour*. 2023;23:694–704.
79. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2019;20:1160–6.
80. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37:1530–4.
81. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40:e49.

### Publisher's Note

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