Acupuncture and Herbal Medicine

OPEN

Characterization of *Rheum palmatum* mitochondrial genome and comparative analysis among Caryophyllales species

Longlong Gao[1], Lijun Hao[1], Wenjie Xu[1], Tianyi Xin[1], Chi Song[2], Yulin Lin[1], Lingping Zhu[2,*], Jingyuan Song[1,*]

Abstract

Objective: This work aimed to report the first complete mitochondrial genome (mitogenome) of *Rheum palmatum*, summarize the features of Caryophyllales mitogenomes, and to reveal the potential of utilizing the mitogenomes of *R. palmatum* and other Caryophyllales species for inferring phylogenetic relationships and species identification.

Methods: Both Illumina short reads and PacBio HiFi reads were utilized to obtain a complete mitogenome of *R. palmatum*. A variety of bioinformatics tools were employed to characterize the *R. palmatum* mitogenome, compare the reported mitogenomes in Caryophyllales and conduct phylogenetic analysis.

Results: The mitogenome of *R. palmatum* was assembled into a single master circle of 302,993 bp, encoding 35 known protein-coding genes, 18 transfer RNA genes, and three ribosome RNA genes. A total of 249 long repeats and 49 simple sequence repeats were identified in this mitogenome. The sizes of mitogenomes in Caryophyllales varied from 253 kb to 11.3 Mb. Among them, 23 mitogenomes were circular molecules, one was linear, and one consisted of relaxed circles, linear molecules, and supercoiled DNA. Out of the total mitogenomes, 11 were single-chromosome structure, whereas the remaining 14 were multi-chromosomal organizations. The phylogenetic analysis is consistent with both the Engler system (1964) and the Angiosperm Phylogeny Group III system.

Conclusions: We obtained the first mitogenome of *R. palmatum*, which consists of a master circle. Mitogenomes in Caryophyllales have variable genome sizes and structures even within the same species. Circular molecules are still the dominant pattern in Caryophyllales. Single-chromosome mitogenomes account for nearly a half of all the mitogenomes in Caryophyllales, in contrast to previous studies. It is feasible to utilize mitochondrial genomes for inferring phylogenetic relationships and conducting species identification.

Keywords: Caryophyllales, Mitochondrial genome, Molecular identification, Phylogenetic analysis, *Rheum palmatum*

Graphical abstract: http://links.lww.com/AHM/A71.

[1] *Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People's Republic of China, Engineering Research Center of Chinese Medicine Resource of Ministry of Education, Institute of Medicinal Plant Development, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China;* [2] *Institute of Herbgenomics, State Key Laboratory of Southwest Chinese Medicine Resources, Pharmacy College, Chengdu University of Traditional Chinese Medicine, Chengdu, China*

**Corresponding authors. Jingyuan Song, Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People's Republic of China, Engineering Research Center of Chinese Medicine Resource of Ministry of Education, Institute of Medicinal Plant Development, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100193, China, E-mail: jysong@implad.ac.cn; Lingping Zhu, Institute of Herbgenomics, State Key Laboratory of Southwest Chinese Medicine Resources, Pharmacy College, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China, E-mail: zhulingping@cdutcm.edu.cn.*

Acupuncture and Herbal Medicine (2023) 3:4

Received 16 June 2023 / Accepted 13 September 2023

http://dx.doi.org/10.1097/HM9.0000000000000078

Introduction

Rheum palmatum L. belonging to the order Caryophyllales is a Chinese endemic plant exhibiting a broad distribution in China[1]. As one of the main sources of Rhei Radix et Rhizoma (or da-huang), it has been widely used as a drastic purgative in traditional Chinese medicine[2–4]. Scientific studies have also discovered lots of beneficial properties of *R. palmatum*. As reported, *R. palmatum* contains a variety of bioactive components such as aloe emodin, chrysophanol, rhein, emodin, and physcion[5], which have been shown to exhibit significant anti-inflammatory[6–8], anti-viral[5,9] and anti-tumor activities[10–12].

The order Caryophyllales consist of approximately 12,500 species encompassing tropical trees to temperate annual herbs, accounting for roughly 6% of the total diversity of flowering plant species[13]. The high diversity of growth patterns and ecological adaptations renders Caryophyllales a prime candidate for exploring the evolution of genes and genomes[14].

Mitochondrion is a crucial organelle in eukaryotic cells, serving as the powerhouses of the cell and orchestrating lots of essential processes. Its indispensable significance not only lies in producing energy to fuel a variety of essential cellular activities including growth, metabolism, and movement but also extends to types

of important metabolic pathways, signaling cascades, cell apoptosis, and genetic inheritance of eukaryotic cells[15–16]. The crucial role of mitochondria in cell vitality dictates that even minor changes at the genetic level may exert significant influences on the speciation and phenotypic traits of organisms. Therefore, the characterization and comparison of mitochondrial genomes (mitogenome) are highly significant and also hold great potential for revealing evolutionary relationships and the molecular identification of various organisms. Due to the rapid advancement of sequencing techniques, numerous mitogenomes of medicinal plants have been released[17–20]. They not only deepened our understanding toward the basic characteristics of the mitogenomes of medicinal plants[21–22] but also provided useful information for evolution history[23–24] and molecular identification research[25]. However, despite abundant phylogenetic and molecular identification studies based on transcriptomes or chloroplast genomes in Caryophyllales[13–14,26], studies utilizing mitogenomes remain scarce, with the limited number of published mitogenomes being a significant contributing factor. So far, only around 20 mitogenome assemblies of species in Caryophyllales have been released, among which only three mitogenomes in Polygonaceae and one mitogenome in *Rheum* have been reported. In particular, mitogenomes for the three Chinese official genuine sources of Rhei Radix et Rhizoma, namely *R. palmatum*, *Rheum tanguticum*, and *Rheum officinale* have not been reported. Among these sources, *R. palmatum* is the most commonly used[27]. Therefore, our study focused on obtaining the complete mitogenome of *R. palmatum* to shed light on its unique mitogenome characteristics and comparing the reported mitogenomes in Caryophyllales to identify their common features. Additionally, this study further explored the feasibility of using the mitochondrial genomes of *R. palmatum* and other Caryophyllales species for inferring phylogenetic relationships and species identification. The findings from this research will provide valuable reference information for effectively utilizing the resources of *R. palmatum* and other species in Caryophyllales.

Materials and methods

Materials

The fresh leaves of *R. palmatum* were collected from Lixian, Gansu, China. The *R. palmatum* plant used in this work was identified by Professor Yulin Lin from the Institute of Medicinal Plant Development, Peking Union Medical College and Chinese Academy of Medical Sciences. A total of 25 mitogenomes for 20 species in Caryophyllales (Table 1), and three mitogenomes for *Actinidia latifolia*, *Saurauia tristyla* and *Saussurea inversa* were obtained from NCBI database (https://www.ncbi.nlm.nih.gov/).

Methods

DNA extraction, sequencing, and assembly

The total genomic DNA were extracted from the fresh leaves of *R. palmatum* using a modified cetyl-trimethylammonium-bromide (CTAB) protocol[28]. The extracted genomic DNA was then made into SMRTbell

Table 1.

Sizes and structures of reported mitogenomes in Caryophyllales

Family	Species	Sequencing platform	Chromosome numbers	Chrromosome structures	Total size, bp	Accession number
Polygonaceae	*Rheum palmatum*	P, I	1	Circular	302,993	OR148905
	Rheum rhabarbarum	I	3	Circular	326,804	OQ754376~8
	Fallopia multiflora	I	2	Circular	312,450	MF611850~1
	Fagopyrum esculentum cultivar Dasha	P	10	Circular	404,063	MT318701~10
Caryophyllaceae	*Agrostemma githago*	R	1	Circular	262,903	NC_057604
	Silene vulgaris-assembly KRA	R	5	Various*	404,739	MH455602~6
	Silene vulgaris-assembly S9L	R	7	Circular	421,903	JQ771310~6
	Silene vulgaris-assembly KOV	R	6	Circular	361,139	JQ771300~5
	Silene vulgaris-assembly MTV	R	4	Circular	428,959	JQ771306~9
	Silene vulgaris-assembly SD2	R	4	Circular	427 138	JF750427~30
	Silene latifolia	R, I	1	Circular	253,413	NC_014487
	Silene noctiflora	R	63	Circular	7,137,666	KP053825~87
	Silene noctiflora	R	59	Circular	6,727,869	JF750431~89
	Silene conica	R	>128	Circular	11,318,000	JF750490~629
Nyctaginaceae	*Mirabilis jalapa*	I	1	Circular	267,334	NC_056991
	Mirabilis himalaica	I	1	Circular	346,363	NC_048974
	Bougainvillea spectabilis	I	1	Circular	343,746	NC_056281
	Bougainvillea glabra	P	3	Circular	322,663	MN432175 MN432177~8
Amaranthaceae	*Chenopodium quinoa*	P, I	1	Circular	315,003	NC_041093
	Spinacia oleracea	I	1	Circular	329,613	NC_035618
	Chenopodium album	P	1	Linear	312,951	OX419235
	Amaranthus tricolor	P	3	Circular	382,432	OP177683~5
	Alternanthera philoxeroides	I	1	Circular	283,258	MN166292
Tamaricaceae	*Myricaria elegans*	P, I	2	Circular	416,354	OP429117~8
Aizoaceae	*Tetragonia tetragonioides*	I	1	Circular	347,227	MW971440

I: Illumina; P: PacBio; R: Roche/454.

*The mitogenome of *Silene vulgaris* KRA *consists of* circles, linear molecules, relaxed circles, and supercoiled DNA.

library using Express Template Prep Kit 2.0 from PacBio according to the manufacturer's protocol and subsequently sequenced on a Sequel II (Pacific Biosciences, USA). The same batch of DNA was used to construct Illumina sequencing libraries and subsequently sequenced on an Novaseq 6000 (Illumina, USA) with 150 base pair (bp) paired-end reads length. To improve assembly reliability, we used two strategies to assemble the *R. palmatum* mitogenome. In the first strategy, the short clean reads were *de novo* assembled with GetOrganelle v1.6.4[29], potential mitochondrial contigs were extracted by aligning against the mitochondrial protein-coding genes (PCGs) from plant mitogenome database (ftp://ftp.ncbi.nlm.nih.gov/refseq/release/mitochondrion/) with BLAST v2.8.1 (National Library of Medicine, USA) + [30]. Then, the putative long mitochondrial reads were baited by mapping the PacBio long reads to the potential mitochondrial contigs using BLASR v5.1 (Pacific Biosciences, USA)[31]. Finally, the putative long mitochondrial reads were assembled by Canu v2.1.1 (Maryland Bioinformatics Labs, USA)[32]. In the second strategy, all PacBio long reads were assembled *de novo* by Canu directly. Subsequently, we used Burrows-Wheeler Alignment tool (BWA)[33] to map the short clean reads to the draft contigs and improved the draft contigs with Pilon v1.22 (Broad Institute of MIT and Harvard, USA)[34]. Then, MUMmer 3.23 (Maryland Bioinformatics Labs, USA)[35] was used to check whether these contigs were circular. Finally, the corrected contigs obtained from the above two assembly strategies were aligned with each other using MUMmer.

Genome annotation

The mitochondrion genes were annotated using the on-line GeSeq[36] tool with default parameters to predict PCGs, transfer RNA (tRNA) genes, and ribosome RNA (rRNA) genes. The position of each coding gene was determined by using BLAST searches against reference mitochondrial genes. Manual corrections of genes for start/stop codons and for intron/exon boundaries were performed in SnapGene Viewer by aligning against the reference mitogenome. The mitogenome map of *R. palmatum* was drawn by using the OrganellarGenomeDRAW (OGDRAW)[37] tool. Functional annotations were performed by using sequence-similarity Blast searches with a typical cutoff E-value of 10^{-5} against several publicly available protein databases[30]: NCBI non-redundant (Nr) protein database, Swiss-Prot, Clusters of Orthologous Groups, Kyoto Encyclopedia of Genes and Genomes, and Gene Ontology terms.

Codon usage analysis

We utilized the cusp software EMBOSS v6.6.0.0 (European Bioinformatics Institute, UK)[38] to analyze the codon usage of *R. palmatum* mitogenome. By calculating Relative Synonymous Codon Usage (RSCU), we obtained codon preference values. A codon's RSCU value of 1.00 indicates no preference; an RSCU value greater than 1.00 signifies relatively high frequency usage, while an RSCU value less than 1.00 indicates relatively low-frequency usage.

Repeats identification

Forward, reverse, palindromic, and complementary repeat sequences were determined by using the REPuter software (https://bibiserv.cebitec.unibielefeld.de/reputer/). The parameters were configured as follows for the long repeat sequence analysis: the Minimal Repeat Size was set to 30 bp, the Hamming distance was set to 3, and the Maximum Computed Repeats was set to 5,000 (equivalent to a significance threshold of 1e-3). The identification of simple sequence repeats (SSRs) was performed by using the on-line Microsatellite identification (MISA) tool (https://webblast.ipk- gatersleben.de/misa/). The parameters were set as follows for SSR analysis: definition (unit size, min repeats): 1-10, 2-5, 3-4, 4-3, 5-3, 6-3. Additionally, a minimum distance of 100 bp was set between two SSRs.

Phylogenetic inference

A total of 13 species from Caryophyllales, Ericales, and Asteraceae were selected to construct a phylogenetic tree, including four species in Polygonaceae (*Fallopia multiflora*, *Fagopyrum esculentum*, *Rheum rhabarbarum,* and *R. palmatum*), two species in Amaranthaceae (*Chenopodium quinoa* and *Spinacia oleracea*), Actinidiaceae (*Actinidia latifolia* and *Saurauia tristyla*), and Nyctaginaceae (*Bougainvillea spectabilis* and *Mirabilis jalapa*), and one species in Caryophyllaceae (*Agrostemma githago*), Aizoaceae (*Tetragonia tetragonoides*), and Asteraceae (*Saussurea inversa*). Among them, *Actinidia latifolia*, *Saurauia tristyla,* and *Saussurea inversa* were used as outgroups. Twenty orthologous PCGs were identified and extracted from the mitogenomes of *R. palmatum* and the other 12 species using PhyloSuite (v1.2.1)[39]. Then, the extracted nucleotide sequences were aligned using MAFFT v7.450 (Research Institute for Microbial Diseases, Japan)[40] implemented in PhyloSuite. These aligned nucleotide sequences were subsequently concatenated and then partitioned by ModelFinder[41]. We conducted a Bayesian inference (BI) analysis using MrBayes v3.2.6 (National Bioinformatics Infrastructure, Sweden)[42], employing the Markov Chain Monte Carlo method for 500,000 generations. Trees were sampled every 100 generations. The initial 25% of trees were discarded as burn-in, whereas the remaining trees were used to generate a consensus tree.

Results

Characteristics of the mitogenome of R. palmatum

Based on Illumina short reads and PacBio HiFi long reads, the mitogenome of *R. palmatum* was assembled into a master circle of 302,993 bp with a GC content of 45.23% (Figure 1). The base composition of the mitogenome was A (27.25%), T (27.52%), C (22.58%), and G (22.65%). The mitogenome of *R. palmatum* contained 35 known PCGs, 18 tRNA genes and 3 rRNA genes (Table 2). PCGs constituted 10.19% (30,885 bp) of the total mitogenome length, with an average gene length of 882 bp. Among 18 tRNA genes, *trnM*-CAU was found quintuplicated, and *trnH*-GUG was duplicated. All three rRNA genes of *R. palmatum* mitogenome were duplicated as well.

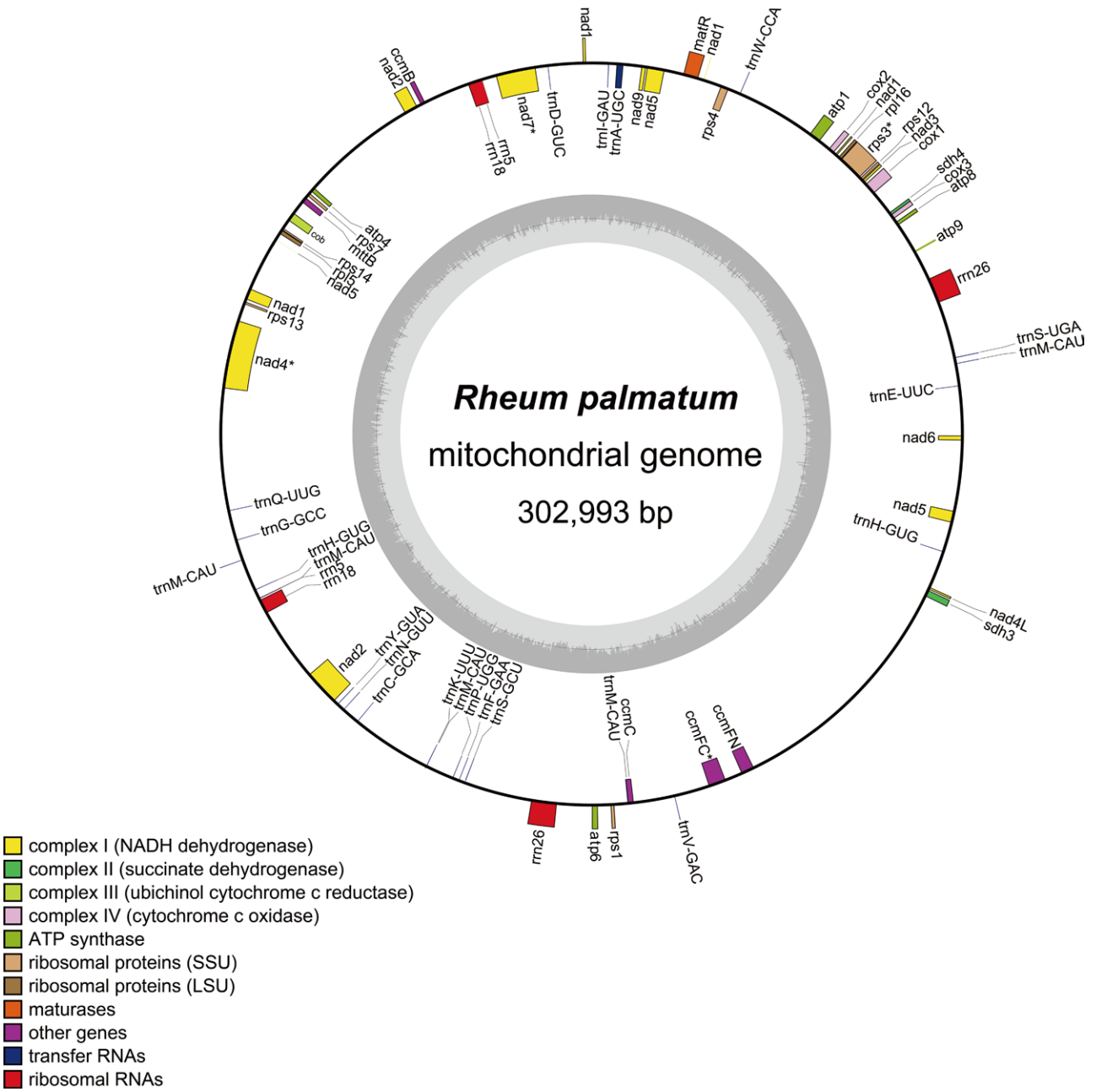

Figure 1. The mitochondrial genome map of *R. palmatum*. Clockwise transcription is observed for genes positioned on the outer side of the circle, whereas counter-clockwise transcription is observed for genes positioned on the inner side.

We computed the codon usage of PCGs in the *R. palmatum* mitogenome based on RSCU values. The results revealed that the mitogenome of *R. palmatum* contained a total of 64 codons, with 61 codons encoding 20 amino acids, and the remaining three serving as stop codons. Among all amino acid codons, those encoding Arginine (Arg, R), Leucine (Leu, L), and Serine (Ser, S) were the most abundant, whereas codons encoding Methionine (Met, M) and Tryptophan (Trp, W) were the least frequent. There were 28 codons with RSCU values > 1, of which 27 ended with A/T bases. Additionally, there were 31 codons with RSCU values < 1, with 28 of them ending with G/C bases. These findings suggested a preference for A/U-ending codons over G/C-ending codons in the *R. palmatum* mitogenome. Codons encoding Met and Trp each had only one form, with an RSCU value of 1, indicating no usage preference (Figure 2).

SSRs, also known as microsatellite sequences, are widely present in the mitogenome. In the current study, we identified 49 SSRs from *R. palmatum* mitogenome, including mono-, di-, tri-, tetra-, and pentanucleotides with numbers of 17, 11, 6, 13, and 2, respectively (Figure 3A). No hexanucleotide was detected. Mononucleotides were the most abundant type of SSRs in the *R. palmatum* mitogenome, among which A/T repeats accounted for 94.1% of all the mononucleotides. In all types of SSRs, A and T were always the most commonly used bases. Repeats with a length of 30 bp or more are called long repeats. Long repeats are important components mediating homologous recombination in plant mitogenomes[43]. In the *R. palmatum* mitogenome, a total of 249 long repeats were identified, comprising 163 forward (or direct) repeats and 86 palindromic repeats (Figure 3B). No reverse or complement repeats were

Table 2.

Gene composition in the mitogenome of *R. palmatum*

Category	Gene type	Gene names
Electron transport and ATP synthesis	Complex I	*nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5, *nad*6, *nad*7, *nad*9
	Complex II	*sdh*3, *sdh*4
	Complex III	*cob*
	Complex IV	*cox*1, *cox*2, *cox*3
	Complex V	*atp*1, *atp*4, *atp*6, *atp*8, *atp*9
	Cytochrome c biogenesis	*ccm*B, *ccm*C, *ccm*FC, *ccm*FN
Transcription and translation	Ribosome large subunit	*rpl*5, *rpl*16
	Ribosome small subunit	*rps*1, *rps*3, *rps*4, *rps*7, *rps*12, *rps*13, *rps*14
RNA genes	Ribosomal RNAs	*rrn*18, *rrn*26, *rrn*5
	Transfer RNAs	*trn*A-UGC, *trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA, *trn*G-GCC, *trn*H-GUG, *trn*I-GAU, *trn*K-UUU, *trn*M-CAU, *trn*N-GUU, *trn*P-UGG, *trn*Q-UUG, *trn*S-GCU, *trn*S-UGA, *trn*V-GAC, *trn*W-CCA, *trn*Y-GUA
Other genes	Maturase	*mat*R
	Methyltransferase	*mtt*B

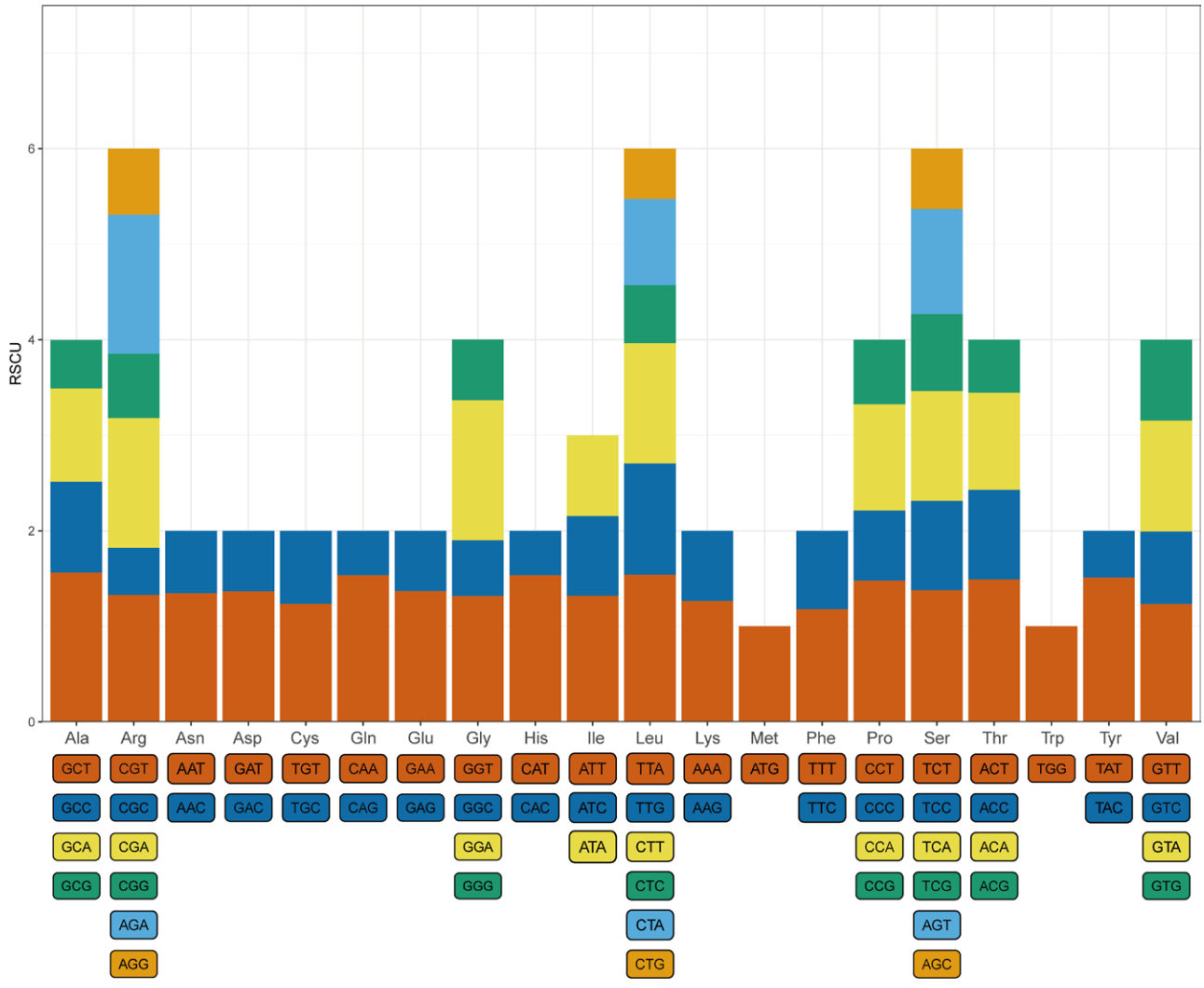

Figure 2. Codon preference map of the mitochondrial genome of *R. palmatum*. The Relative Synonymous Codon Usage (RSCU) is used to reflect the degree of codon preference. Codons are ranked below their corresponding amino acids in descending order of RSCU values.

detected. Among these repeats, there were 234 sequences with lengths less than 100 bp (Figure 3C), 13 sequences ranging from 100 to 400 bp, and two exceptionally long sequences measuring 3,344 and 8,993 bp, respectively (Figure 3D, Table 3). The length distribution map of long repeats in the *R. palmatum* mitogenome demonstrated that repeats with lengths less than 35 bp accounted for the largest part in the mitogenome.

Comparative analysis of species in Caryophyllales

Only one mitogenome assembly of *R. rhabarbarum* has been previously reported within the *Rheum* genus

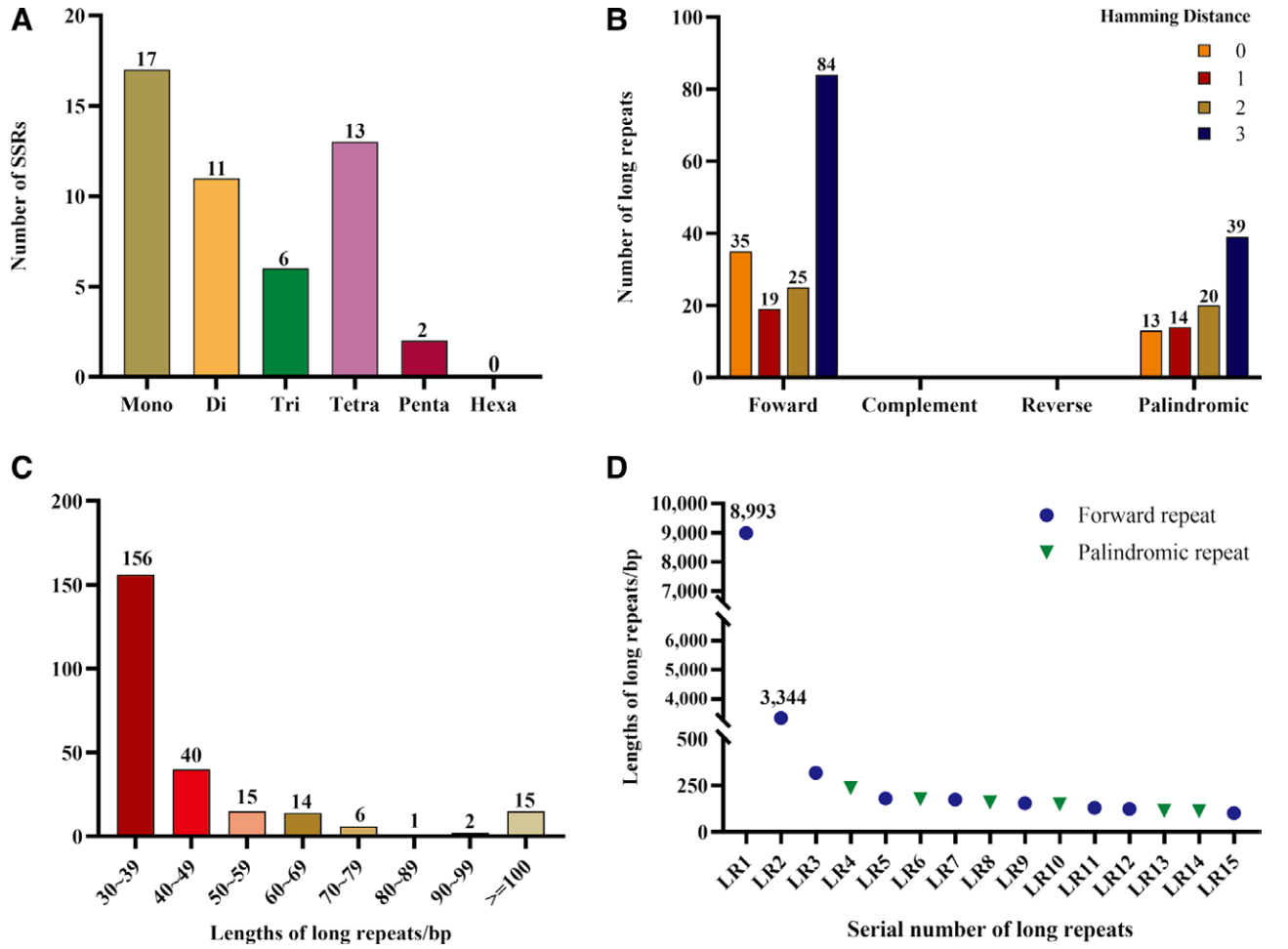

Figure 3. Distribution of different types or lengths of repeats in the mitogenome of *Rheum palmatum*. (A) Distribution of different types of SSRs. (B) Distribution of different types of long repeats. The numerical value of the Hamming distance represents the count of different bases between long repeat pairs. (C) Distribution of different lengths of long repeats. (D) Length distribution of long repeats longer than 100 bp. Di: Dinucleotide repeats; Hexa: Hexanucleotide repeats; Mono: Mononucleotide repeats; Penta: Pentanucleotide repeats; SSRs: Simple sequence repeats; Tetra: Tetranucleotide repeats; Tri: Trinucleotide repeats.

(Table 1). Compared to the mitogenome of *R. rhabarbarum* (327 kb), the *R. palmatum* mitogenome is smaller in size. Correspondingly, the mitogenome structure of *R. palmatum* (one circular molecule) is simpler than *R. rhabarbarum* (three circular molecules). Despite the structural differences, they share almost the same types of PCGs in their mitogenomes, whereas *R. rhabarbarum* exhibited more copies number of several genes including *nad*1, *nad*2, *nad*5, *rrn*5, and *sdh*4.

In the family Polygonaceae, the mitogenome assemblies of *Fal. multiflora*[44] and *Fag. esculentum*[45] have been published previously, with genome sizes of 312 and 404 kb and consisting of two and 10 chromosomes, respectively. Compared to the three previously reported mitogenomes, the *R. palmatum* mitogenome possessed the smallest size and fewest chromosomes. In addition, the four Polygonaceae species shared similar types of conserved PCGs, excepting that *Fal. Multiflora* has lost the *sdh*3 gene, and *Fag. esculentum* has lost the *sdh*3 and *rpl*16 genes. Moreover, they exhibit large variations in the copies number of many genes, including *ccm*FC, *nad*1, *nad*2, *nad*5, *nad*7, *rps*1, *rps*3, *rrn*5, *rrn*18, *rrn*26, and *sdh*4 (Table 4).

In the order Caryophyllales, the sizes of mitogenomes varied from 253 kb[46] to 11.3 Mb[47]. Among the 25 reported mitogenome assemblies for 20 species in Caryophyllales, 23 genome assemblies were circular molecules, one (*Chenopodium album*) was linear, and one (*Silene vulgaris KRA*)[48–49] consisted of a combination of circles, relaxed circles, linear molecules, and supercoiled DNA. Out of the total mitogenome assemblies, 11 were single-chromosome organization, whereas the remaining 14 were multi-chromosomal organizations. There were five assemblies reported for *Si. vulgaris*[48–51]. However, their mitogenome sizes of them varied from 361 to 429 kb. Four of the assemblies demonstrated that the mitogenome of *Si. vulgaris* was composed of several circular molecules, whereas one assembly showed that it consisted of circles, relaxed circles, linear molecules, and supercoiled DNA as mentioned above. Chromosome numbers of the five assemblies were also different, which were four, four, five, six, and seven, respectively. In addition, *Si. noctiflora* also had two mitogenome assemblies with different sizes and chromosome numbers (Table 1).

Phylogenetic inference

A well-supported phylogeny of selected mitogenomes of species in Caryophyllales was reconstructed with BI posterior probabilities equal to one for the majority

Table 3.

Distribution of repeat pairs (>100 bp) in the mitogenome of *R. palmatum*

ID	Length/bp	Type	Repeat A-start	Repeat A-end	Repeat B-start	Repeat B-end
R1	8993	F	13,415	22,407	215,327	224,319
R2	3344	F	89,928	93,271	173,564	176,907
R3	318	F	23,914	24,231	279,962	280,279
R4	236	P	115,615	115,850	237,308	237,543
R5	181	F	35,679	35,859	281,927	282,107
R6	177	P	200,413	200,589	296,042	296,218
R7	174	F	69,240	69,413	117,141	117,314
R8	158	P	115,693	115,850	237,308	237,465
R9	155	F	35,705	35,859	281,953	282,107
R10	149	P	115,702	115,850	237,308	237,456
R11	130	F	226,940	227,069	281,040	281,169
R12	124	F	41,617	41,740	280,306	280,429
R13	114	P	191,075	191,188	292,192	292,305
R14	112	P	43,340	43,451	69,201	69,312
R15	102	F	32,156	32,257	115,143	115,244

F: Forward repeat; P: palindromic repeat.

Table 4.

Comparison of protein-coding genes of mitogenomes in Polygonaceae

Names of genes	*R. palmatum*	*R. rhabarbarum*	*Fal. multiflora*	*Fag. esculentum*
*atp*1	√	√	√	√
*atp*4	√	√	√	√
*atp*6	√	√	√	√
*atp*8	√	√	√	√
*atp*9	√	√	√	√
*ccm*B	√	√	√	√
*ccm*C	√	√	√	√
*ccm*FC	√ (2)	√ (2)	√	√ (2)
*ccm*FN	√	√	√	√
cob	√s	√	√	√
*cox*1	√	√	√	√
*cox*2	√	√	√	√
*cox*3	√	√	√	√
*mat*R	√	√	√	√
*mtt*B	√	√	√	√
*nad*1	√	√ (5)	√ (2)	√ (5)
*nad*2	√	√ (2)	√	√ (5)
*nad*3	√	√	√	√
*nad*4	√	√	√	√
*nad*4L	√	√	√	√
*nad*5	√	√ (5)	√ (2)	√ (5)
*nad*6	√	√	√	√
*nad*7	√ (5)	√ (5)	√	√ (5)
*nad*9	√	√	√	√
*rpl*15	√	√	√	√
*rpl*16	√	√	√	×
*rps*1	√	√	√	√ (2)
*rps*3	√ (2)	√ (2)	√	√ (2)
*rps*4	√	√	√	√
*rps*7	√	√	√	√
*rps*12	√	√	√	√
*rps*13	√	√	√	√
*rps*14	√	√	√	√
*rrn*5	√ (2)	√ (3)	√	√
*rrn*18	√ (2)	√ (2)	√	√
*rrn*26	√ (2)	√ (2)	√	√ (2)
*sdh*3	√	√	×	×
*sdh*4	√	√ (2)	√	√

"√ (2/3/5)" means that this gene has two or three or five copies.

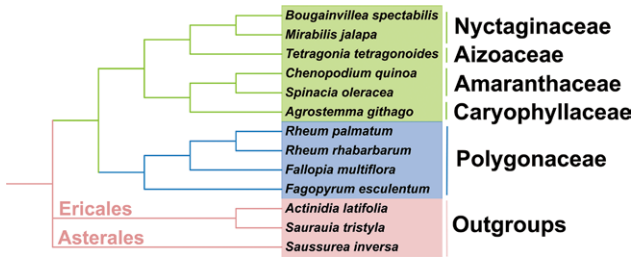

Figure 4. Phylogenetic relationships of *Rheum palmatum* with other nine plants in Caryophyllales. This tree was constructed based on the nucleotide sequences of 20 conserved mitochondrial protein-coding genes, including *atp*1, *atp*4, *atp*6, *atp*8, *atp*9, *ccm*B, *ccm*C, *ccm*FC, *ccm*FN, *cox*2, *cox*3, *nad*3, *nad*4, *nad*4L, *nad*5, *nad*6, *nad*7, *nad*9, *mat*R, and *rps*3. *Actinidia latifolia* and *Saurauia tristyla* from Ericales and *Saussurea inversa* from Asterales were used as outgroups.

of nodes. The phylogenetic tree demonstrated a clustering of *R. palmatum*, *R. rhabarbarum*, *Fag. esculentum*, and *Fal. multiflora*, all of which belonged to the family Polygonaceae. Other clades were also consistent with both the Engler system (1964) and the Angiosperm Phylogeny Group III (APG III) system[52] (Figure 4).

Discussion

In this study, we reported the first complete mitogenome of *R. palmatum*, which consists of a single master circle. The mitogenome contains a complete set of 24 highly conserved mitochondrial genes involved in electron transport and ATP synthesis[16] and seven variable genes encoding ribosomal proteins[21]. Despite the presence of two significantly large repeat pairs in the mitogenome, no recombination events were observed. This is in contrast to its closely related species, *R. rhabarbarum*, which was reported to possess a flexible mitogenome consisting of three circular chromosomes. Previous studies have shown that the lengths of repeats are positively correlated with the recombination activity[53–54]. In *R. palmatum* mitogenome, the prevalence of shorter repeats (Figure 3C) may contribute to its low recombination activity. Another notable characteristic of the *R. palmatum* mitogenome is the prevalence of A/T SSRs. A/T repeats accounts for 94.1% of all the mononucleotides

and in all types of SSRs, A and T are always the most commonly used bases. This bias toward A/T bases is consistent with the chloroplast genome of *R. palmatum*, suggesting a similar base composition preference between the two organelles[55].

The comparative analysis of mitogenomes in Caryophyllales revealed highly diverse sizes and structures. The sizes of plant mitogenomes are generally 200–700 kb[43]. Most of the reported mitogenomes in Caryophyllales fall within the range except for *Si. noctiflora* and *Si. conica*. Two reported mitogenome assemblies of *Si. noctiflora* exhibited mitogenome sizes of 7.1 and 6.7 Mb, and the size of *Si. conica* mitogenome is up to 11.3 Mb. Despite their huge mitogenomes, 99% of their genome expansion comes from intergenic sequences, and only 1% of these sequences come from chloroplasts and genomes, with the origin of the remaining sequences still a question needed to be further researched[22,47]. While recent research suggested that genome-sized circular molecules are rare and mitogenomes with various isoforms are more common, this study found that 45.8% of mitogenomes in Caryophyllales exhibit a single circular structure. It is worth noting that more than a half of these single circular mitogenomes were obtained using Illumina sequencing with relatively short read lengths, which may not have captured the complete isoform information[20–21]. Therefore, further clarification of this issue may require more long-read sequencing data, comparative analysis and experimental evidence. Additionally, despite the presence of linear, branched, and supercoiled DNA molecules observed in this study and previous research, circular structures remain dominant in the Caryophyllaceae mitogenomes[21–22].

In our study, it is demonstrated that the phylogenetic relationships established based on the mitogenome of *R. palmatum* and other nine mitogenomes in Caryophyllales were in concordance with the commonly used Engler system (1964) and APG III system, indicating the feasibility of analyzing the evolutionary relationships by using mitochondrial genomes. Besides, due to the lower evolution rate of protein-coding sequences in the plant mitogenomes compared to the chloroplast and nuclear genomes, mitogenomes were long considered unsuitable for molecular identification markers[25,56–57]. However, this perspective overlooked the substantial intergenic regions that constituted a major portion of the mitogenomes. Our study on the Caryophyllales demonstrated significant structural and size differences in the mitochondrial genomes even among closely related species within the same genus. These differences were primarily attributed to variations in the intergenic regions[22,47]. Therefore, this study suggested that exploring markers for molecular identification of Caryophyllales species within the "dark matter" of mitochondrial genomes, the intergenic regions, holds great potential. For closely related species that were challenging to differentiate due to hybridization, gene introgression, and other factors, employing the mitochondrial genome as a "super-barcode" could also be an effective approach[25].

In conclusion, this study characterized the first complete mitogenome of *R. palmatum*, provided important insights into the features of mitogenomes in Caryophyllales and revealed the feasibility of utilizing the mitochondrial genomes of *R. palmatum* and other Caryophyllales species for inferring phylogenetic relationships and species identification. These findings will serve as a valuable reference for further research on *R. palmatum* and other species in Caryophyllales.

Conflict of interest statement

The authors declare no conflict of interest.

Funding

This study was financially supported by the National Natural Science Foundation of China (81874339) and Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2022-I2M-1-018).

Author contributions

Jingyuan Song designed the research. Lingping Zhu and Chi Song provided the Illumina and PacBio sequencing data of *R. palmatum*. Yulin Lin identified the *R. palmatum* plant used in this work. Longlong Gao analyzed the data and wrote the manuscript. Lingping Zhu, Tianyi Xin, Wenjie Xu, and Jingyuan Song revised the manuscript. All authors read and approved the manuscript.

Ethical approval of studies and informed consent

Not applicable.

Acknowledgments

None.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] Wang XM, Hou XQ, Zhang YQ, et al. Distribution pattern of genuine species of rhubarb as traditional Chinese medicine. *J Med Plant Res* 2010;4(18):1865–1876.
[2] Chen D, Wang Y, Shi W, et al. Analysis of endophyte diversity of Rheum palmatum among different tissues and ages. *Arch Microbiol* 2022;205(1):14.
[3] Feng TS, Yuan ZY, Yang RQ, et al. Purgative components in rhubarbs: adrenergic receptor inhibitors linked with glucose carriers. *Fitoterapia* 2013;91:236–246.
[4] Xin T, Li R, Lou Q, et al. Application of DNA barcoding to the entire traditional Chinese medicine industrial chain: a case study of Rhei Radix et Rhizoma. *Phytomedicine* 2022;105:154375.
[5] Chang SJ, Huang SH, Lin YJ, et al. Antiviral activity of Rheum palmatum methanol extract and chrysophanol against Japanese encephalitis virus. *Arch Pharm Res* 2014;37(9):1117–1123.
[6] Wen Q, Miao J, Lau N, et al. Rhein attenuates lipopolysaccharide-primed inflammation through NF-κB inhibition in RAW264.7 cells: targeting the PPAR-γ signal pathway. *Can J Physiol Pharmacol* 2020;98(6):357–365.
[7] Shang X, Dai L, He J, et al. A high-value-added application of the stems of Rheum palmatum L. as a healthy food: the nutritional value, chemical composition, and anti-inflammatory and antioxidant activities. *Food Funct* 2022;13(9):4901–4913.

[8] Nguyen LTH, Ahn SH, Shin HM, et al. Anti-psoriatic effect of Rheum palmatum L. and its underlying molecular mechanisms. *Int J Mol Sci* 2022;23(24):16000.

[9] Chen Y, Zhu J. Anti-HBV effect of individual traditional Chinese herbal medicine *in vitro* and *in vivo*: an analytic review. *J Viral Hepat* 2013;20(7):445–452.

[10] Manimaran A, Manoharan S. Tumor preventive efficacy of emodin in 7,12-dimethylbenz[a]anthracene-induced oral carcinogenesis: a histopathological and biochemical approach. *Pathol Oncol Res* 2018;24(1):19–29.

[11] Rezapour Kalkhoran M, Kazerouni F, Omrani MD, et al. Cytotoxic effect of emodin on growth of SKBR3 breast cancer cells. Research Article. *Int J Cancer Manag* 2017;10(4): e8094.

[12] El-Saied MA, Sobeh M, Abdo W, et al. Rheum palmatum root extract inhibits hepatocellular carcinoma in rats treated with diethylnitrosamine. *J Pharm Pharmacol* 2018;70(6): 821–829.

[13] Smith SA, Brown JW, Yang Y, et al. Disparity, diversity, and duplications in the Caryophyllales. *New Phytol* 2018;217(2): 836–854.

[14] Yang Y, Moore MJ, Brockington SF, et al. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol* 2015;32(8):2001–2014.

[15] Pfanner N, Warscheid B, Wiedemann N. Mitochondrial proteins: from biogenesis to functional networks. *Nat Rev Mol Cell Biol* 2019;20(5):267–284.

[16] Møller IM, Rasmusson AG, Van Aken O. Plant mitochondria—past, present and future. *Plant J* 2021;108(4):912–959.

[17] Li J, Li J, Ma Y, et al. The complete mitochondrial genome of okra (Abelmoschus esculentus): using nanopore long reads to investigate gene transfer from chloroplast genomes and rearrangements of mitochondrial DNA molecules. *BMC Genomics* 2022;23(1):481.

[18] Li J, Xu Y, Shan Y, et al. Assembly of the complete mitochondrial genome of an endemic plant, Scutellaria tsinyunensis, revealed the existence of two conformations generated by a repeat-mediated recombination. *Planta* 2021;254(2):36.

[19] Hong Z, Liao X, Ye Y, et al. A complete mitochondrial genome for fragrant Chinese rosewood (Dalbergia odorifera, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers. *BMC Genomics* 2021;22(1):672.

[20] Gao L, Xu W, Xin T, et al. Application of third-generation sequencing to herbal genomics. *Front Plant Sci* 2023;14: 1124536.

[21] Kozik A, Rowan BA, Lavelle D, et al. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. *PLoS Genet* 2019;15(8):e1008373.

[22] Smith DR, Keeling PJ. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci U S A* 2015;112(33):10177–10184.

[23] Dong S, Chen L, Liu Y, et al. The draft mitochondrial genome of Magnolia biondii and mitochondrial phylogenomics of angiosperms. *PLoS One* 2020;15(4):e0231020.

[24] Qu XJ, Zhang XJ, Cao DL, et al. Plastid and mitochondrial phylogenomics reveal correlated substitution rate variation in Koenigia (Polygonoideae, Polygonaceae) and a reduced plastome for Koenigia delicatula including loss of all NDH genes. *Mol Phylogenet Evol* 2022;174:107544.

[25] Ślipiko M, Myszczyński K, Buczkowska K, et al. Supermitobarcoding in plant species identification? It can work! The case of leafy liverworts belonging to the genus Calypogeia. *Int J Mol Sci* 2022;23(24):15570.

[26] Kojoma M, Kurihara K, Yamada K, et al. Genetic identification of cinnamon (Cinnamomum spp.) based on the trnL-trnF chloroplast DNA. *Planta Med* 2002;68(1):94–96.

[27] Li YX, Gong XH, Li Y, et al. The influence of Aconitum carmichaelii Debx. on the pharmacokinetic characteristics of main components in Rheum palmatum L. *Phytother Res* 2015;29(8):1259–1264.

[28] Arseneau JR, Steeves R, Laflamme M. Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. *Mol Ecol Resour* 2017;17(4):686–693.

[29] Jin JJ, Yu WB, Yang JB, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* 2020;21(1):241.

[30] Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990;215(3):403–410.

[31] Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinform* 2012; 13:238.

[32] Koren S, Walenz BP, Berlin K, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 2017;27(5):722–736.

[33] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25(14):1754–1760.

[34] Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;9(11):e112963.

[35] Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes. *Genome Biol* 2004;5(2):R12.

[36] Tillich M, Lehwark P, Pellizzer T, et al. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* 2017;45(W1):W6–W11.

[37] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 2019;47(W1):W59–W64.

[38] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000;16(6): 276–277.

[39] Zhang D, Gao F, Jakovlić I, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour* 2020;20(1):348–355.

[40] Rozewicki J, Li S, Amada KM, et al. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res* 2019;47(W1):W5–W10.

[41] Kalyaanamoorthy S, Minh BQ, Wong TKF, et al. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14(6):587–589.

[42] Ronquist F, Teslenko M, van der Mark P, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 2012;61(3):539–542.

[43] Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol* 2017;68:225–252.

[44] Kim CK, Kim YK. The multipartite mitochondrial genome of Fallopia multiflora (Caryophyllales: Polygonaceae). *Mitochondrial DNA B Resour* 2018;3(1):155–156.

[45] Logacheva MD, Schelkunov MI, Fesenko AN, et al. Mitochondrial genome of Fagopyrum esculentum and the genetic diversity of extranuclear genomes in buckwheat. *Plants (Basel)* 2020;9(5):618.

[46] Sloan DB, Alverson AJ, Storchová H, et al. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm Silene latifolia. *BMC Evol Biol* 2010;10:274.

[47] Sloan DB, Alverson AJ, Chuckalovcak JP, et al. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol* 2012;10(1):e1001241.

[48] Krüger M, Abeyawardana OAJ, Juříček M, et al. Variation in plastid genomes in the gynodioecious species Silene vulgaris. *BMC Plant Biol* 2019;19(1):568.

[49] Štorchová H, Stone JD, Sloan DB, et al. Homologous recombination changes the context of Cytochrome b transcription in the mitochondrial genome of Silene vulgaris KRA. *BMC Genomics* 2018;19(1):874.

[50] Sloan DB, Müller K, McCauley DE, et al. Intraspecific variation in mitochondrial genome sequence, structure, and gene content in Silene vulgaris, an angiosperm with pervasive cytoplasmic male sterility. *New Phytol* 2012;196(4):1228–1239.

[51] Sloan DB, Alverson AJ, Wu M, et al. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus Silene. *Genome Biol Evol* 2012;4(3):294–306.

[52] GROUP TAP. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* 2009;161(2):105–121.

[53] Skippington E, Barkman TJ, Rice DW, et al. Miniaturized mitogenome of the parasitic plant Viscum scurruloideum is extremely divergent and dynamic and has lost all NAD genes. *Proc Natl Acad Sci U S A* 2015;112(27):E3515–E3524.

[54] Shi Y, Liu Y, Zhang S, et al. Assembly and comparative analysis of the complete mitochondrial genome sequence of Sophora japonica "JinhuaiJ2". *PLoS One* 2018;13(8):e0202485.
[55] Li RJ, W L, Xin TY, et al. Analysis of chloroplast genomes and development of specific DNA barcodes for identifying the original species of Rhei Radix et Rhizoma. *Acta Pharmaceutica Sinica* 2022;57(5):1495–1505.
[56] Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 1987;84(24):9054–9058.
[57] Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogenet Evol* 2008;49(3):827–831.

How to cite this article: Gao LL, Hao LJ, Xu WJ, Xin TY, Song C, Lin YL, Zhu LP, Song JY. Characterization of *Rheum palmatum* mitochondrial genome and comparative analysis among Caryophyllales species. Acupunct Herb Med 2023;3(4):323–332. doi: 10.1097/HM9.0000000000000078