BMC Plant Biology

RESEARCH

Open Access

Conserved genome structure and phylogenetic insights for the heterogeneous subfamily of Convallarioideae (Asparagaceae) revealed from plastome data

Shao-De Wu[1], Ran Meng[1], Ze-Long Nie[1], Ming-Yang Song[1], Xing-Ru Chen[1], Jun Wen[2] and Ying Meng[1*]

Abstract

Background Convallarioideae, a subfamily of Asparagaceae, encompasses a wide range of morphologically diverse lineages previously classified under different traditional families and holds significant economic value. Despite its importance, chloroplast genome data for Convallarioideae remain limited, hindering a comprehensive understanding of their genome structural evolution and phylogenetic relationships. This study aims to provide a detailed characterization of chloroplast genome features and to conduct robust phylogenetic analyses of this subfamily using an expanded dataset of chloroplast genomes.

Results The plastomes of the subfamily exhibit a typical circular quadripartite structure with conserved genomic organization and gene content. However, variations were observed in genome size, SSRs, and codon usage across the subfamily. Nine highly variable regions and positive selection genes were identified. Phylogenetic analyses based on complete plastid genomes resolved the non-monophyly of Polygonateae. Compared to *Eriospermum mackenii*, the chloroplast genomes of tribe Rusceae, tribe Dracaeneae, and the *Polygonatum-Disporopsis* lineage showed significant size reduction.

Conclusions Chloroplast genomes across Convallarioideae exhibit remarkable structural conservation. The phylogenetic analyses revealed weakly resolved backbone relationships among core members of this subfamily. Indels in the LSC region and gene loss were identified as key drivers of structural divergence in plastome size. These results clarify the interplay between genomic architecture and phylogenetic discordance, advancing our understanding of genomic evolution within Convallarioideae.

Keywords Convallarioideae, Plastid genomes, Phylogenetic, Genomic evolution

*Correspondence:
Ying Meng
mengying@jsu.edu.cn
[1] Hunan Provincial key Laboratory of Ecological Conservation and Sustainable Utilization of Wulingshan Resources, College of Biology and Environmental Sciences, Jishou University, Jishou, Hunan 416000, China
[2] Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013-7012, USA

Background

As one of the seven subfamilies of the broadly-defined family of Asparagaceae, Convallarioideae is a complex group comprising approximately 23 genera and 600 species, with a high level of morphological heterogeneity and world-wide distribution [1–3]. Previously classified under various names and ranks such as Nolinoideae, Ruscaceae sensu lato, or Convallariaceae sensu lato, it is currently recognized as Convallarioideae [2] according to the APG IV [4]. This subfamily encompasses seven morphologically distinct tribes (i.e., Convallarieae, Dracaeneae, Eriospermeae, Nolineae, Ophiopogoneae, Polygonateae, and Rusceae). Historically, these tribes were distributed across multiple families and orders in different classification systems [1, 5–7].

Except for Eriospermeae, which had consistently been resolved as sister to the rest of the subfamily with robust support, deep relationships among the remaining Convallarioideae taxa remain unresolved based on limited Sanger sequencing data [5, 6, 8]. Meng et al. [9] utilized transcriptomic data to investigate deep relationships within the subfamily, which significantly enhanced our understanding of this phylogenetically challenging group. However, the taxonomic sampling at the genus level was insufficient to adequately address the issues of tribe-level monophyly or deep relationships within the subfamily.

In all cases, previous studies used traditional or genomic molecular data to provide us with a good starting point for clarifying the evolution of Convallarioideae, but the evolutionary history of the tribes still remains unresolved and controversial. To further address these questions more extensive genomic data of Convallarioideae is needed. In recent years, the rapid advancement and decreasing costs of high-throughput sequencing technology have facilitated the acquisition and assembly of complete plastomes. The chloroplast genome provides multiple benefits, such as a distinctive inheritance pattern, a stable genomic structure, and an intermediate evolutionary rate [10].

Owing to these unique characteristics, chloroplast genomes are being widely used in plant phylogenetic studies. Ji et al. [3] extracted 68 plastid protein-coding genes from 88 complete plastomes in Asparagaceae to reconstruct the phylogeny of Convallarioideae. The results were largely consistent with previous findings, with weak support among tribes. However, the study primarily focused on phylogenetic reconstruction based on coding genes, resulting in low support values for the deep relationships within the subfamily [3]. Current information on comparative plastomes of Convallarioideae is very few, except for some from Chen et al. [11], which conducted comparative plastid genome analyses of seven species representing seven clades of the subfamily. Their results showed that all plastomes exhibited conserved quadripartite structures with 137 genes, including 87 protein-coding genes, 38 transfer RNAs and 8 ribosomal RNA genes [11].

Although some studies have utilized plastids for phylogenetic and comparative genomics analyses, researches on the chloroplast genome structure and comparative genomics analyses of Convallarioideae are very few. Therefore, in the study, we sequenced the complete chloroplast genomes of the 34 Convallarioideae species and one species of *Asparagus* L. belongs to Asparagoideae to explore the features of chloroplast structure and phylogenetic relationships among Convallarioideae species. This investigation is based on a broader sampling from all the seven tribes and the two isolated genera of *Theropogon* Maxim. and *Comospermum* Rauschert, which can improve our understanding of plastome evolution and phylogenetic relationships within Convallarioideae.

Methods

Plant materials and sequencing

A total of 34 species including 22 genera of Convallarioideae and one species (*Asparagus nelsii* Schinz) from Asparagoideae were sampled, with a complete coverage of representatives from all the seven tribes and the two unclassified genera of *Theropogon* and *Comospermum* (Table S1). Voucher specimens were desposited in the United States National Herbarium (US), Jishou University Herbarium (JIU) and Kunming Institute of Botany of Chinese Academy of Sciences Herbarium (KUN). Total DNA was obtained using the modified CTAB method [12] and DNA libraries were generated with the NEBNext Ultra DNA Kit (New England Biolabs, Ipswich, MA, USA) following the manufacturer's protocol. Prepared libraries were sequenced on the Illumina Hiseq 2000 platform using a 150 paired-end run. Trimmomatic 0.39 was used to remove adaptors and to filter low-quality reads with default parameters.

Chloroplast assembly and annotation

Chloroplast sequences were recovered from filtered Illumina sequencing reads using GetOrganelle 1.7.5 [13] with the default settings. Chloroplast genomes were first annotated using GeSeq [14]. The junctions between the large single-copy (LSC), a small single-copy (SSC), and a pair of inverted repeat (IR) regions in each plastome were examined visually and adjusted manually by comparing them to reference sequences in Geneious 9.0.2. A reference closest to each genome was obtained through BLASTing in GenBank. Circle maps were drawn using OGDRAW [15].

Comparative analyses of plastid genomes

We used REPuter [16] to detect different repeats, including forward, reverse, complementary, and palindromic. The minimum repetition was set to 30 bp and minimum repetition sequence length distance to 3. We also used MISA to predict simple sequence repeats (SSR) with parameters as follows: mononucleotide unit repetition number ≥10; dinucleotide unit repetition number ≥5; trinucleotide unit repetition number ≥4; and tetraconucleotide, pentanucleotide, and hexanucleotide unit repetition number ≥3 [17]. Indels were pairwise detected using snippy [18].

CodonW 1.4.2 was employed to estimate the amino acid usage frequency and relative synonymous codon usage (RSCU) [19]. Rearrangement analyses of the chloroplast DNA for the 34 Convallarioideae species were performed using the Mauve alignment tool within Geneious 9.0.2. To illustrate interspecific variation, the chloroplast genomes were compared using mVISTA [20] with the Shuffle-LAGAN model and the reference genome of *Eriospermum mackenii* (Hook.f.) Baker. Visualization of the contraction and expansion at the four boundaries of the plastomes was produced using CPJSdraw 1.0.0 [21]. To determine nucleic acid variations among the genomes, DnaSP 6.12.03 was utilized with a sliding window size of 600 bp.

Protein-coding genes (PCG) derived from whole chloroplast genomes were analyzed to calculate synonymous (Ks) and nonsynonymous (Ka) substitution rates using KaKs_Calculator 2.0 [22] with the genetic code table 11 (bacterial and plant plastid code) and method of MLWL calculation.

Phylogenetic analyses

Two datasets were applied for phylogenetic analyses, including the PCG and the complete chloroplast genome sequences (CGS). Multiple sequence alignment was performed using MAFFT 7.427 [23] with an auto mode. Trimmed alignments were generated using trimAl 1.4 [24]. Phylogenetic analyses were conducted using both maximum likelihood (ML) and Bayesian inference (BI) approaches. The nucleotide substitution models were tested using jModeltest 2.1.10 [25] and the best fit model was selected using the corrected Akaike Information Criterion (AICc). The ML phylogenetic tree was constructed using RAxML 8.2.10 [26], with the GTRGAMMA model and rapid bootstrap analysis (bootstrap = 1000).

The BI was performed by MrBayes 3.2.7 [27] with the Markov chain Monte Carlo run for 10,000,000 generations. The GTR + I + G model was chosen based on the Bayesian information criterion through model testing. Trees were sampled every 1000 generations. Stationarity was deemed to be achieved when the average standard deviation of split frequencies fell below 0.01. Trees generated prior to achieving likelihood and topological convergence were discarded as burn-in. The remaining trees were then used to construct a 50% majority-rule consensus tree.

Results

Annotation and structure of chloroplast genomes

The chloroplast genomes of Convallarioideae consistently exhibited a standard quadripartite structure, consisting of a LSC, a SSC, and a pair of IR regions (Fig. S1). The gene content was also identical with a total of 137 genes, including 87 PCG, 8 rRNA, and 38 tRNA genes with the exception of *Ruscus* L. (Table 1). Fifteen genes (*ndhA*, *ndhB*, *petB*, *petD*, *atpF*, *rps16*, *rpl*2, *rpl16*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC* , and *rpoC*1) contained only one intron and three genes (*rps12*, *clpP*, and *ycf3*) included two introns (Table 2). The GC content was very similar for the whole chloroplast genomes of all the species (37.5–37.8%).

The chloroplast genomes of the subfamily showed similar structures, particularly in the IR regions (Figs. 1, S2) and no rearrangement found in gene organization (Fig. S3). The LSC/IRb boundary was located between *rpl22* and *rps19* genes, and the IRa/LSC boundary in most of the Convallarioideae species was found between the *rps19* and *psbA* gene (Fig. 1). Moreover, the IRb/SSC boundary was located in the overlapping region of the *ycf1* and *ndhF* genes, and the *ycf1* and *trnN* genes were found on both sides of IRa-SSC boundary (Fig. 1).

Variation of plastome length and repeated sequences

The length variation of plastid genomes was found within the subfamily, spanning from 153,882 to 157,193 base pairs (Table 1; Fig. 2). *Maianthemum* F. H. Wigg. showed the longest average length of 156,909 bp within the genus, ranging from 156,799 to 157,115 bp. In contrast, *Ruscus* had the shortest average length of 153,986 bp, ranging from 153,882 to 154,089 bp. The length variation of the IR and SSC regions fluctuated within a range of 600 bp, whereas the LSC region showed much greater variation, exceeding 2600 bp in length. The IR regions were found to be 26,127 and 26,678 bp (Table 1; Fig. 2).

The insertions and deletions in the chloroplast genome were identified through comparison with *Eriospermum mackenii* (Fig. 2). *Maianthemum stellatum* (L.) Link genome had the longest insertion of 385 bp, and the *Dracaena* Vand. ex L. genome presented the longest deletion of 448 bp (Fig. 2).

Four types of repetitions (forward, palindromic, complement, and reverse) were found in the subfamily (Fig. S4). Palindromic units were the commonest type (1209),

Table 1 The chloroplast genomes features of 34 species in 22 genera from subfamily Convallarioideae

Taxa	Length (bp)				GC content (%)				Number of genes			
	Total	LSC	IR	SSC	LSC	IR	SSC	Overall	tRNA	rRNA	CDS	Total number of genes
Maianthemum henryi	157115	85692	26523	18377	35.5	42.9	31.6	37.6	38	8	87	137
Maianthemum lichiangense	156799	85404	26521	18353	35.6	43.0	31.5	37.6	38	8	87	137
Maianthemum racemosum	156842	85515	26398	18531	35.7	43.0	31.6	37.7	38	8	87	137
Maianthemum stellatum	156882	85529	26311	18731	35.7	43.1	31.6	37.7	38	8	87	137
Beaucarnea sp.	156930	85386	26433	18678	35.6	43.0	31.3	37.6	38	8	87	137
Dasylirion longistylum	156547	84960	26453	18681	36.0	43.0	31.2	37.8	38	8	87	137
Nolina cespitifera	156836	85310	26440	18646	35.6	43.0	31.4	37.6	38	8	87	137
Liriope platyphylla	156765	85113	26489	18674	35.6	43.0	31.3	37.6	38	8	87	137
Ophiopogon planiscapus	156979	85326	26473	18707	35.7	43.0	31.4	37.7	38	8	87	137
Peliosanthes teta	156778	85444	26386	18562	35.7	43.1	31.4	37.7	38	8	87	137
Polygonatum acuminatifolium	155273	84318	26289	18351	35.8	43.0	31.7	37.7	38	8	87	137
Polygonatum odoratum	155348	84238	26282	18455	35.7	43.0	31.6	37.7	38	8	87	137
Polygonatum sibiricum	155512	84535	26281	18415	35.7	43.0	31.7	37.7	38	8	87	137
Polygonatum cirrhifolium	155533	84283	26391	18468	35.7	42.9	31.5	37.7	38	8	87	137
Polygonatum tessellatum	155698	84496	26318	18566	35.7	42.9	31.4	37.6	38	8	87	137
Polygonatum kingianum	155690	84501	26324	18541	35.7	43.0	31.6	37.7	38	8	87	137
Heteropolygonatum altelobatum	155835	84792	26199	18645	35.6	43.0	32.1	37.7	38	8	87	137
Heteropolygonatum pendulum	155943	84976	26224	18519	35.6	42.9	31.2	37.6	38	8	87	137
Disporopsis fuscopicta	156012	84993	26236	18553	35.7	43.1	31.7	37.7	38	8	87	137
Disporopsis pernyi	156141	85048	26277	18539	35.7	43.1	31.6	37.7	38	8	87	137
Aspidistra fenghuangensis	156385	85119	26516	18234	35.6	43.0	31.6	37.6	38	8	87	137
Aspidistra oblanceifolia	156374	85115	26518	18223	35.7	43.0	31.7	37.7	38	8	87	137
Tupistra urceolata	157059	85505	26512	18530	35.6	43.0	31.5	37.6	38	8	87	137
Reineckea sp.	157052	85467	26428	18729	35.6	43.0	31.5	37.6	38	8	87	137
Rohdea sp.	157193	85573	26482	18656	35.6	43.0	31.4	37.6	38	8	87	137
Speirantha gardenii	156868	85362	26436	18634	35.6	43.0	31.6	37.7	38	8	87	137
Comospermum yedoense	155953	85207	26127	18492	35.7	43.0	31.4	37.6	38	8	87	137
Theropogon pallidus	156581	85504	26436	18205	35.6	43.0	31.5	37.7	38	8	87	137
Chrysodracon forbesii	155011	83760	26477	18297	35.5	42.9	31.3	37.6	38	8	87	137
Dracaena sp.	155291	83752	26525	18489	35.6	42.9	31.1	37.6	38	8	87	137
Sansevieria aethiopica	154810	83609	26510	18181	35.4	42.9	31.2	37.5	38	8	87	137
Ruscus hypophyllum	154089	83212	26261	18355	35.9	43.1	31.5	37.8	38	8	86	136
Ruscus sp.	153882	83006	26261	18354	35.9	43.1	31.5	37.8	38	8	86	136
Eriospermum mackenii	157001	85119	26678	18526	35.7	42.9	31.5	37.7	38	8	87	137

followed by the forward (791), the reverse (453), and the complement (52). The numbers of SSR ranged from 39 to 89 (Fig. S4). Mononucleotide repeats were the most abundant (1420 repeats), followed by dinucleotides (493), tetranucleotides (293), trinucleotides (127), pentanucleotides (76), and hexanucleotide (7).

Nucleotide diversity, codon usage, and selective pressure

The single copy regions showed higher levels of variation and diversity than the IR regions, while non-coding regions demonstrated greater variability relative to coding regions (Fig. 3). Six highly variable regions were identified in non-coding sequences and three were found in coding regions (Fig. 3). The coding sequences, intergenic regions, and introns longer than 200 bp showed highly variable regions, including *trnS-GCU*, *petD*, and *infA* in PCG with nucleotide diversity of Pi > 0.03, and *psbI-trnS-GCU*, *psbE-petL*, *rps3-rpl22*, *ccsA-ndhD*, *ndhD-psaC*, and *rps15-ycf1* in non-coding regions with Pi > 0.05 (Fig. 3).

The codon number of PCG in the Convallarioideae ranged from 28,982 to 31,112, with RSCU values falling

Table 2 Gene composition of chloroplast genome of 34 species from subfamily Convallarioideae

Categroy	Group of genes	Name of genes
Photosynthesis	Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
	NADH dehydrogenase	*ndhA*[a], *ndhB*[a](×2), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Cytochrome b/f complex	*petA*, *petB*[a], *petD*[a], *petG*, *petL*, *petN*
	ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*[a], *atpH*, *atpI*
Self-replication	Ribosomal proteins (SSU)	*rps*2, *rps*3, *rps*4, *rps*7(×2), *rps*8, *rps*11, *rps*12[b](×2), *rps*14, *rps*15, *rps*16[a], *rps*18, *rps*19(×2)
	Ribosomal proteins (LSU)	*rpl*2[a](×2), *rpl*14, *rpl*16[a], *rpl20*, *rpl*22, *rpl*23(×2), *rpl*32, *rpl*33, *rpl*36
	Ribosomal RNAs	*rrn4.5*(×2), *rrn*5(×2), *rrn*16(×2), *rrn*23(×2)
	Transfer RNAs	*trnA-UGC*[a](×2), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG-UCC*[a](×2), *trnH-GUG*(×2), *trnI-CAU*(×2)/*GAU*[a](×2), *trnK-UUU*[a], *trnL-CAA*(×2)/*UAA*[a]/*UAG*, *trnM-CAU*(×2), *trnN-GUU*(×2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*(×2)/*UCU*, *trnS-GCU*/*GGA*/*UGA*, *trnT-GGU*/*UGU*(×2), *trnV-GAC*(×2)/*UAC*[a], *trnW-CCA*, *trnY-GUA*
Other genes	DNA-dependent RNA polymerase	*rpoA*, *rpoB*, *rpoC1*[a], *rpoC2*
	Maturase	*matK*
	Protease	*clpP*[b]
	Envelope membrane protein	*cemA*
	Subunit acetyl-CoA-carboxylase	*accD*
	c-Type cytochrome synthesis gene	*ccsA*
	Subunit of Rubisco	*rbcL*
	Translation initiation factor	*infA*
Hypothetical chloroplast reading frames	Conserved open reading frames	#*ycf1*(×2), #*ycf2*(×2), #*ycf3*[b], #*ycf4*, #*ycf*15(×2), #*ycf*68(×2)

[a] Represents a gene with one intron

[b] Represents a gene with two introns

[#] Represents pseudogene

(×2) represents Number of copies of multi-copy genes

between 0.2883 (CGC) and 2.2787 (AGA). Leucine was the most frequently amino acid, comprising 99,091 codons. Conversely, tryptophan was presented as a rare amino acid with just 17,315 codons. Among the remaining codons, 31 showed biased usage, with RSCU exceeding 1 (Fig. S5).

Significant differences in Ka/Ks rates were observed, with 9 out of 79 protein genes exhibiting Ka/Ks > 1 (i.e., *infA*, *matK*, *petD*, *psbN*, *rpl14*, *rpl20*, *rpl32*, *rpoC1*, *rps16*) (Fig. 4). The *rpl14* gene in *Tupistra urceolata* N. Tanaka & W.J. Kress show the highest Ka/Ks ratio of 2.38001. The remaining 70 protein-coding genes have a Ka/Ks ratio less than one and most of them have a Ka/Ks ratio of less than 0.5 (Fig. 4).

Phylogenetic relationships within Convallarioideae

The CGS data produced similar topological relationships with well-supported bootstrap supports (BS) and posterior probability (PP) based on the BI and ML analyses (Figs. 5A, S6). All the tribes were supported to be monophyletic with the exception of Polygonateae, which was isolated into two lineages, the *Polygonatum* Mill.-*Disporopsis* Hance clade and the *Maianthemum* clade (Figs. 5A, S6). Following the first diverged lineage of tribe Eriospermeae, a woody clade including Draceaneae and Rusceae (BP=100%, PP=1.00) was placed as the second lineage of the subfamily (Figs. 5A, S6). The core clade was split into four clades (Figs. 5A, S6), corresponding to (1) *Maianthemum* clade + Nolineae + Ophiopogoneae (BS = 50%, PP = 0.93), (2) *Polygonatum-Disporopsis* clade (BS = 100 %, PP = 1.00), (3) Convallarieae + *Comospermum* (BS = 100 %, PP = 1.00), and (4) *Theropogon*.

The PCG data produced similar topological relationships based on the BI and ML analyses (Figs. 5B, S7). Both analyses revealed the monophyletic status of all the tribes except for Polygonateae with *Theropogon* nested within the tribe (Fig. 5B). Two major clades were recognized within the core subfamily: one is the woody clade including Draceaneae sister to Rusceae (BP=100%, PP=1), and the other is the core clade (BP=100%, PP=1.00). Within the latter, our analyses recovered three subclades: (1) Polygnateae + *Theropogon* (BS = 50%, PP = 0.93), (2) Convallarieae + *Comospermum* (BS = 100%, PP = 1.00), and (3) Nolineae + Ophiopogoneae (BS = 100%, PP = 1.00).

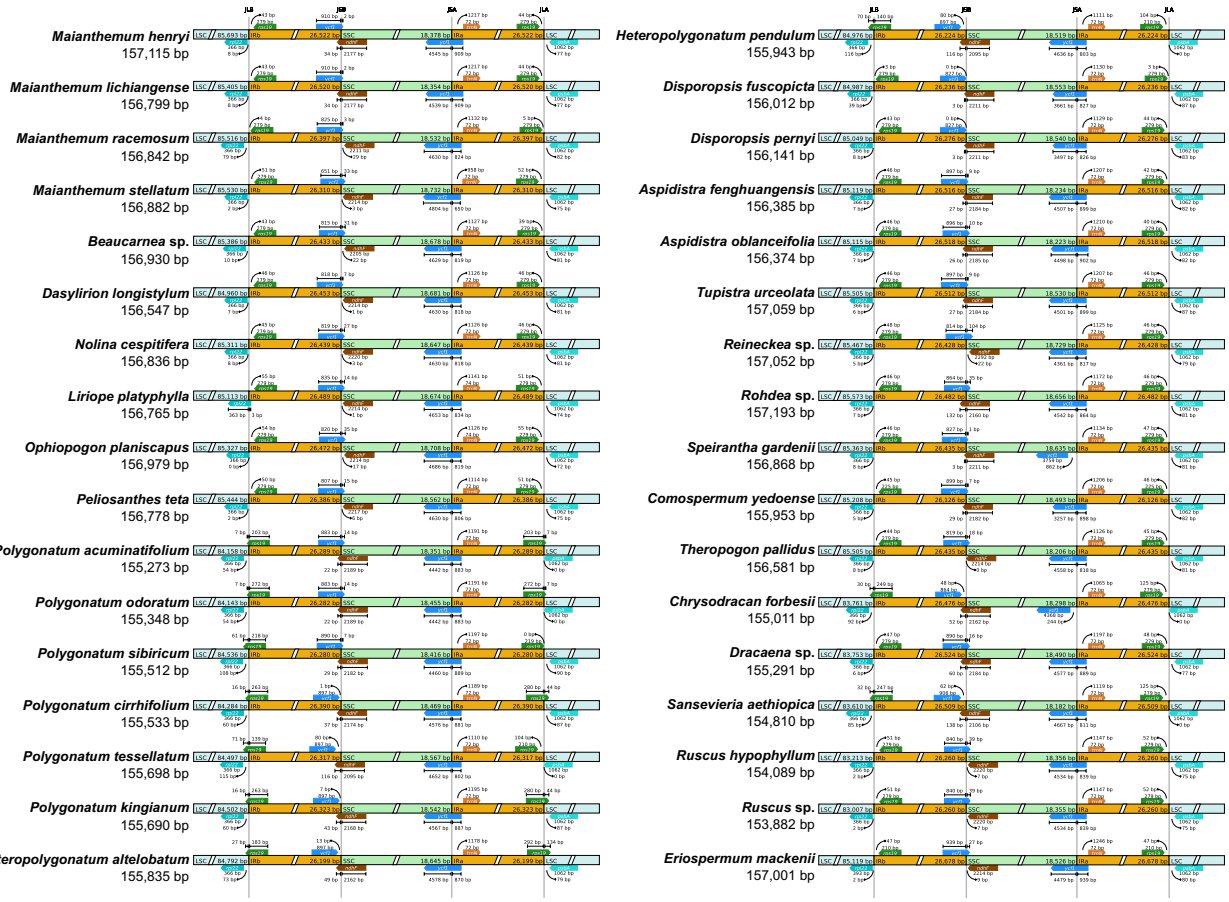

Fig. 1 Comparative analysis of gene order and junction sites of the 34 plastomes from Convallarioideae, showing connection sites of LSC, IRb, SSC, and IRa. The T bars above or below the genes indicate the extent of their parts with their corresponding values in the base pair. The plotted genes and distances in the vicinity of the junction sites are the scaled projection of the genome. JLB (IRb/LSC), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC) represent the junction sites between two adjacent regions in the genome

Discussion

Conserved chloroplast genome structure of Convallarioideae

All of the chloroplast genomes of Convallarioideae exhibited a typical quadripartite structure (Fig. S1), which were consistent with those of other Convallarioideae species [11] and other angiosperm [28]. The subfamily also showed a remarkable consistency in the number of PCG, rRNA, and tRNA genes (Table 1). The difference in GC content among chloroplast genomes was also subtle (Table 1), which may be caused by the presence of a large amount of rRNA in the IR regions [29] The chloroplast genome features in our study were in align with the findings reported by Chen et al. [11], indicating a high level of similarity in the chloroplast genomes of the subfamily.

Plastid genomes in most angiosperms are typically maternally inherited with limited recombination, thereby maintaining a highly conserved structure among closely related species [10, 30]. The junction regions of chloroplast genomes from Convallarioideae sequenced in this study had high similarity (Fig. 1) and the distribution of genes at region boundaries was highly consistent among Convallarioideae species with only minor contractions and expansions observed in the IR regions (Figs. S2, S3). Therefore, it seemed that the chloroplast genome structure of the Convallarioideae was extremely conserved, though the subfamily is composed of taxa with a high level of morphological heterogeneity and global distribution from tropical to temperate regions [3, 9].

While the chloroplast genome is typically regarded as highly conserved across flowering plants, some closely related species display regions with elevated mutation rates in their sequences [31]. Nine regions were identified with high variability (*trnS-GCU*, *petD*, *infA* in coding sequences with Pi > 0.03, and *psbI-trnS-GCU*, *psbE-petL*, *rps3-rpl22*, *ccsA-ndhD*, *ndhD-psaC*, *rps15-ycf1* in untranslated regions with Pi > 0.05) exhibiting elevated rates of variation (Fig. 3). These regions of high mutation

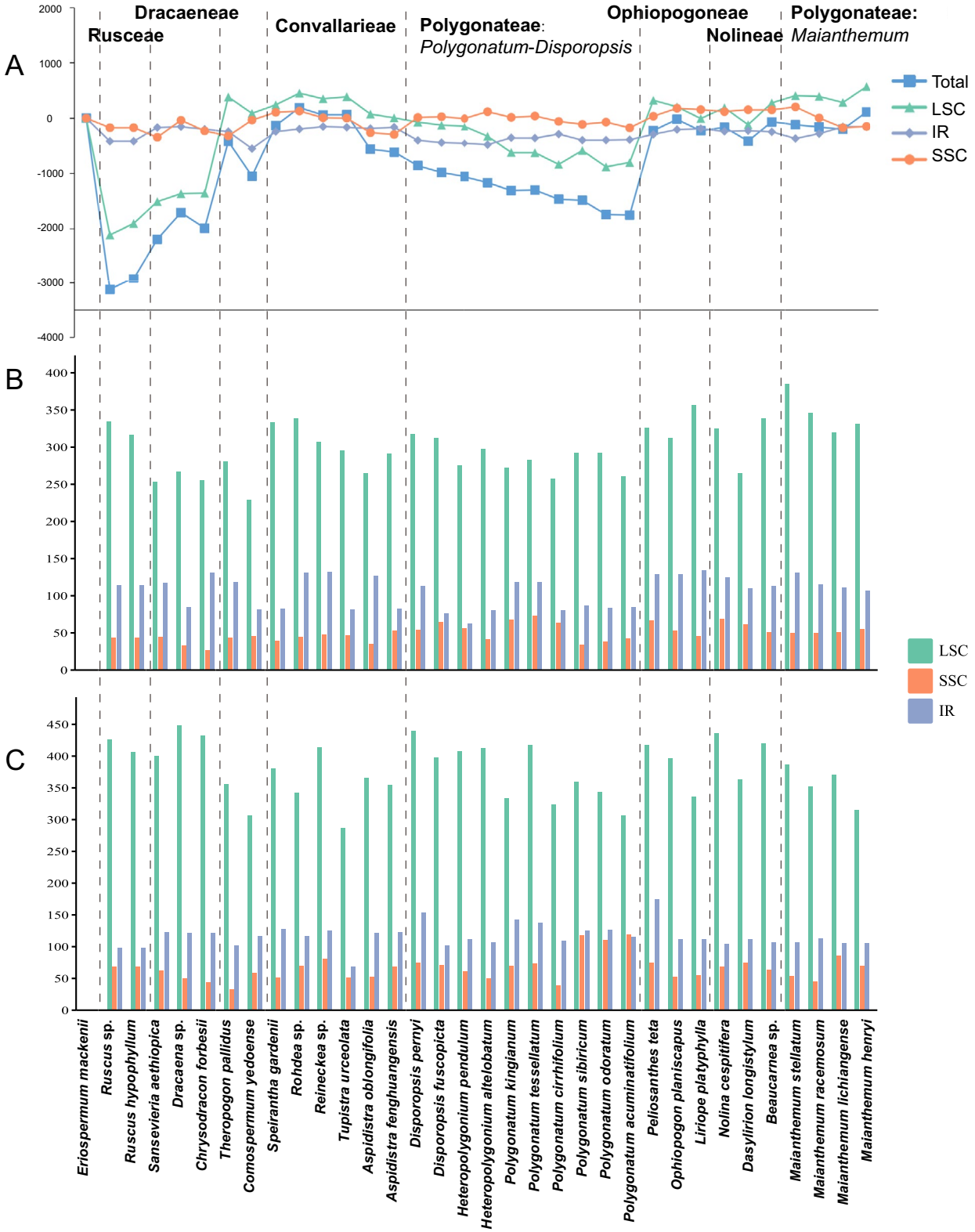

Fig. 2 Length variation of chloroplast genomes (**A**), insertions (**B**), and deletions (**C**) of the Convallarioideae taxa compared with *Eriospermum mackenii*

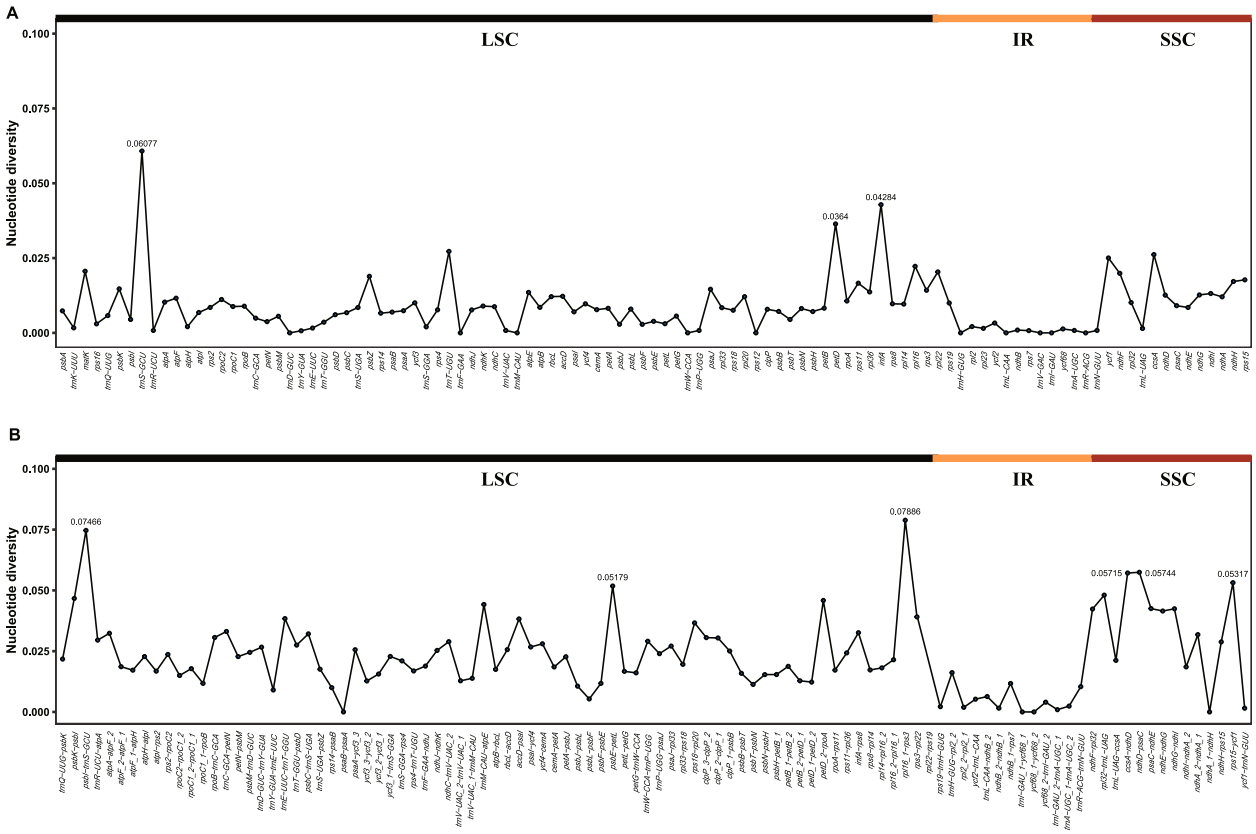

Fig. 3 Nucleotide diversity (Pi) values of different regions of 34 plastomes. **A** Pi values of PCG regions; **B** Pi values of non-coding regions

rates are frequently employed in studies related to plant evolutionary relationships, genetic diversity within populations, and DNA-based species identification.

Only 9 positive selected PCG were identified through all the Convallarioidea and there was no PCG with positive selection specific to any tribes within the subfamily (Fig. 4). These genes are crucial for a range of biological functions, including fatty acid synthesis, tRNA modification, protein degradation, protein synthesis, and photosynthesis. For instance, the *infA* and *matK* genes are involved in fatty acid synthesis and tRNA modification, respectively, contributing to chloroplast structure and function [32, 33]. The PCG under positive selection might be key factors in promoting the adaptive divergence and evolutionary diversification of Convallarioideae species.

Codon usage bias was commonly observed among various species and associated with gene expression and nucleotide composition [34]. In Convallarioideae, the majority of codons with the RSCU value > 1 terminated with A or U, whereas those ending with C or G typically having RSCU values below 1 (Fig. S5). Consistent with previous studies [35], leucine and isoleucine codons were the most frequent, while tryptophan codons were the least frequent in the Convallarioideae.

Deep relationships revealed from the chloroplast genomes

Both CGS and PCG data provided useful implications for the deep relationships of the subfamily (Fig. 5). Congruent with previous findings [3, 7, 11], the tribe Eriospermeae was consistently recognized as the first split lineage of the subfamily and the two woody tribes of Dracaeneae and Rusceae were clustered together as the second lineage of the subfamily [3, 11]. Similar results were suggested by the transcriptome data but they are grouped as a grade other than a clade [3, 9]. Dracaeneae was initially thought to share a close evolutionary relationship with Nolineae, another woody-like tribe, due to their shared traits such as linear leaves, succulent foliage on thickened woody stems formed by a secondary thickening meristem, and the presence of tenuinucellate ovules [36]. However, this overall morphological similarity is probably due to convergence in dry habit, which was not supported by either plastid nor nuclear data. In fact, leaf anatomy is distinctly different between Nolineae and Dracaeneae. In Nolineae, sclerenchyma supports extend from the majority of vascular bundles to the outer layer, whereas in Dracaeneae, vascular bundles are dispersed within the mesophyll tissue [37]. Furthermore, cytogenetic comparisons reveal greater chromosomal correspondence

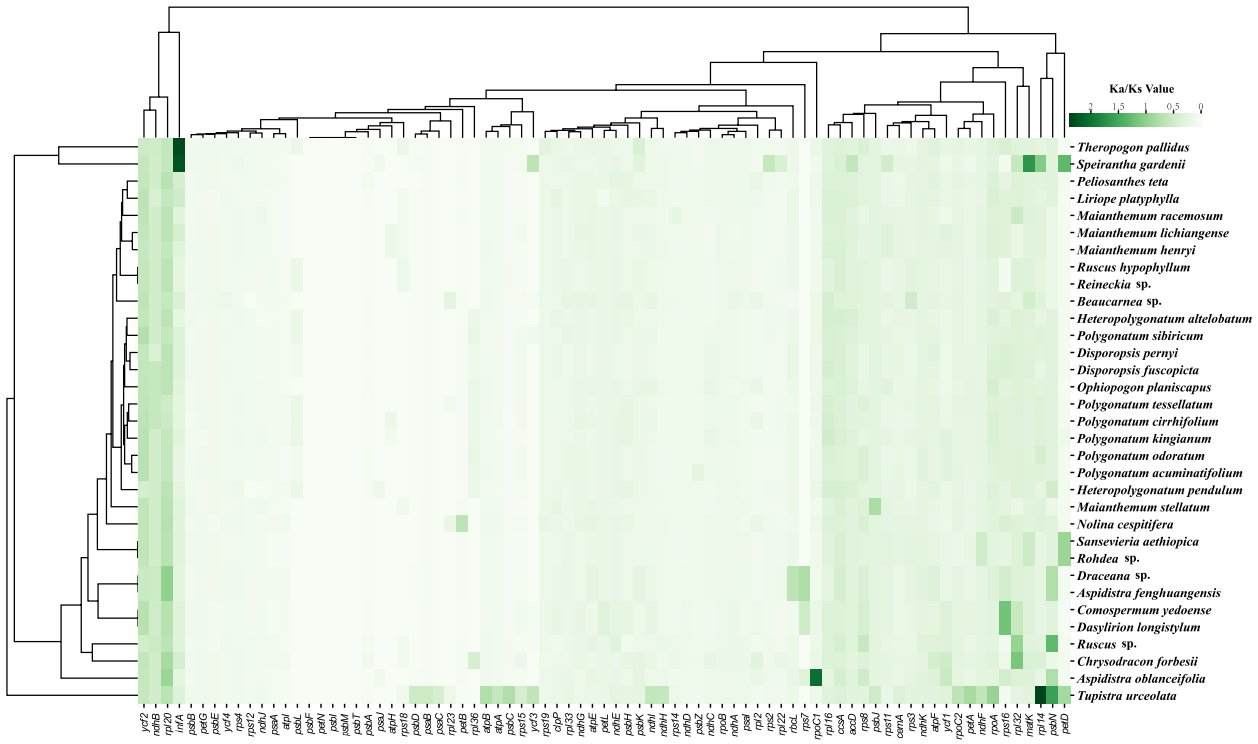

Fig. 4 Pairwise Ka/Ks ratios of 34 species from subfamily Convallarioideae in different genes with *Eriospermum mackenii* as a reference. The color closer to green represents the gene has a higher Ka/Ks ratio

between Dracaeneae and Rusceae (both x = 20) compared to Dracaeneae and Nolineae (x = 19) [5, 9].

All the other taxa, including the woody-like Nolineae, the four herbaceous tribes, and the two isolated genera, were grouped together forming the majority of the subfamily, the core clade of Convallarioideae (Fig. 5), which were also supported by previous works based on both nuclear and plastid data [3, 9, 11]. Our phylogenetic tree based on PCG yielded results largely consistent with the study published by Ji et al. [3] using the similar plastid PCG dataset. However, different phylogenetic relationships within the core clade were found from the CGS data, which is probably due to the fast divergence of the core clade with lower variation of PCG sequences. The CGS was comprised of both protein-coding and non-coding regions (such as introns and inter-genic regions) whereas PCG includes only protein-coding sequences. The non-coding regions usually had greater variability than coding regions (Fig. 4), which evolved more rapidly than coding regions and provided significant insights for phylogenetic analyses across taxa [38, 39].

Phylogenetic analyses had proposed a closer evolutionary relationship between Nolineae and the herbaceous Convallarieae rather than with lignified groups of Dracaeneae and Rusceae [3]. However, Ophiopogoneae and Nolineae clustered together, forming a clade close to Convallarieae by the PCG data, while they were grouped as a grade close to *Maianthemum* in the CGS data (Fig. 5A). The transcriptome data suggested Nolineae as the first lineage within the core clade [9]. The Nolineae comprised woody perennials displaying unusual lignification patterns, characterized by apical leaf clusters and non-splitting fruits, primarily distributed in arid North American habitats. In contrast, Ophiopogoneae members developed sympodial underground stems, precociously dehiscent fruits containing fleshy-coated seeds, and a fundamental chromosome complement of x = 18 [6]. Identifying clearly shared and derived characteristics among these taxa was proved to be challenging, though cytogenetic evidence offered potential diagnostic markers, with Nolineae typically demonstrating x = 19 while Convallarieae members predominantly having x = 18-19 [9].

Although most previous molecular studies had consistently supported the monophyly of Polygonateae [7, 9, 37, 40–42], the PCG dataset suggested the non-monophyly of the tribe with *Theropogon* nested within it, which is similar to Ji et al. [3]. However, the CGS data recognized two lineages of the tribe (Fig. 5), which was consistent with previous results based on whole plastome data [11, 43]. *Maianthemum*, which had a base chromosome number of x = 18, was sister to Nolineae with x = 19, rather

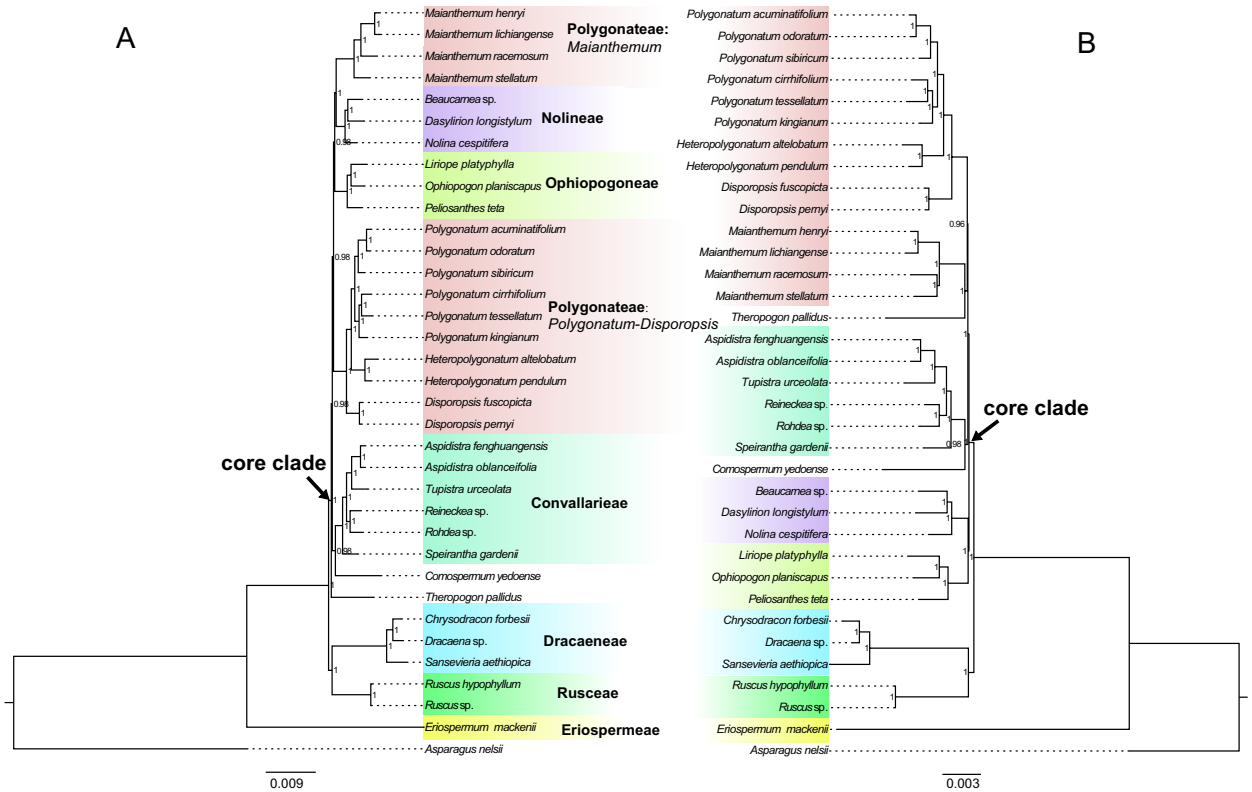

Fig. 5 Phylogenetic trees of Convallarioideae using the Bayesian inference method based on CGS (**A**) and PCG data (**B**). Numbers above branches indicate posterior probability (PP)

than being closely related to the *Polygonatum-Disporopsis* lineage with a base chromosome number ranging from x = 9 to 20 [9].

The phylogenetic position of *Theropogon* remained unclear in previous studies [6, 7, 44]. Moreover, Ji et al. [3] placed *Theropogon* within a larger clade as a sister group to *Maianthemum*, and Chen et al. [11] identified it as a sister group to Convallarieae. Our analysis suggested that *Theropogon* was the first lineage of the core clade from the CGS dataset (Fig. 5), but close to *Maianthemum* by the PCG data which was consistent with Ji et al. [3]. *Theropogon* had a unique base chromosome number (x = 20) apart from Convallarieae and Polygonateae, though it shares the characteristic of septal nectaries, a trait otherwise observed solely in Convallarieae. Additional genetic and morphological analyses are essential to unravel the phylogenetic uncertainties and clarify the evolutionary connections related to *Theropogon*.

Comospermum was one of several isolated asparagoid taxa with obscure affinities, which was traditionally ascribed to Anthericaceae [45], but embedded in a large clade of Convallariaceae sensu lato based on analysis of *rbcL* sequence [44]. The phylogenetic position of *Comospermum* within the core group remained uncertain in previous studies [6, 7, 44]. The genus was supposed to be close to Rusceae and Dracaeneae [36], which share common cytological characteristics, including tenuinucellate ovules with parietal tissue organization and a uniform x=20. It seemed to be also doubtful for a close relationship between *Comospermum* and *Eriospermum* based on early *rbcL* data [44] because their seed hair types were not homologous [46]. In our study, utilizing complete chloroplast genomic data, we found that the genus was close to Convallarieae (Fig. 5A), both of which are distributed in East Asia. This novel finding suggested it as the sister taxon to the Convallarieae, which develop a single dominant rhizome system accompanied by abbreviated stems, often producing fleshy fruits [47].

Length and structural evolution within Convallarioideae

The plastid genomes of the subfamily showed a constant size, except for the two woody tribes of Rusceae and Dracaeneae characterized with short genome length (Fig. 2A). For the herbaceous taxa, *Polygonatum-Disporopsis* lineage also had short genome size (Fig. 2A). This result indicated that variations in genome length across plastid genomes are likely associated with phylogenetic relationships.

Changes in genome length can result from gene loss or even the complete deletion of a plastid partition [48]. Our study identified the loss of the *rps16* gene in *Ruscus* (Table 1), which may contribute to its shorter genome size and was related to its adaptation as a shade-tolerant shrub. The lower photosynthetic requirements of *Ruscus* may lead to a simplification of chloroplast functions, potentially resulting in gene loss [49]. The expansion and contraction of IR was also suggested to play a crucial role in the size variation of chloroplast genomes [50]. No apparent expansion or contraction of the IR boundary was found (Fig. 1), which can not be considered as the main cause of the length variation in plastome size in the subfamily.

It was common that insertions and deletions were typically correlated with variations length of plastomes [51]. The size differences in the LSC regions of Convallarioideae plastomes mainly corresponded to variations in complete chloroplast genome size (Fig. 2), similar as other finding [52]. Moreover, genomic repetitive elements played a crucial role in facilitating the occurrence of insertion and deletion events [53]. In Gnaphalieae (Asteraceae) and Ampelopsideae (Vitaceae), the size variation of chloroplast genomes were influenced by the presence of repeat sequences and the frequency of indels [54, 55]. Our findings also indicated that the reduced chloroplast genome lengths observed in these three clades may result from an elevated frequency of deletions coupled with a reduced incidence of insertions in the LSC region (Fig. 2). Environmental factors during the evolutionary history of plants may have led to the loss of certain gene fragments [56], which could account for the observed higher frequency of deletions compared to insertions in these clades in Convallarioideae with reduced genome size. Furthermore, the Rusceae tribe demonstrates the lowest abundance of long sequence repeats among the analyzed groups, a characteristic potentially linked to its comparatively compact genome size (Fig. S4).

Conclusions

In this study, representatives of 22 out of the 23 genera currently accepted in the Convallarioideae were sampled for the comparative analyses of plastome structure and phylogenetic relationships, including *Comospermum* which was rarely sampled in previous studies. The analyzed plastid genomes maintained consistent structural organization, with preservation of genomic dimensions, gene content, sequential arrangement, and functional organization. Nine highly variable regions were identified as potential DNA barcodes for phylogeography and genetic diversity studies in Convallarioideae. Eriospermeae was consistently recognized as the first split lineage of the subfamily, followed with the two woody tribes of Dracaeneae and Rusceae as the second lineage. *Comospermum* was strongly supposed to be sister to the tribe Convallarieae, but some conflict relationships were observed for the other taxa within the core clade of Convallarioideae. Structural variations of plastid genomes provided important insights into the evolution of the subfamily. The early lineages of Rusceae and Dracaeneae showed relatively small genome size than the core group. Furthermore, the *Polygonatum-Disporopsis* lineage exhibited reduced genome length than the *Maianthemum* taxa, which support possible remote relationships of them.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06711-7.

Supplementary Material 1: Fig. S1. Structure and characteristics of the complete chloroplast genomes of Convallarioideae species. Genes inside and outside the circle are transcribed clockwise and counterclockwise separately. Darker and lighter grey in the inner circle each represent GC and AT content.

Supplementary Material 2: Fig. S2. Visualization of genome alignment of 34 Convallarioideae taxa. Eriospermum mackenii genome is shown at the top as the reference. Orientation of genes is pointed out by arrows up the alignments. Purple, blue, pink, and grey bars correspond to exons, untranslated regions, non-coding sequences, and mRNA, respectively. The Y-axis shows different species names, and sequence similarity of aligned regions is displayed as horizontal bars, which expresses as a percentage within 50%-100%.

Supplementary Material 3: Fig. S3. MAUVE alignment of 34 Convallarioideae species chloroplast genomes. The Eriospermum mackenii as the reference genome.

Supplementary Material 4: Fig. S4. Summary of sequence repeats across Convallarioideae chloroplast genomes. (A) Dispersed repeats types; (B) Base composition of SSR.

Supplementary Material 5: Fig. S5. RSCU values of the chloroplast genomes of Convallarioideae taxa. Color depth represents the Euclidean distance.

Supplementary Material 6: Fig. S6. ML phylogenetic tree of Convallarioideae based on the CGS sequences. Numbers above branches indicate supported values from bootstrap analyses.

Supplementary Material 7: Fig. S7. ML phylogenetic tree of Convallarioideae based on the PCG data. Numbers above branches indicate supported values from bootstrap analyses.

Supplementary Material 8.

Acknowledgments

The authors acknowledge the Smithsonian High Performance Cluster (SI/HPC; https://doi.org/10.25572/SIHPC) for providing computational resources for the data analyses.

Authors' contributions

This study was conceived and designed by S-D Wu and Z-L N. Collection and identification of field material were performed by J W, Z-L N and Y M. Paper writing and data analysis were performed by S-D W, X-R, R M and M-Y S. Later editions of the manuscript were contributed by S-D W, Y M and Z-L N.

Funding

This study was supported by grants from Natural Sciences Foundation of China (32160052, 31760055, 31570211) and Natural Sciences Foundation of

Hunan Province (2019JJ40232). The experimental work was conducted at the Smithsonian Laboratories of Analytical Biology of the National Museum of Natural History. The authors acknowledge the Smithsonian High Performance Cluster (SI/HPC; https://doi.org/10.25572/SIHPC) for providing computational resources for the data analyses.

Data availability

All chloroplast genomes in this study are available from the National Center for Biotechnology Information (NCBI) (accession numbers: PV087142-PV087176).

Declarations

Ethics approval and consent to participate

Experimental research and field studies on plants complies with relevant institutional, national, and international guidelines and legislation. All plant species are not rare species for protection. We have the permission to collect the plant samples by the local government. Our work reported here complies with the IUCN Policy Statement on Research Involving Species at Risk of Extinction.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 February 2025 Accepted: 13 May 2025
Published online: 27 May 2025

References

1. Chase MW, Reveal JL, Fay MF. A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. Bot J Linn Soc. 2009;161(2):132–6.
2. Tanaka N, Nguyen KS. Nolinoideae (Asparagaceae) in APG III needs replacing with Convallarioideae. Phytotaxa. 2023;583(3):297–9.
3. Ji Y, Landis JB, Yang J, Wang S, Zhou N, Luo Y, Liu H. Phylogeny and evolution of Asparagaceae subfamily Nolinoideae: new insights from plastid phylogenomics. Ann Bot. 2023;131(2):301–12.
4. Group TAP, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181(1):1–20.
5. Yamashita J, Tamura MN. Molecular phylogeny of the Convallariaceae (Asparagales). In: Wilson KL, DA M, editors. Monocots: systematics and evolution. Melbourne: CSIRO Publishing; 2000. p. 387–400.
6. Kim JH, Kim DK, Forest F, Fay MF, Chase MW. Molecular phylogenetics of Ruscaceae sensu lato and related families (Asparagales) based on plastid and nuclear DNA sequences. Ann Bot. 2010;106(5):775–90.
7. Chen S, Kim DK, Chase MW, Kim JH. Networks in a large-scale phylogenetic analysis: reconstructing evolutionary history of Asparagales (Lilianae) based on four plastid genes. PLoS ONE. 2013;8(3):e59472.
8. Meng R, Meng Y, Yang YP, Nie ZL. Phylogeny and biogeography of *Maianthemum* (Asparagaceae: Nolinoideae) revisited with emphasis on its divergence pattern in SW China. Plant Divers. 2021;43(2):93–101.
9. Meng R, Luo LY, Zhang JY, Zhang DG, Nie ZL, Meng Y. The deep evolutionary relationships of the morphologically heterogeneous Nolinoideae (Asparagaceae) revealed by transcriptome data. Front Plant Sci. 2021;11:584981.
10. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17:1–29.
11. Chen XR, Wu SD, Nie ZL, Wen J, Meng Y. Comparative analyses of complete plastid genomes from seven representative species of Convallarioideae (Asparagaceae). Phytotaxa. 2024;638(1):20–32.
12. Doyle J. DNA protocols for plants. In: Hewitt GM, Johnston AWB, Young JPW, editors. Molecular techniques in taxonomy. Berlin: Springer; 1991. p. 283–93.
13. Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:1–31.
14. 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(W1):W6–11.
15. 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):W59–64.
16. 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(22):4633–42.
17. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.
18. Taouk ML, Featherstone LA, Taiaroa G, Seemann T, Ingle DJ, Stinear TP, Wick RR. Exploring SNP filtering strategies: the influence of strict vs soft core. Microb Genomics. 2025;11(1):001346.
19. Sun X, Yang Q, Xia X. An improved implementation of effective number of codons (N c). Mol Biol. 2013;30(1):191–6.
20. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32(suppl_2):W273–9.
21. Li H, Guo Q, Xu L, Gao H, Liu L, Zhou X. CPJSdraw: analysis and visualization of junction sites of chloroplast genomes. PeerJ. 2023;11:e15326.
22. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genom Proteom Bioinform. 2010;8(1):77–80.
23. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. 2019;20(4):1160–6.
24. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
25. Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008;25(7):1253–6.
26. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–90.
27. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.
28. Bautista MAC, Zheng Y, Boufford DE, Hu Z, Deng Y, Chen T. Phylogeny and taxonomic synopsis of the genus *Bougainvillea* (Nyctaginaceae). Plants. 2022;11(13):1700.
29. Zhao W, Guo L, Yang Y, Wang Y, Yang L, Wei C, Guo J, Yan K, Chen H, Yang Z. Complete chloroplast genome sequences of *Phlomis fruticosa* and *Phlomoides strigosa* and comparative analysis of the genus *Phlomis* sensu lato (Lamiaceae). Front Plant Sci. 2022;13:1022273.
30. Liu L, Wang Y, He P, Li P, Lee J, Soltis DE, Fu C. Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. BMC Genomics. 2018;19:1–17.
31. Kim SH, Yang J, Park J, Yamada T, Maki M, Kim SC. Comparison of whole plastome sequences between thermogenic skunk cabbage Symplocarpus renifolius and nonthermogenic *S. nipponicus* (Orontioideae; Araceae) in East Asia. Int J Mol Sci. 2019;20(19):4678.
32. Hirose T, Ideue T, Wakasugi T, Sugiura M. The chloroplast *infA* gene with a functional UUG initiation codon. FEBS Lett. 1999;445(1):169–72.
33. Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76:273–97.
34. Wang Y, Wen F, Hong X, Li Z, Mi Y, Zhao B. Comparative chloroplast genome analyses of *Paraboea* (Gesneriaceae): insights into adaptive evolution and phylogenetic analysis. Front Plant Sci. 2022;13:1019831.
35. Shahzadi I, Mehmood F, Ali Z, Ahmed I, Mirza B. Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae. Genomics. 2020;112(2):1454–63.
36. Rudall PJ, Furness CA, Chase MW, Fay MF. Microsporogenesis and pollen sulcus type in Asparagales (Lilianae). Can J Bot. 1997;75(3):408–30.

37. Rudall PJ, Conran JG, Chase MW. Systematics of Ruscaceae/Convallariaceae: a combined morphological and molecular investigation. Bot J Linn Soc. 2000;134(1–2):73–92.
38. Li X, Xu H, Yu J, Han Y, Han S, Niu Y, Zhang F. Comparative and phylogenetic analysis of *Potentilla* and *Dasiphora* (Rosaceae) based on plastid genome. BMC Plant Biol. 2025;25(1):176.
39. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot. 2007;94(3):275–88.
40. Yamashita J, Tamura MN. Phylogenetic analyses and chromosome evolution in Convallarieae (Ruscaceae sensu lato), with some taxonomic treatments. J Plant Res. 2004;117:363–70.
41. Meng Y, Wen J, Nie ZL, Sun H, Yang YP. Phylogeny and biogeographic diversification of *Maianthemum* (Ruscaceae: Polygonatae). Mol Phylogenet Evol. 2008;49(2):424–34.
42. Meng Y, Nie ZL, Deng T, Wen J, Yang YP. Phylogenetics and evolution of phyllotaxy in the Solomon's seal genus *Polygonatum* (Asparagaceae: Polygonateae). Bot J Linn Soc. 2014;176(4):435–51.
43. Floden A, Schilling E. Using phylogenomics to reconstruct phylogenetic relationships within tribe Polygonateae (Asparagaceae), with a special focus on *Polygonatum*. Mol Phylogenet Evol. 2018;129:202–13.
44. Chase MW. Molecular phylogenetics of Lilianae. In: Rudall PJ, Cribb PJ, Cutler DF, CJ H, editors. Monocotyledons: systematics and evolution. London: Royal Botanic Gardens, Kew; 1995. p. 109–37.
45. Liliaceae Krause K. In: Engler A, Prantl K, editors. Die natürlichen Pflanzenfamilien. Leipzig: Engelman; 1930. p. 227–386.
46. Rudall PJ, Campbell G. Flower and pollen structure of Ruscaceae in relation to Aspidistreae and other Convallariaceae. Flora. 1999;194(2):201–14.
47. Conran J, Tamura M. Convallariaceae. In: Kubitzki K, editor. The families and genera of vascular plants. Berlin: Springer; 1998. p. 186–98.
48. Guzmán Díaz S, Núñez FAA, Veltjen E, Asselman P, Larridon I, Samain M-S. Comparison of Magnoliaceae plastomes: adding neotropical *Magnolia* to the discussion. Plants. 2022;11(3):448.
49. Mohanta TK, Mishra AK, Khan A, Hashem A, Abd Allah EF, Al-Harrasi A. Gene loss and evolution of the plastome. Genes. 2020;11(10):1133.
50. Wicke S, Naumann J. Molecular evolution of plastid genomes in parasitic flowering plants. Adv Bot Res. 2018;85:315–47.
51. Cai J, Ma PF, Li HT, Li DZ. Complete plastid genome sequencing of four *Tilia* species (Malvaceae): a comparative analysis and phylogenetic implications. PLoS ONE. 2015;10(11):e0142705.
52. Wang J, Fu GF, Tembrock LR, Liao XZ, Ge S, Wu ZQ. Mutational meltdown or controlled chain reaction: the dynamics of rapid plastome evolution in the hyperdiversity of Poaceae. J Syst Evol. 2023;61(2):328–44.
53. Bennetzen JL, Ma J, Devos KM. Mechanisms of recent genome size variation in flowering plants. Ann Bot. 2005;95(1):127–32.
54. Lao XL, Meng Y, Wu J, Wen J, Nie ZL. Plastid genomes provide insights into the phylogeny and chloroplast evolution of the paper daisy tribe Gnaphalieae (Asteraceae). Gene. 2024;901:148177.
55. Zhang L, Meng Y, Wang D, He GH, Zhang JM, Wen J, Nie ZL. Plastid genome data provide new insights into the dynamic evolution of the tribe Ampelopsideae (Vitaceae). BMC Genomics. 2024;25(1):247.
56. Xu S, Shao S, Feng X, Li S, Zhang L, Wu W, Liu M, Tracy ME, Zhong C, Guo Z. Adaptation in unstable environments and global gene losses: small but stable gene networks by the May-Wigner Theory. Mol Biol Evol. 2024;41(4):msae059.

Publisher's Note