

Comparative and Phylogenetic Analysis of the Complete Chloroplast Genomes of Six *Polygonatum* Species (Asparagaceae)

Dongjuan Zhang

Chinese Academy of Sciences

Jing Ren

Chinese Academy of Sciences

Hui Jiang

Chinese Academy of Sciences

Vincent Okelo Wanga

Chinese Academy of Sciences

Xiang Dong

Chinese Academy of Sciences

Guangwan Hu (✉ guangwanhu@wbcas.cn)

Chinese Academy of Sciences

Research Article

Keywords: Polygonatum, Heteropolygatum, chloroplast genome, phylogenetic analysis, Asparagaceae

Posted Date: May 31st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1669614/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: *Polygonatum* Miller is the largest genus in the tribe Polygonateae of Asparagaceae, and the horizontal creeping fleshy roots of several species in this genus serve as traditional Chinese medicine. Previous studies have been concerned mainly with the size and gene contents of the plastome, no intensive plastid genomic comparative analysis of this genus has been conducted, and there are still some species whose chloroplast genome information has not been reported.

Results: In this study, the complete plastomes of six *Polygonatum* were sequenced and assembled, among which, the chloroplast genome of *P. campanulatum* and *P. franchetii* were reported for the first time. Comparative and phylogenetic analyses were then conducted with the published plastomes of three related species. Results indicated that the whole plastome length of the *Polygonatum* species ranged from 154,565 bp (*P. multiflorum*) to 156,028 bp (*P. stenophyllum*) having a quadripartite structure of LSC and SSC separated by two IR regions. A total of 113 unique genes were detected in each of the species. Comparative analysis revealed that gene content, protein-coding genes and total GC content in these species were highly identical. No significant contraction or expansion was observed in the IR boundaries among all the species except *P. sibiricum*, in which the *rps19* gene was pseudogenized owing to incomplete duplication. Abundant long dispersed repeats and SSRs were detected in each genome. There were eight remarkably variable regions and 14 positively selected genes were identified among *Polygonatum* and *Heteropolygonatum*. Phylogenetic results based on chloroplast genome illustrated that *P. campanulatum* with alternate leaves were strongly supported locating in sect. *Verticillata*, a group characterized by whorled leaves. Moreover, *P. verticillatum* and *P. cyrtoneura* were displayed as paraphyletic.

Conclusions: This study revealed that the characters of plastomes in *Polygonatum* and *Heteropolygonatum* maintained a high similarity. Eight highly variable regions were found to be potential specific DNA barcodes in *Polygonatum*. Phylogenetic results suggested that leaf arrangement was not suitable as a basis for delimitation of subgeneric groups in *Polygonatum* and the definitions of *P. cyrtoneura* and *P. verticillatum* require further study.

1. Introduction

According to APG IV [1], *Polygonatum* Miller is a monotypic genus of Asparagaceae. The species of this genus are perennial herbs with horizontal creeping fleshy roots and unbranched stems [2]. This genus comprises approximately 80 species in the world (<https://wcsp.science.kew.org/>), accessed 30 March 2022). According to Chen and Tamura [2], 39 species have been recorded in China with 20 of them being endemic to the region. *Polygonatum* is widely distributed in Northern Hemisphere, with the center of diversity in East Asia, especially in the Hengduan Mountains of southwest China and the eastern Himalayas [3, 4]. It has a significant medicinal value, with several species of this genus, such as *Polygonatum kingianum* and *P. sibiricum* being used as traditional Chinese medicine due to their properties of tonifying Qi, nourishing Yin, strengthening the spleen, moistening the lung and benefiting the kidney [5].

Phylogenetic relationships reconstructed using ribosomal ITS and plastid DNA sequence suggested the monophyly of *Polygonatum* and its sister relationship to *Heteropolygonatum* M.N. Tamura & Ogisu [6–9]. In terms of infrageneric classification of this genus, it received considerable attention from researchers in history owing to the wide phenotypic variation within and among species. Baker [10] subdivided *Polygonatum* into three sections according to the leaf arrangement: the sect. *Alternifolia* with alternate leaves, sect. *Oppositifolia* with opposite leaves and sect. *Verticillata* with whorled leaves. However, phyllotaxy types in this genus were considered to be unstable in subsequent studies [7]. On account of morphological traits like leaf arrangement, bract size and texture, length of the perianth tube, perianth shape, anther length and ovary shape, Tang Y.C. and his colleagues proposed eight series for *Polygonatum* distributing in China [11]. Based on karyological and micromorphological characters, Tamura [12] sub-divided *Polygonatum* into the sects. *Polygonatum* and *Verticillata*. Recently, Meng and Nie reconstructed the phylogenetic relationship among this genus using four chloroplast (cp) genes, *rbcl*, *trnK*, *trnC-petN* and *psbA-trnH*, and they proposed a new group on the basis of Tamura's work, namely sect. *Sibirica* [7]. By then, *Polygonatum* was divided into sect. *Polygonatum*, sect. *Verticillata* and sect. *Sibirica*. This infrageneric classification system was most widely accepted and was demonstrated by Floden's research based on the complete cp genomes of *Polygonatum* [13].

The chloroplast is a unique organelle found in green plants that is responsible for photosynthesis. It has a separate genome from the nuclear and the mitochondria genomes, and is mostly inherited matrilineally in angiosperms. Compared to the nuclear and the mitochondrial genomes, plastomes are small, less vulnerable to recombination, with low nucleotide substitution rates as well as generally more conserved in terms of gene structure and organization, and therefore can provide unique genetic information [14, 15]. Among most higher plants, the cp genome possesses a typical tetrad structure comprising a small single-copy (SSC), a large single-copy (LSC), and two inverted repeats (IRs) [16]. Most cp genomes examined in plants have a constrained size varies from 120–160 kb, and this discrepancy is mainly related to expansion/contraction or even loss of IR [14, 17, 18]. Considerable genetic information is involved in the cp genome, which encodes about 120–130 genes [19]. These genes can be divided into three categories. The first is those involved in chloroplast gene expression, the second is genes related to photosynthesis, and the third is genes whose function is not yet clear [20]. The speed of molecular evolution between the coding and non-coding regions of chloroplast genomes differs noticeably, which is suitable for systematic studies at different levels [21].

Benefiting from advances in next-generation sequencing technologies, cp genomes can be obtained more efficiently and economically. In the National Center for Biotechnology Information (NCBI) organelle genome database, there are over 8000 cp genomes of plants have been published currently. Among angiosperms, plenty of cp genomes have been successfully employed to address the issues of phylogenetic relationships and species identification at different taxonomic levels [22–26]. There are about 115 complete cp genome sequences (ca. 40 species) in *Polygonatum* have been reported by NCBI (2022/4/26), but previous studies have been concerned mainly with the size and gene contents of the plastid genome, and detailed comparative genomic analysis are lacking [13, 27]. Although, chloroplast gene fragments and complete cp genomes between species in *Polygonatum* have been adopted for phylogenetic analysis recently [7, 13, 27]. there are still some species whose complete chloroplast genome data have not been published, and the phylogenetic placement of them is not well understood.

In this study, we reported the initial complete chloroplast genomes of *Polygonatum campanulatum* and *P. franchetii*, together with the sequencing of *P. cyrtoneura*, *P. filipes*, *P. zanzlanense* and *P. sibiricum*, and then compared them with three related species i.e. *P. kingianum* (MW373517), *Heteropolygonatum alternicirrhosum* (MZ150832), *H. ginfushanicum* (MW363694). *P. campanulatum* is a critically endangered species discovered by professor Guangwan Hu in Yunnan Province in 2011. No molecular information about this species has been reported before. Except for the two species in *Heteropolygonatum*, seven species in *Polygonatum* (including six newly sequenced species) were selected for plastid genomic comparative analysis, which cover the three subgroups of *Polygonatum*, as well as the major branches. There are 14 plastomes in *Polygonatum* failed to be verified in NCBI (Table S1). Manual checking found that among the 14 unverified plastomes, the two IR regions has different length and the discrepancy occurred only in no-coding regions. Therefore, these unverified plastomes have been used to reconstruct phylogenetic relationships only, and not for comparative analysis of the cp genome. Another 51 published cp genome sequences (47 from *Polygonatum*; three from *Heteropolygonatum*; *Maianthemum henryi* was chosen as outgroup) obtained from the NCBI database were employed to reconstruct phylogenetic tree. The aims of this study were to 1) conduct a comprehensive analysis of the chloroplast genome among the six *Polygonatum* and its related species; 2) explore hotspots regions of *Polygonatum* from the cp genomes; 3) infer the phylogenetic relationships of *Polygonatum* species and determine the taxonomic status of *P. campanulatum*, *P. franchetii*, *P. cyrtoneura*, *P. filipes*, *P. zanzlanense* and *P. sibiricum* based on cp genome.

2. Materials And Methods

2.1 Sample collection, Total DNA Extraction and Sequencing

Table 1
Specimen collection information of the six *Polygonatum* samples.

Species	Voucher specimen number	Date	Locality	Decimal Latitude	Decimal Longitude
<i>Polygonatum franchetii</i>	HGW-1223	2019-09-02	Gaowangjie National Nature Reserve, Guzhang County, Hunan Province, China	-/	-/
<i>Polygonatum zanlanscianense</i>	HGW-1357	2021-05-03	Malinyaozu Village, Xinning County, Hunan Province, China	26°27'N	110°38'E
<i>Polygonatum filipes</i>	HGW-1359	2021-05-03	Malinyaozu Village, Xinning County, Hunan Province, China	26°27'N	110°38'E
<i>Polygonatum sibiricum</i>	HGW-1379	2021-05-19	Donggang District, Rizhao City, Shandong Province, China	-/	-/
<i>Polygonatum campanulatum</i>	HGW-Z-2259	2019-07-27	Xima Township, Yingjiang County, Yunnan Province, China	24°47'N	97°40'E
<i>Polygonatum cyrtoneura</i>	HGW-Z-2364	2020-08-19	Tiantangzhai Township, Jinzhai County, Anhui Province, China	31°13'N	115°42'E

The six newly sequenced *Polygonatum* species (*Polygonatum campanulatum*, *P. filipes*, *P. franchetii*, *P. zanlanscianense*, *P. cyrtoneura*, *P. sibiricum*) were collected by Guangwan Hu in China during 2019 to 2021. And detailed field collection information of them was described in Table 1. These samples were collected with the permission of Gaowangjie National Nature Reserve, Hunan Forestry Bureau, Shandong Forestry Bureau, Yunnan Forestry Bureau and Anhui Forestry Bureau. The collected species were identified and verified by professor Guangwan Hu, from Wuhan Botanical Garden, Chinese Academy of Science. Voucher specimens were deposited at the Herbarium of Wuhan Botanical Garden, CAS (HIB) (China), with voucher specimen numbers listed in Table S1. Total genome DNA was extracted from dry leaves preserved in silica gels, using a modified cetyltrimethylammonium bromide (CTAB) method, and then sequenced based on the Illumina HiSeq X Ten platform, 150 bp paired-end reads (PE150) at Novogene Co., Ltd. (Beijing, China)

2.2 Assembly and Annotation of Chloroplast Genome

Chloroplast genome assembling was accomplished by Get Organelle v1.7.5 [28] with default parameters, and gene annotation was completed by PGA (Plastid Genome Annotator) software [29] with *Amborella trichopoda* as a reference. To ensure the reliability of the data used for subsequent analysis, all chloroplast genome download from NCBI was annotated over again by PGA. Manual checking and adjustment of the annotation results, including positions of initiation and termination codons and boundaries of IR repeat regions, were performed in Geneious v10.2.3 [30]. Annotated chloroplast genome sequences of the six species were submitted to GenBank (Table S1) in NCBI. Further, the circular chloroplast genome map was drawn online by OGDRAW [31]

2.3 Comparative Analysis of the Whole Chloroplast Genome

Geneious v10.2.3 [30] was employed to analyze length and guanine-cytosine (GC) content of the whole chloroplast genome, LSC, SSC and IR regions, together with numbers of genes and genes categories. Multiple genome alignment analysis was performed in MAFFT program [32]. Comparative chloroplast genomes divergence was conducted and visualized by mVISTA [33] with the annotation of *Polygonatum campanulatum* as a reference in Shuffle-LAGAN mode. To detect the contraction or expansion at the boundaries, the SC/IR boundary analysis of the chloroplast genomes was carried out by IRscope [34]. Mauve [35] was adopted to perform the analyses of cp genome rearrangement based on default settings.

2.4 Codon usage, and repeated sequences analysis

Relative synonymous codon usage (RSCU) value was detected using MEGA v7.0 [36]. RSCU is defined as the ratio of the observed frequency of a codon to the expected frequency without preference. The values greater than 1.0 mean that the particular codons are used more frequently than normal while the reverse indicates the opposite [37].

Long dispersed repeats were identified using REPuter [38] with a hamming distance equal to 3 bp, and repeat size no less than 30 bp. Simple sequence repeats (SSRs) were identified using MicroSatellite identification tool (MISA) [39] with minimum parameters being set as 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotides SSR motifs, respectively. **2.5 Nucleotide diversity analysis and selected pressure**

DnaSP [40] was adopted to analyze the nucleotide diversity (π) with the window length of 600 bp and the step size of 200 bp. Given that DnaSP v6 cannot recognize degenerate bases, like M, K, and Y, dashes were used to take the place of these letters. Further, the figure was generated in Excel and optimized in Adobe Illustrator.

To identify the positive selection loci of coding sequences (CDS) in the cp genome, the dN/dS values were calculated by employing EasyCodeML v1.12 [41]. Each single-copy CDS were extracted from the complete chloroplast genome using Geneious v10.2.3 [30], after aligning under the codon model, they were finally combined into one matrix. The input tree was an ML tree reconstructed by IQ-TREE [42]. Four site models (i.e. M0 vs. M3, M1a vs. M2a, M7 vs. M8, and M8a vs. M8) along with a likelihood ratio test (LRT) were used to perform the analyses. Naive Empirical Bayes (NEB) and Bayes Empirical Bayes (BEB) [43] analyses were conducted under the M8 model to identify positive selection loci and the selected genes.

2.6 Phylogenetic Analysis

The phylogenetic analysis was performed based on the complete chloroplast genomes of 54 *Polygonatum* taxa and five *Heteropolygonatum* taxa. *Maianthemum henryi* was set as an outgroup. The chloroplast genomes of all species were obtained from GenBank (Table S1), except for *Polygonatum campanulatum*, *P. filipes*1, *P. franchetii*, *P. zanzlanscianense* 1, *P. cyrtoneuma* 1, and *P. sibiricum* 1. The total matrix was aligned under MAFFT [32]. ModelFinder [44] was adopted to select the best-fit model according to the Bayesian information criterion (BIC). Maximum likelihood (ML) phylogenetic tree was reconstructed using IQ-TREE [42] under the K81u + R6 + F model for 5000 ultrafast bootstraps [45]. BI (Bayesian inference) analysis was conducted using MrBayes v3.2.6 [46] based on GTR + F + I + G4 model. Two independent Markov Chain Monte Carlo (MCMC) run for 1,000,000 generations, trees were sampled every 100 generations, and the initial 25% of sampled data were discarded as burn-in. The two output trees were visualized and improved by Figtree v1.4 (<http://github.com/rambaut/figtree/>)

3. Results

3.1 Chloroplast genome structure and characteristics analyses

The complete chloroplast genomes of the species in *Polygonatum* displayed closed circular and common tetrad structures. The length of cp genomes ranging from 154,565 bp (*P. multiflorum*) to 156,028 bp (*P. stenophyllum*) in *Polygonatum* and 155,508–155,944 bp in *Heteropolygonatum* (Table 2). Each genome included a large single-copy (LSC), a small single-copy (SSC) and a pair of inverted repeats (IRa and IRb) that separated the LSC and SSC regions (Fig. 1). LSCs of the species ranged from 83,486 bp (*P. odoratum*) to 84,968 bp (*H. alternicirrhosum*), while SSC varied between 18,303 bp (*P. cyrtoneuma*) to 18,576 bp (*P. sibiricum*), and the sizes of IRs were between 52,174 bp (*P. inflatum*) and 52,832 bp (*P. sibiricum*) (Table 2). The Guanine-Cytosine (GC) content of the total genome ranged from 37.6–37.8%. Further, GC content exhibited an unbalanced distribution both in the cp genomes of *Polygonatum* and *Heteropolygonatum*. The SSCs presented the lowest GC content of 31.4–31.7%, followed by LSCs (35.6–35.8%), whereas the IRs have the highest GC content from 42.9–43% (Table 2).

All the species were detected with 132 genes (113 unique genes) in the same order in the complete cp genome. The whole genomes included 87 protein-coding genes (PCGs), 38 transfer RNA (tRNA) genes and 8 ribosomal RNA (rRNA) genes (Table 2). Moreover, a total of 19 genes, comprising seven PCGs (*rps19*, *rpl2*, *rpl23*, *ycf2*, *ndhB*, *rps7*, *rps12*), 8 tRNA genes (*trnN-GUU*, *trnR-ACG*, *trnA-UGC*, *trnI-GAU*, *trnV-GAC*, *trnL-CAA*, *trnI-CAU*, *trnH-GUG*) and 4 rRNA genes (*rnn5*, *rnn4.5*, *rnn23*, *rnn16*) were duplicated in the pair of inverted repeats. We also observed the pseudogenization of *rps19* gene located in the boundary regions between IRa and LSC in *P. sibiricum* due to incomplete duplication of the normal copy, while *rps19* genes located inside the IRs with complete duplication in the other species analyzed in this study. In addition, a total of 18 genes (*trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*, *rps12*, *rps16*, *rpl2*, *rpl16*, *rpoC1*, *petB*, *petD*, *atpF*, *ndhA*, *ndhB*, *clpP*, *ycf3*) and six tRNA contained at least one intron in the complete cp genome, in which *clpP* and *ycf3* included two introns. Particularly, *rps12* gene was a trans-spliced gene with the 5' exon situated in the LSC region and the two copies of 3' exon and intron sitting in the IRs. The longest intron was

identified in *trnK-UUU* with the length of 2568–2584 bp and the *matK* gene was placed inside the intron (Table S2). All of the functional genes can be divided into three categories, i.e. self-replication genes, photosynthesis genes, and other genes (Table 3).

Table 2

General information and comparison of chloroplast genomes of the 25 *Polygonatum* and two *Heteropolygonatum* species.

Species	Genome length (bp)				GC content (%)				Gene number			
	Total	LSC	SSC	IR	Total	LSC	SSC	IR	Total	PCG	tRNA	rRNA
<i>P. acuminatifolium</i>	155354	84271	18455	52628	37.7	35.7	31.6	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. campanulatum</i>	155487	84458	18373	52656	37.6	35.7	31.7	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. cirrhifolium</i>	155944	84568	18546	52830	37.6	35.7	31.5	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. curvistylum</i>	155939	84563	18546	52830	37.6	35.7	31.5	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. cyrtonema</i>	155509	84448	18303	52758	37.7	35.7	31.7	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. filipes</i>	155361	84307	18454	52600	37.7	35.7	31.6	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. franchetii</i>	155962	84722	18566	52674	37.6	35.7	31.5	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. hirtum</i>	155490	84385	18419	52686	37.7	35.7	31.6	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. hookeri</i>	155976	84600	18546	52830	37.6	35.7	31.5	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. hunanense</i>	155609	84438	18427	52744	37.7	35.7	31.5	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. inflatum</i>	154898	84270	18454	52174	37.7	35.7	31.6	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. kingianum</i>	155824	84632	18570	52622	37.7	35.7	31.5	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. macropodum</i>	154610	83554	18464	52592	37.7	35.8	31.6	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. multiflorum</i>	154564	83515	18457	52592	37.7	35.8	31.5	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. nodosum</i>	155205	84052	18422	52640	37.7	35.7	31.6	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. odoratum</i>	154569	83486	18459	52624	37.8	35.8	31.6	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. prattii</i>	155915	84538	18547	52830	37.7	35.7	31.5	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. punctatum</i>	155657	84542	18423	52692	37.7	35.7	31.5	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. sibiricum</i>	155514	84664	18576	52832	37.7	35.7	31.5	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. stenophyllum</i>	156028	84677	18561	52790	37.7	35.7	31.6	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. tessellatum</i>	155688	84488	18564	52636	37.6	35.7	31.4	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. uncinatum</i>	155797	84614	18531	52652	37.7	35.7	31.6	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)

Species	Genome length (bp)				GC content (%)				Gene number			
<i>P. verticillatum</i>	155856	84545	18513	52798	37.7	35.7	31.6	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. zanolanscianense</i>	155787	84418	18539	52830	37.6	35.7	31.6	42.9	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>P. involucreatum</i>	155370	84280	18450	52640	37.7	35.7	31.6	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>H. alternicirrhosum</i>	155944	84968	18520	52456	37.6	35.6	31.5	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)
<i>H. ginfushanicum</i>	155508	84552	18528	52428	37.6	35.6	31.4	43	132 (19×2)	86 (7×2)	38 (8×2)	8 (4×2)

Table 3
The annotated genes in the chloroplast genomes of *Polygonatum*

Category	Gene Group	Gene Name
Self-replication	Ribosomal RNA	<i>rrn4.5^c, rrn5^c, rrn16^c, rrn23^c</i>
	Transfer RNA	<i>trnA-UGC^{a,c}, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC^a, trnH-GUG^c, trnI-CAU^c, trnI-GAU^{a,c}, trnK-UUU^a, trnL-CAA^c, trnL-UAA^a, trnL-UAG, trnM-CAU, trnM-CAU, trnN-GUU^c, trnP-UGG, trnQ-UUG, trnR-UCU, trnR-ACG^c, trnS-UGA, trnS-GCU, trnS-GGA, trnT-GGU, trnT-UGU, trnV-UAC^a, trnV-GAC^c, trnW-CCA, trnY-GUA</i>
	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7^c, rps8, rps11, rps12^{a,c}, rps14, rps15, rps16^a, rps18, rps19^c</i>
	Large subunit of ribosome	<i>rpl2^{a,c}, rpl14, rpl16^a, rpl20, rpl22, rpl23^c, rpl32^a, rpl33, rpl36</i>
	RNA polymerase subunits	<i>rpoA, rpoB, rpoC1^a, rpoC2</i>
Photosynthesis	Photosystem ☒	<i>psaA, psaB, psaC, psal, psaJ</i>
	Photosystem ☒	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunits of cytochrome	<i>petA, petB^a, petD^a, petG, petL, petN</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF^a, atpH, atpI</i>
	NADH-dehydrogenase	<i>ndhA^a, ndhB^{a,c}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
Other genes	Rubisco large subunit	<i>rbcL,</i>
	Translational initiation	<i>infA</i>
	Maturase	<i>matK</i>
	Envelope membrane protein	<i>cemA</i>
	Acetyl-CoA-carboxylase	<i>accD</i>
	Proteolysis	<i>clpP^b</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
Conserved open reading frames	<i>ycf1, ycf2^c, ycf3^b, ycf4</i>	

^a, genes with one intron; ^b, genes with two introns; ^c, two genes copied in IR regions

3.2 Relative synonymous codon usage analysis

Given that codon usage is closely related to genome-wide protein and mRNA levels, it is an essential feature of gene expression, and the same codon presents different frequencies in different organisms. The codon usage frequencies of *Polygonatum*

campanulatum, *P. filipes*, *P. franchetii*, *P. zanolanscianense*, *P. cyrtonea* and *P. sibiricum* were computed based on protein-coding genes of the complete chloroplast genome. The total codons in these six species varied from 26,512 (*P. cyrtonea*, *P. sibiricum*) to 26,651 (*P. zanolanscianense*). The most abundant amino acid (AA) was leucine (Leu), with the proportion of 10.2% of all AA in *P. cyrtonea* and 10.3% in other species studied here, followed by serine (Ser) accounting for 7.8%-7.9% (Table S3). In contrast, cystine (Cys) possessed the lowest number of codons (306–308) in all the six species when terminal codons were not considered. The AGA codon, encoding arginine (Arg), presented the highest RSCU (relative synonymous codon usage) value of 1.92–1.96, while AGC codon, encoding serine (Ser), showed the lowest RSCU value with 0.31–0.33 (Table S3). Additionally, CGC encoding Arginine (Arg) shared the same lowest RSCU value of 0.31 in *P. campanulatum*, *P. filipes* and *P. sibiricum*. Figure 2 illustrated the summary statistics for amino acid frequency and relative synonymous codon usage. Among the 64 codons, there were 31 codons with RSCU values less than 1 (RSCU < 1), which showed a lower usage frequency than expected. Meanwhile, 30 codons were used more frequently than expected in *P. campanulatum* and *P. filipes* with RSCU values greater than 1 (RSCU > 1), while 31 codons in *P. franchetii*, *P. zanolanscianense*, *P. cyrtonea* and *P. sibiricum*. Furthermore, the RSCU values of AUG and UGG in all the six species were equal to one (RSCU = 1) appearing without usage preference, while UCC only showed the same characteristics in *P. campanulatum* and *P. filipes*. Particularly, methionine (AUG) and tryptophan (UGG) were encoded by only one codon. All codons with RSCU > 1 were characterized by Adenine-Thymine ending in the six species apart from UUG and the UCC in *P. franchetii*, *P. zanolanscianense*, *P. cyrtonea* and *P. sibiricum*. On the contrary, 28 of the 31 codons with RSCU < 1 were detected ending with Guanine-Cytosine (GC) in each species. The comparative analysis of RSCU value among the six *Polygonatum* appeared almost no discrepancies, showing relative stability in codon usage bias of *Polygonatum* (Fig. 2).

3.3 Long dispersed Repeats and microsatellites analysis

A total of 378 long dispersed repeats were observed in the seven *Polygonatum* and two *Heteropolygonatum* species, consisting of 191 palindromic repeats, 177 forward repeats, nine reverse repeats and one complementary repeat (the palindromic repeat of IR regions itself was excluded in all the nine species) (Table S4). Obviously, palindromic repeats were the dominant repeat type (from 47.2% in *P. filipes* to 53.5% in *P. zanolanscianense*), while complementary repeats were the least frequent one which was only detected in *P. campanulatum* (2.7%). Likewise, *P. franchetii* and *H. ginfushanicum* did not possess any reverse repeats. On the other hand, the species that harbor the highest number of long repeats was *P. zanolanscianense* (49), and the species with the lowest number was *P. kingianum* (35) (Fig. 3, A). In *H. ginfushanicum*, the length of the longest repeat sequence was 66 bp while in the rest eight species were 71 bp, and all of them were forward repeats. Furthermore, among all repeats detected in the nine species, the length of repeats ranging from 30–34 bp accounted for the majority (260, 68.1%) (Fig. 3, B; Table S5). The most repeats were detected in the CDS, followed by IGS regions, some repeats were also identified between CDS, IGS, tRNA and introns (Fig. 3, C; Table S6). Most of the repeat sequences were located in the IR regions except for *P. campanulatum* and *P. filipes*, which harbored the highest number of repeats in LSC region (Fig. 3, D; Table S7).

In this study, we observed 507 SSRs among the nine species in total, comprising 303 mono-, 91 di-, 27 tri-, 63 tetra-, 20 penta-, and two hexa-nucleotide repeats (Table S8). Moreover, a total of two mono-, three di-, four tri-, eight tetra-, four penta- types and two hexa-nucleotide repeats types were identified. And one tri-, two tetra-, three penta- and two hexa-nucleotide types were observed only once in only one species (Table S9). Most SSRs were mononucleotide and dinucleotide repeats, besides, the rest of SSRs showed lower frequencies. As shown in Fig. 4, A, mono-nucleotide repeats were the most frequent type ranging from 55.9% (*Polygonatum kingianum*) to 61.8% (*Heteropolygonatum ginfushanicum*). The number of SSRs of *H. alternicirrhosum* reached a peak value of 64 among the nine species. On the other hand, *P. sibiricum* possessed the least number of SSRs of 50 (Fig. 4, A, Table S9). The most dominant SSRs were A/T polymers (Fig. 4, B–J), suggesting a remarkable base preference. And the majority of the microsatellites were located in the LSC region (Table S10). These results indicate that there were no distinctive differences in SSRs between *Polygonatum* and *Heteropolygonatum*, and the identified SSRs will provide valuable genetic information for the phylogeny and population genetics of *Polygonatum* in the future.

3.4 Comparative Genome Analysis and Sequence Variation

To identify highly variable regions among the nine *Polygonatum* and *Heteropolygonatum* species, multiple sequence alignment of the cp genomes was carried out. The annotation of *Polygonatum campanulatum* was set as a reference. It can be seen from the data in Fig. 5 that coding regions were much more conserved than non-coding regions, with almost no significant variations

except for *ycf1*. Additionally, we detected that some intergenic spacer region and introns appeared considerable variations, including *rps16-trnQ*, *trnS-trnG*, *atpF-atpH*, *atpH-atpI*, *petA-psbJ*, *ndhF-rpl32*, *rpl32-trnL* and *rpl16*. These regions were potential tools for the development of molecular markers for the identification of *Polygonatum*. Another significant result was that compared with the IRs regions, LSC and SSC regions showed higher variation, which was consistent with the result of nucleotide polymorphisms analysis (Fig. 5). Apart from *ycf1*, all highly divergent regions mentioned above were located in single-copy regions. With respect to tRNA and rRNA, they were strongly conserved without evident variations. Additionally, collinearity detection analysis found that there were no interspecific or intraspecific rearrangements in the nine species (Fig. 6).

3.5 Expansion and contraction of IRs

Figure 7 Comparative analysis of the LSC, IR and SSC boundary regions in the nine chloroplast genomes.

A comprehensive comparison of boundaries between single-copy and the IRs regions was carried out. We observed that the complete cp genome structure of the nine species varied from each other slightly. Apart from *Polygonatum sibiricum*, the junctions of LSC/IRb sit between *rpl22* gene and *rps19* gene among the other eight species. The *rpl22* gene was located in the LSC region completely with 26 bp to 34 bp away from LSC/IRb border, while the *rps19* genes within IR regions were close to two IR/LSC boundaries. Furthermore, in *P. sibiricum*, two *rps19* genes extended into the LSC region due to the contraction of IRs (Fig. 7), leading to the one located at IRa/LSC junction being a pseudogene. Apart from this special case, *rps19* in the other species was quite conservative with the same length of 279 bp. Likewise, *rpl22* gene was also very conserved with the same length of 366 bp in all the nine species. Moreover, *ndhF* gene was located in the boundaries of IRb/ SSC and expanded to the IRb region by 22, 29, or 34 bp. And *trnN* gene was close to the IRs/ SSC boundaries with the whole gene within IRs regions. The *ycf1* gene ranges from 4454–4573 bp and straddled the SSC/IRa boundary, with 883–895 bp distributed in the IRa region and the rest in the SSC region (Fig. 7). In terms of IRa-LSC boundary, *rps19* gene was located on the left side while *psbA* gene was on the right, and *psbA* gene was highly conserved with a steady length of 1062 bp. The distances between *psbA* and the IRa/LSC junction varied from 87 to 94 bp.

Together these results provided important insights into contractions and expansions of IR region borders in *Polygonatum* and *Heteropolygonatum*. The structures and gene orders of the two genera were relatively conserved except for *P. sibiricum*, in which a slight expansion and contraction occurred between IRs and LSC.

3.6 Nucleotide Diversity and Selective Pressure Analysis

Nucleotide diversity of nine chloroplast genomes of *Polygonatum* and *Heteropolygonatum* was calculated to detect divergence hotspots. It turned out that the pair of inverted repeats were relatively conserved regions with an average Pi value of 0.00113, while LSC and SSC showed higher nucleotide diversity with a mean Pi value of 0.00492 and 0.00674 respectively. Significant variations (Pi > 0.014) were found in the following regions: *trnK-UUU-rps16*, *trnC-GCA-petN*, *trnT-UGU-trnL-UAA*, *ccsA-ndhD* and *ycf1* (Fig. 8), in which the most divergent region was *trnK-UUU-rps16*, with the Pi value of 0.01565. Of these five regions, 80 percent (4) were intergenic genes, whereas protein-coding regions accounted for 20% (1), indicating that non-coding regions harbored more variations and coding region was more stable and more conservative. What's more, all the divergent hotspots might be potential molecular markers for DNA barcodes adopted into species identification and phylogenetic studies in the future.

Synonymous substitutions in the nucleotide preserve the same amino acids, on the contrary, non-synonymous substitutions will change the amino acids. The substitution rates of nonsynonymous (dN) and synonymous (dS) have been widely used for quantifying adaptive molecular evolution in the chloroplast genome [47]. In the current study, according to BEB methods, a total of 14 genes corresponding to 65 sites were detected under positive selection, among them, four genes (*rpoC2*, *rpoB*, *psaA*, *ndhK*) were identified under significant positive selection, and ten genes (*psbA*, *psbK*, *atpA*, *rpoC1*, *psbD*, *psbC*, *psbZ*, *psaB*, *rps4*, *ndhJ*) under positive selection (Table S11). All the selected genes were located in LSC regions, and 10 of them were related to photosynthesis. We observed that *rpoC2* harbored the highest number of sites that under positive selection (13), followed by *psaA* (12) and *rpoB* (11).

3.7 Phylogenetic analysis of *Polygonatum*

A total of 60 cp genome sequences of *Polygonatum* and its related species were selected to reconstruct phylogenetic relationships among this genus. *Maianthemum henryi* was chosen as an outgroup own to its closer distances and more basic position to *Polygonatum* and *Heteropolygonatum*. The 60 species comprise six newly sequenced data (i.e. *Polygonatum campanulatum*, *P. filipes*, *P. franchetii*, *P. zanolanscianense*, *P. cyrtoneuma*, *P. sibiricum*) and 54 cp genome published in NCBI (Table S1). The topologies of Maximum likelihood (ML) and Bayesian inference (BI) were highly identical both in tree structure and species position with generally strong support (Fig. 9). The only difference was the position of *Polygonatum inflatum*, it appeared as sister to *P. filipes* + *P. yunnanense* + *P. nodosum* in the ML tree, while as sister to *P. filipes* + *P. yunnanense* + *P. nodosum* + *P. odoratum* + *P. multiflorum* + *P. macropodum* + *P. acuminatifolium* + *P. involucreatum* + *P. orientale* according to BI analysis. Both *Polygonatum* and *Heteropolygonatum* exhibited monophyletic relationships and shared the most recent common ancestor. *Polygonatum* was divided into two main lineages including sect. *Verticillata* and the clade consisting of sect. *Polygonatum* and sect. *Sibirica*. Phylogenetic analysis suggested that sect. *Sibirica* comprise only one species, i.e. *P. sibiricum*. Moreover, we also observed that *Polygonatum verticillatum* and *P. cyrtoneuma* were paraphyletic. One sample of *P. verticillatum* was sister to *P. zanolanscianense* (BS = 100, PP = 1.00), while the other appeared as sister clade to *P. curvistylum* + *P. pratti* + *P. stewartianum* (BS = 100, PP = 1.00). Four samples of *P. cyrtoneuma*, including the newly sequenced one, appeared as sister to *P. hunanense* (BS = 100, PP = 1.00) and this clade locates at the base of sect. *Polygonatum*. Whereas the other two samples present as sister clade to *P. hirtum* with significantly high Bayesian posterior probability and bootstrap support (BS = 100, PP = 1.00). With respect to *P. franchetii*, it was sister clade to *P. stenophyllum* (BS = 100, PP = 1.00). Furthermore, *P. filipes* was strongly supported to be included in sect. *Polygonatum* and being sister to *P. yunnanense* plus *P. nodosum* (BS = 99, PP = 1.00). Surprisingly, *P. campanulatum* with alternate leaves located in sect. *Verticillata*, a group characterized by whorled leaves, and formed a sister clade with *Polygonatum tessellatum* plus *Polygonatum oppositifolium* (BS = 100, PP = 1.00), which suggested that leaf arrangement is not suitable as the basis for delimitation of subgeneric groups in *Polygonatum*.

4. Discussion

Features of Complete Chloroplast Genome and Comparative Analyses

In the current study, we reported the initial complete cp genomes for two *Polygonatum* species, *Polygonatum campanulatum* and *P. franchetii*. Additionally, the complete cp genomes of another four species were newly sequenced (*P. cyrtoneuma*, *P. filipes*, *P. zanolanscianense*, *P. sibiricum*) using Illumina sequencing technology. Besides, cp genomic comparative analyses were carried out among the six species plus another three related species (*P. kingianum*, *Heteropolygonatum alternicirrhosum*, *H. ginfushanicum*) to understand potential genetic information of *Polygonatum*. The cp genome showed a typical quadripartite structure, with the length between 155,361 bp and 155,962 bp in *Polygonatum*, and 155,50–155,944 bp in *Heteropolygonatum*. The range of chloroplast genome length variation in these two species was similar to other Asparagaceae and higher plants reported previously [48–52]. And the size changes are partially caused by elongation or contraction of inverted repeat regions.

Our study revealed that gene content and gene order in the cp genome of *Polygonatum* and *Heteropolygonatum* were highly conserved, only slight variations in gene size and position were found. This result is similar to other species of Asparagaceae [53]. All plastomes contained 132 genes comprising 86 protein-coding genes, 38 tRNA and eight rRNA. Among these genes, 18 included intron and 19 were duplicated in IR regions. One interesting finding was that one of the *rps19* genes in *P. sibiricum* presented to be a pseudogene. Owing to its location at IR/LSC boundary, the gene lost the ability to duplicate completely. Likewise, this phenomenon has also been observed in *Polygonatum verticillatum* (NC_028523). Expression of the *rps19* gene is relatively unstable among species of Asparagaceae, the pseudogenization of *rps19* has also been reported in *Behnia reticulata*, *Hesperaloe parviflora* and *Hosta ventricosa*, while *Camassia scilloides* and *Chlorophytum rhizopendulum* missed this gene completely [54]. The *rps2*, *infA* and other pseudogenes reported previously in Asparagaceae were not detected in this study [54, 55]. In addition, although there were no remarkable variations in GC content among different species, the distribution of GC content was identified as asymmetrical. The higher GC content in IRs means a more stable structure in that GC pairs include three hydrogen bonds and AT pairs have two [56]. And this may be attributed to the presence of the four rRNA genes, which possess high-level GC nucleotide percentages. Similar results have been found in the chloroplast genomes of other angiosperms [57–59].

The pattern of codon usage is a vital genetic characteristic of the organism, related to the occurrence of mutation, selection and other molecular evolutionary phenomena [60]. Our results demonstrated that Leucine (Leu) presented the highest frequency of all amino acids in *Polygonatum campanulatum*, *P. filipes*, *P. franchetii*, *P. zanlanscianense*, *P. cyrtonema* and *P. sibiricum*. On the contrary, cystine (Cys) was the least abundant amino acid except for stop codons, which was also found in other angiosperm taxa [23, 61]. Furthermore, The result of RSCU analysis illustrated that most codons ended with A or U when RSCU value was greater than one, likewise, most codons ended with C or G when RSCU value was less than one. This phenomenon revealed that codon usage was biased towards A and U at the third codon position in *Polygonatum*, which coincided with previous studies [53, 58, 62].

Long dispersed repeats are essential for the rearrangement and stability of the chloroplast genome, and relevant to copy number differences among species [63]. Identifying their number and distribution plays a key role in genomic studies [64]. The current study found that palindromic repeats were the most common repeat type, followed by forward repeats. Whereas complementary repeat was identified only in *P. campanulatum*, and *P. franchetii* and *H. ginfushanicum* did not harbor any reverse repeats. In the plastomes of the nine species reported here, the length of repeats ranging from 30–39 bp are dominant, which is commonly observed in other angiosperm lineages [49, 65, 66]. Our study also revealed that the repetitive sequences were not randomly allocated in the cp genome of the nine *Polygonatum* and *Heteropolygonatum* species, they were mainly identified in the LSC region (48.7%) and CDs (51.9%).

SSR (Simple Sequence Repeats) is a kind of significant codominant DNA molecular marker, with the advantages of high abundance, random distribution throughout the genome and ample polymorphism information [67, 68]. Therefore, it provides important insights into many fields, such as species identification, phylogeography and population genetics [69, 70]. A total of 507 SSRs were detected in the current study, with *H. alternicirrhosum* containing the most. Further, among the nine cp genomes of *Polygonatum* and its related species, six categories of SSRs were observed in total. Mononucleotide SSRs showed the highest frequency in each genome, with A/T as the predominant motif type. Similar results had been reported in numerous taxa [50, 58, 71]. By contrast, hexanucleotide SSRs were the rarest type, with only one such element being observed in *P. cyrtonema* and *P. filipes*. In addition, SSRs lying within LSC regions accounted for the majority (72.4%), which was in agreement with previous studies [62, 66]. In summary, the microsatellites identified in this study will be developed as markers for *Polygonatum*, and contribute to species identification and evolutionary studies of this genus in the future.

The results of multiple sequence alignment revealed the similarities of cp genome in structure, content, and order among *Polygonatum* and its related species. Consistent with previous reports [72–74], we also found out that no coding regions harbored more distinctive variation than coding regions in this study, and two single-copy regions exhibited higher sequence divergence than the IRs. The following seven intergenic regions, i.e. *rps16-trnQ*, *trnS-trnG*, *atpF-atpH*, *atpH-atpI*, *petA-psbJ*, *ndhF-rpl32*, *rpl32-trnL* and two genes, i.e. *ycf1* and *rpl16* were detected as the most divergent. These regions can be adopted as potential molecular markers, which is a promising issue for future research. Comparative analysis among *Polygonatum* and its related species discovered that the cp genomes presented highly conserved, and no interspecific or intraspecific rearrangement was detected.

Contraction and expansion in IRs regions led to variations in cp genome size, which was observed in the evolutionary history of terrestrial plants commonly [59]. The size of IR regions was relatively similar in *Polygonatum* and *Heteropolygonatum*, ranging from 26,214 bp in *H. ginfushanicum* to 26,415 bp in *P. zanlanscianense*. Despite that, all the cp genomes showed similarity in the overall gene order and structures, several variations were identified at the junctions of IR/SC. The current study demonstrated that boundary genes in *Polygonatum* were mainly *rpl22*, *rps19*, *trnN*, *ndhF*, *ycf1* and *psbA*, which is also identified with *Heteropolygonatum* and *Hosta* [53]. It further confirms that boundary features are relatively stable across closely related species [75]. The LSC/IRb boundary was traversed by *rps19* gene in *P. sibiricum*, whereas the junctions located between *rpl22* and *rps19* in the other species. Incomplete duplication of the normal copy resulting in pseudogenization of the *rps19* gene located at IRa/LSC boundary, and this phenomenon has also been reported in other taxa of Asparagaceae, such as *Behnia reticulata*, *Hesperaloe parviflora* and *Hosta ventricosa* [54]. Excluding *rps19*, the other genes situated at SC/IR boundaries exhibited relative stability across the six *Polygonatum* and two *Heteropolygonatum* species studied in this work, and only *ndhF* and *ycf1* had slight variations in size. The high resemblances in boundaries between SC/IR also demonstrate that all the species share the same genes. Besides, the total number of genes does not change as a result of IR contraction and expansion [76].

We detected *trnK-UUU-rps16*, *trnC-GCA-petN*, *trnT-UGU-trnL-UAA*, *ccsA-ndhD* and *ycf1* were prominent divergent regions, with nucleotide diversity greater than 0.014. The result indicated that divergent regions located in LSC were in the majority, and the IR regions displayed relatively poor diversity, which was agree with the results of multiple sequence alignment conducted by mVISTA. The same phenomenon has been observed in many taxa [23, 65]. The regions detected in nucleotide diversity analysis might also provide additional genetic information for DNA barcodes in *Polygonatum*, but this required the support of further experiments.

The non-synonymous (dN) and synonymous (dS) substitution rates are beneficial in inferring the adaptive evolution of genes [24, 77]. The analysis of dN/dS was carried out owing to its popularity and reliability in quantifying selective pressure [78, 79]. Results indicated that several positively selected genes existed in *Polygonatum*, and these genes are relevant to photosynthesis and self-replication activities (Table S11), which has a positive effect on comprehension of the mechanisms that generate selection pressure.

Phylogenetic analysis

Phylogenetic analysis based on complete cp genome demonstrated that both *Polygonatum* and *Heteropolygonatum* were monophyly. Coinciding with the results of previous studies [7, 13, 27], *Polygonatum* was composed of three major clades, sect. *Verticillata*, sect. *Sibirica* and its sister clade sect. *Polygonatum*. In the current study, we observed that sect. *Sibirica* contained only one species, *P. sibiricum*, which was consistent with Xia and Meng's findings [7, 27], but data from Floden [13] suggest that one sample of *Polygonatum verticillatum* was sister to *Polygonatum sibicirum* within sect. *Sibirica*. Moreover, previous studies indicated that *P. verticillatum* was paraphyletic, potentially as a result of its wide geographic distribution and diverse morphological variations [13, 27]. The similar result was presented in this study. One sample of *P. verticillatum* exhibited as sister clade to *P. zanzlancianense* while the other was sister to *P. curvistylum* + *P. pratti* + *P. stewartianum*. With similarities to previous findings [27], *P. cyrtoneuma* was either recovered as paraphyletic in this study given that four samples including the newly sequenced one appeared as sister to *P. hunanense*, while the other two samples presented being sister relationship with *P. hirtum*. All the clades were supported highly. It suggests that the definition of *P. cyrtoneuma* requires further study.

Another important finding was that *Polygonatum franchetii* was strongly supported as sister clade to *P. stenophyllum*. The cp genome of *P. franchetii* was reported for the first time in this work. Before this, only Meng's team [7] reported the phylogenetic relationships included in *P. franchetii* using four chloroplast fragments (*rbcL*, *psbA-trnH*, *trnK* and *trnC-petN*). Regrettably, the branch structure which *P. franchetii* belonged to was ambiguous, making it difficult to recognize the relationship between *P. franchetii* and its close taxa. Furthermore, *P. filipes* presented sister clade to *P. yunnanense* plus *P. nodosum* within sect. *Polygonatum*. This finding is contrary to that of Xia et al. [27] who found *P. filipes* was sister to the clade consisting of *P. inflatum* + *P. multiflorum* + *P. odoratum* + *P. macropodum* + *P. involucreatum* + *P. acuminatifolium* + *P. arisanense* + *P. orientale* + *P. yunnanense* + *P. nodosum* with high support, however, the clade composed of *P. yunnanense* + *P. nodosum* was weakly supported as sister to the rest species in the sister clade of *P. filipes*.

One unanticipated finding was that phylogenetic tree strongly supported *Polygonatum campanulatum* placed in sect. *Verticillata*, though *P. campanulatum* arise alternately leaves and sect. *Verticillata* was characterized by whorled or opposite leaves. *P. campanulatum* was compared to *P. gongshanense* and *P. franchetii* when it was first published, but material for *P. gongshanense* was not available in this work. Furthermore, phylogenetic analysis indicated that *P. franchetii* and *P. campanulatum* presented in separate branches whereas *P. tessellatum* + *P. oppositifolium* were highly supported as sisters to *P. campanulatum* (BS = 100, PP = 1.00). Despite *P. campanulatum*, *P. tessellatum* and *P. oppositifolium* sharing similar lustrous and lanceolate leaves [2, 80], they differ in leaf arrangement, filament structure and floescence, etc. In detail, *P. campanulatum* is characterized by alternate leaves with a retrorse spur at the filament apex and flowers in October, while *P. tessellatum* and *P. oppositifolium* differ in whorled or opposite leaves without a retrorse spur at the filament apex and flower in May [2, 80]. Moreover, previous studies discover that leaf arrangement is labile and the whorled leaves have arisen from the alternate state at least twice [7, 81]. In conclusion, we infer that phyllotaxis appears not suitable as basis for delimitation of subgeneric groups in *Polygonatum*. Additionally, flower color and pollen exine sculpture were also used as the characters to subgrouping *Polygonatum* in previous studies [7, 12, 82]. Section *Polygonatum* is characterized by perforated pollen exines and greenish-white or yellow perianths, whereas *Verticillata* presented mostly reticulate pollen exines and purple or pink perianths [7, 82]. However, *P. campanulatum* placed in *Verticillata* has perforate

reticulate ornaments and yellowish-green, or greenish-white perianths [80]. The controversy of flower color has been reported in the study of Xia and her team [27]. From this we can see that flower color and pollen exine sculpture may be unrelated with phylogeny and not ideal as basis for subgenus classification of *Polygonatum* either. Moreover, further research about the information on base chromosome numbers and karyotypes of *P. campanulatum* requires to be undertaken. This work will contribute to a more insightful understanding of the infrageneric classification of *Polygonatum* and demonstrate that cp genome is an efficient tool for resolving specific level phylogeny.

5. Conclusion

In the current study, we sequenced and annotated the cp genomes of *Polygonatum campanulatum*, *P. franchetii*, *P. filipes*, *P. zanzlanscianense*, *P. cyrtoneura* and *P. sibiricum*. Comparative analyses of the chloroplast genome of the six taxa and three related species were conducted. The genome size, gene content, gene order and G-C content maintained a high similarity except that *P. sibiricum* lost a *rps19* gene due to incompletely duplicated. No interspecific or intraspecific rearrangements were detected. Eight highly variable regions were found to be potential specific DNA barcodes. Fourteen genes were revealed under positive selection and a large variety of repetitive sequences were identified. Seventy-nine protein-coding genes were utilized for phylogenetic analyses. The phylogenetic results illustrated that *Polygonatum* can be divided into two significant clades, sect. *Verticillata* and sect. *Sibirica* plus sect. *Polygonatum*. Further, *P. campanulatum* and *P. tessellatum* were strongly supported being sister relationship and located in sect. *Verticillata*, suggesting that leaf arrangement appears not suitable as basis for delimitation of subgeneric groups in *Polygonatum*. Additionally, *P. franchetii* is sister to *P. stenophyllum* within sect. *Verticillata*, too. With high morphological and karyological diversity, *Polygonatum* has attracted much attention in phylogenetic and taxonomic research. Our analysis provides more chloroplast genomic information of *Polygonatum* and contributes to improving species identification and phylogenetic studies in further work.

Abbreviations

LSC: large single-copy; SSC: small single-copy; IR: inverted repeat; BI: Bayesian inference; ML: Maximum Likelihood; PCGs: protein-coding genes; CDS: coding sequence; rRNA: ribosomal RNA; tRNA: Transfer RNA; GC: Guanine-Cytosine; SSR: simple sequence repeats; cp: chloroplast; CTAB: cetyltrimethylammonium bromide; RSCU: Relative synonymous codon usage; NGS: next-generation sequencing.

Declarations

Acknowledgments

We acknowledge Jiaxin Yang and Miao Liao for giving suggestions on the paper.

Authors' contributions

GWH collected these materials and identified species. GWH and DJZ designed and supervised the research. DJZ performed the experiment, conducted the analyses and wrote the manuscript. JR, HJ and VOW repeatedly proof-read the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Science & Technology Fundamental Resources Investigation Program of China (2019FY101800), Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0502) and National Natural Science Foundation of China (31970211)

Availability of data and materials

All the newly sequenced sequences in this study are available from the National Center for Biotechnology Information (NCBI) (accession numbers in Additional Table 1: Table S1). Information for other samples used for phylogenetic analysis download

from GenBank can be found in Additional Table 1: Table S1

Ethical approval and consent to participate

The authors have complied with the relevant institutional, national and international guidelines in collecting biological materials for the study. The study contributes to facilitating future studies in population genetics and species identification.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. APG IV. 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.
2. Chen SC, Tamura MN: *Polygonatum* Miller. *Flora of China*. Beijing & China and St Louis: Science Press & Missouri Botanical Garden Press; 2000. p. 223–232.
3. Therman E. Chromosomal evolution in the genus *Polygonatum*. *Hereditas.* 1953;39(1–2):277–288.
4. Tamura MN, Schwarzbach AE, Kruse S, Reski R. Biosystematic studies on the genus *Polygonatum* (Convallariaceae) IV. Molecular phylogenetic analysis based on restriction site mapping of the chloroplast gene *tmK*. *Feddes Repertorium.* 1997;108(3–4):159–168.
5. Pharmacopoeia Commission. *The Pharmacopoeia of the People's Republic of China*. 10th ed. Beijing, China: China Medical Science Press; 2015.
6. Zhao LH, Zhou SD, He XJ. A phylogenetic study of Chinese *Polygonatum* (Polygonateae, Asparagaceae). *Nord J Bot.* 2019;37(2).
7. 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):435–451.
8. Tamura MN, Ogisu M, Xu J-M. *Heteropolygonatum*, a new genus of the tribe Polygonateae (Convallariaceae) from West China. *Kew Bulletin.* 1997;52(4):949–956.
9. Szczecinska M, Sawicki J, Polok K, Holdynski C, Zielinski R. Comparison of three *Polygonatum* species from Poland based on DNA markers. *Ann Bot Fenn.* 2006;43(5):379–388.
10. Baker JG. Revision of the Species and Genera of Asparagaceae. *Bot J Linn Soc.* 1875;14:508–629.
11. Tang YC: *Polygonatum* Mill. In: Wang FT, Tang T, editors. *Flora Reipublicae Popularis Sinicae*. Beijing: Science Press; 1978. p. 52–80.
12. Tamura MN. Biosystematic studies on the genus *Polygonatum* (Liliaceae): III. Morphology of staminal filaments and karyology of eleven Eurasian species. *Bot Jahrb Syst.* 1993;115(1):1–26.
13. Floden AJ, Schilling EE. Using phylogenomics to reconstruct phylogenetic relationships within tribe Polygonateae (Asparagaceae), with a special focus on *Polygonatum*. *Mol Phylogenet Evol.* 2018;129:202–213.
14. Zheng X, Wang J, Li F, Liu S, Pang H, Qi L, et al. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci Rep.* 2017;7(1):1555.
15. Ravi V, Khurana JP, Tyagi AK, Khurana P. An update on chloroplast genomes. *Plant Syst Evol.* 2007;271(1–2):101–122.
16. Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, et al. Methods for obtaining and analyzing whole chloroplast genome sequences. *Method in Enzymol.* Academic Press; 2005. p. 348–384.
17. Jiang H, Tian J, Yang J, Dong X, Zhong Z, Mwachala G, et al. Comparative and phylogenetic analyses of six Kenya *Polystachya* (Orchidaceae) species based on the complete chloroplast genome sequences. *BMC Plant Biology.* 2022;22(1).

18. Palmer JD. Comparative organization of chloroplast genomes. *Ann Rev Genet* 1985;19(1):325–354.
19. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. *Methods Mol Biol.* 2014;1132:3–38.
20. Zhang YJ, Li DZ. Advances in phylogenomics based on complete chloroplast genomes. *Plant Diversity and Resources.* 2011;33(4):365–375.
21. Clegg MT, Gaut BS, Learn GH, JR., Morton BR. Rates and patterns of chloroplast DNA evolution. *P Natl Acad Sci USA.* 1994;91(15):6795–6801.
22. Yan MH, Fritsch PW, Moore MJ, Feng T, Meng AP, Yang J, et al. Plastid phylogenomics resolves infrafamilial relationships of the Styracaceae and sheds light on the backbone relationships of the *Ericales*. *Mol Phylogenet Evol.* 2018;121:198–211.
23. Yang JX, Hu GX, Hu GW. Comparative genomics and phylogenetic relationships of two endemic and endangered species (*Handeliodendron bodinieri* and *Eurycorymbus cavaleriei*) of two monotypic genera within Sapindales. *BMC Genomics.* 2022;23(1):27.
24. Xie DF, Tan JB, Yu Y, Gui LJ, Su DM, Zhou SD, He XJ. Insights into phylogeny, age and evolution of *Allium* (Amaryllidaceae) based on the whole plastome sequences. *Ann Bot.* 2020;125(7):1039–1055.
25. Zhang R, Wang Y, Jin J, Stull G, Bruneau A, Cardoso D, al. e. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Syst Biol.* 2020;69(4):613–622.
26. Liu B, Liu G, Hong D, Wen J. *Eriobotrya* belongs to *raphiolepis* (maleae, rosaceae): Evidence from chloroplast genome and nuclear ribosomal DNA data. *Front Plant Sci.* 2019;10:1731.
27. Xia MQ, Liu Y, Liu JJ, Chen DH, Shi Y, Chen ZX, et al. Out of the Himalaya-Hengduan Mountains: Phylogenomics, biogeography and diversification of *Polygonatum* Mill. (Asparagaceae) in the Northern Hemisphere. *Mol Phylogenet Evol.* 2022;169:107431.
28. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 2020;21(1):241.
29. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods.* 2019;15:50.
30. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647–1649.
31. 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.
32. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–780.
33. Mayor C, Brudno M, Schwartz J, Poliakov A, Rubin E, Frazer K, et al. VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics.* 2000;16(11):1046–1047.
34. Amiryousefi A, Hyvonen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics.* 2018;34(17):3030–3031.
35. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* 2010;5(6):e11147.
36. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016;33(7):1870–1874.
37. Gupta SK, Bhattacharyya TK, Ghosh TC. Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. *J Biomol Struct Dyn.* 2004;21(4):527–536.
38. 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):4633–4642.
39. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33(16):2583–2585.

40. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol.* 2017;34(12):3299–3302.
41. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol Evol.* 2019;9(7):3891–3898.
42. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–274.
43. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005;22(4):1107–1118.
44. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–589.
45. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 2013;30(5):1188–1195.
46. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61(3):539–542.
47. dos Reis M. How to calculate the non-synonymous to synonymous rate ratio of protein-coding genes under the Fisher-Wright mutation-selection framework. *Biol Lett.* 2015;11(4):20141031.
48. Wu Z, Liao R, Yang T, Dong X, Lan D, Qin R, et al. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). *BMC Genomics.* 2020;21(1):621.
49. Wu L, Nie L, Xu Z, Li P, Wang Y, He C, et al. Comparative and phylogenetic analysis of the complete chloroplast genomes of three *Paeonia* section moutan species (Paeoniaceae). *Front Genet.* 2020;11:980.
50. Alzahrani DA, Yaradua SS, Albokhari EJ, Abba A. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genomics.* 2020;21(1):393.
51. Singh NV, Patil PG, Sowjanya RP, Parashuram S, Natarajan P, Babu KD, et al. Chloroplast genome sequencing, comparative analysis, and discovery of unique cytoplasmic variants in *Pomegranate* (*Punica granatum* L.). *Front Genet.* 2021;12:704075.
52. Raman G, Park S. The complete chloroplast genome sequence of the *Speirantha gardenii*. Comparative and adaptive evolutionary analysis. *Agronomy.* 2020;10(9).
53. Lee SR, Kim K, Lee BY, Lim CE. Complete chloroplast genomes of all six *Hosta* species occurring in Korea: molecular structures, comparative, and phylogenetic analyses. *BMC Genomics.* 2019;20(1):833.
54. McKain MR, McNeal JR, Kellar PR, Eguiarte LE, Pires JC, Leebens-Mack J. Timing of rapid diversification and convergent origins of active pollination within Agavoideae (Asparagaceae). *Am J Bot.* 2016;103(10):1717–1729.
55. Munyao JN, Dong X, Yang JX, Mbandi EM, Wanga VO, Oulo MA, et al. Complete chloroplast genomes of *Chlorophytum comosum* and *Chlorophytum gallabatense*: genome structures, comparative and phylogenetic analysis. *Plants.* 2020;9(3):296.
56. Jia Q, Wu H, Zhou X, Gao J, Zhao W, Aziz J, et al. A "GC-rich" method for mammalian gene expression: a dominant role of non-coding DNA GC content in regulation of mammalian gene expression. *Sci China Life Sci.* 2010;53(1):94–100.
57. Du Z, Lu K, Zhang K, He Y, Wang H, Chai G, et al. The chloroplast genome of *Amygdalus* L. (Rosaceae) reveals the phylogenetic relationship and divergence time. *BMC Genomics.* 2021;22(1):645.
58. Luo C, Huang W, Sun H, Yer H, Li X, Li Y, et al. Comparative chloroplast genome analysis of *Impatiens* species (Balsaminaceae) in the karst area of China: insights into genome evolution and phylogenomic implications. *BMC Genomics.* 2021;22(1):571.
59. Wen F, Wu X, Li T, Jia M, Liu X, Liao L. The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China. *BMC Genomics.* 2021;22(1):161.
60. Dong S, Ying Z, Yu S, Wang Q, Liao G, Ge Y, et al. Complete chloroplast genome of *Stephania tetrandra* (Menispermaceae) from Zhejiang Province: insights into molecular structures, comparative genome analysis, mutational hotspots and phylogenetic relationships. *BMC Genomics.* 2021;22(1).

61. Somaratne Y, Guan DL, Wang WQ, Zhao L, Xu SQ. The complete chloroplast genomes of two *Lespedeza* species: Insights into codon usage bias, RNA editing sites, and phylogenetic relationships in Desmodieae (Fabaceae: Papilionoideae). *Plants*. 2019;9(1).
62. Xu J, Liu C, Song Y, Li M. Comparative analysis of the chloroplast genome for four *Pennisetum* species: Molecular structure and phylogenetic relationships. *Front Genet*. 2021;12:687844.
63. Park M, Park H, Lee H, Lee BH, Lee J. The complete plastome sequence of an antarctic bryophyte *Sanionia uncinata* (Hedw.) Loeske. *Int J Mol Sci*. 2018;19(3).
64. Timme RE, Kuehl JV, Boore JL, Jansen RK. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *Am J Bot*. 2007;94(3):302–312.
65. Ren J, Tian J, Jiang H, Zhu X-X, Mutie FM, Wang VO, et al. Comparative and phylogenetic analysis based on the chloroplast genome of *Coleanthus subtilis* (Tratt.) Seidel, a protected rare species of monotypic genus. *Front Plant Sci*. 2022;13.
66. Li DM, Zhao CY, Liu XF. Complete chloroplast genome sequences of *Kaempferia galanga* and *Kaempferia elegans*: Molecular structures and comparative analysis. *Molecules*. 2019;24(3).
67. Hirano R, Ishii H, Htun Oo T, Gilani SA, Kikuchi A, Watanabe KN. Propagation management methods have altered the genetic variability of two traditional *Mango* varieties in Myanmar, as revealed by SSR. *Plant Genet Resour-C* 2011;9(3):404–410.
68. Chen CX, Zhou P, Choi YA, Huang S, Gmitter FG. Mining and characterizing microsatellites from citrus ESTs. *Theor Appl Genet*. 2006;112(7):1248–1257.
69. Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol*. 2001;16(3):142–147.
70. Jiao Y, Jia HM, Li XW, Chai ML, Jia HJ, Chen Z, et al. Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genomics*. 2012;13:201.
71. Zhou T, Ruhsam M, Wang J, Zhu H, Li W, Zhang X, et al. The complete chloroplast genome of *Euphrasia regelii*, pseudogenization of *ndh* genes and the phylogenetic relationships within Orobanchaceae. *Front Genet*. 2019;10:444.
72. Gu CH, Tembrock LR, Johnson NG, Simmons MP, Wu ZQ. The complete plastid genome of *Lagerstroemia fauriei* and loss of *rp12* intron from *Lagerstroemia* (Lythraceae). *PLOS ONE*. 2016;11(3).
73. Khayi S, Gaboun F, Pirro S, Tatusova T, El Mousadiq A, Ghazal H, et al. Complete chloroplast genome of *Argania spinosa*: Structural organization and phylogenetic relationships in sapotaceae. *Plants*. 2020;9(10).
74. Dong F, Lin Z, Lin J, Ming R, Zhang W. Chloroplast genome of *Rambutan* and comparative analyses in Sapindaceae. *Plants*. 2021;10(2).
75. Liu L, Wang Y, He P, Li P, Lee J, Soltis DE, et al. Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. *BMC Genomics*. 2018;19(1):235.
76. Abdullah, Mehmood F, Rahim A, Heidari P, Ahmed I, Poczai P. Comparative plastome analysis of *Blumea*, with implications for genome evolution and phylogeny of Asteroideae. *Ecol Evol*. 2021;11(12):7810–7826.
77. Li X, Zuo Y, Zhu X, Liao S, Ma J. Complete chloroplast genomes and comparative analysis of sequences evolution among seven *Aristolochia* (aristolochiaceae) medicinal species. *Int J Mol Sci*. 2019;20(5).
78. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. *Plos Genet*. 2008;4(12):e1000304.
79. Mugal CF, Wolf JB, Kaj I. Why time matters: codon evolution and the temporal dynamics of dN/dS. *Mol Biol Evol*. 2014;31(1):212–231.
80. Cai XZ, Hu GW, Kamande EM, Ngumbau VM, Wei N. *Polygonatum campanulatum* (Asparagaceae), a new species from Yunnan, China. *Phytotaxa*. 2015;236(1):94–96.
81. Noltie HJ. Including a record of plants from Sikkim and Darjeeling. *Flora of Bhutan*. Edinburgh: Royal Botanic Garden; 1994. p. 38–46.
82. Jeffrey C. The genus *Polygonatum* (Liliaceae) in Eastern Asia. *Kew Bull*. 1980;34(3):36

Figures

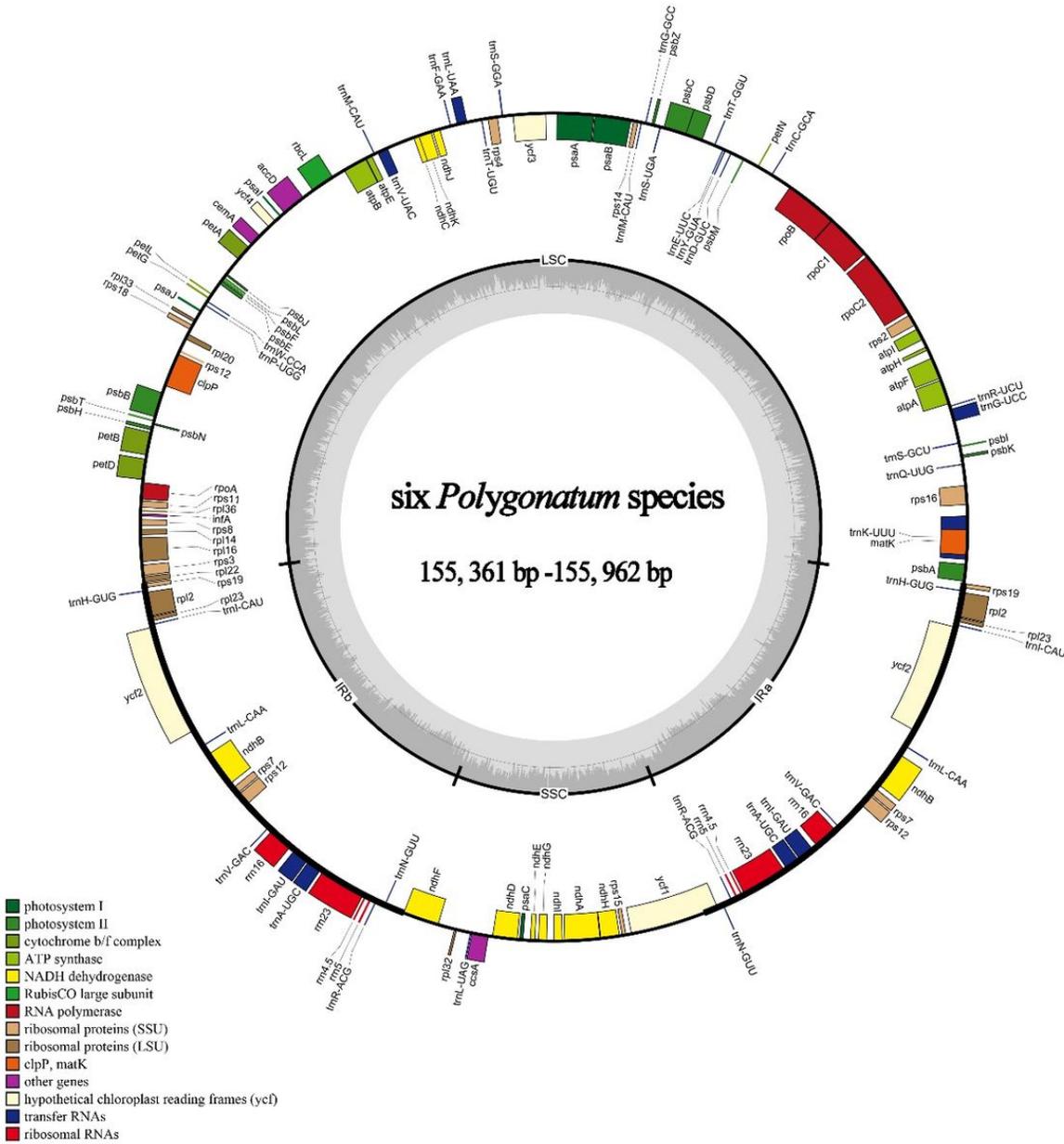


Figure 1

Gene map of the chloroplast genome among the *Polygonatum* species. Genes inside and outside the circle transcribed in counter-clockwise and clockwise respectively. The dark gray and light gray areas inside the inner circle indicates GC content and AT content respectively. LSC (Large single-copy), SSC (Small single-copy) and the inverted repeats (IRa, IRb) were denoted inner the circle.

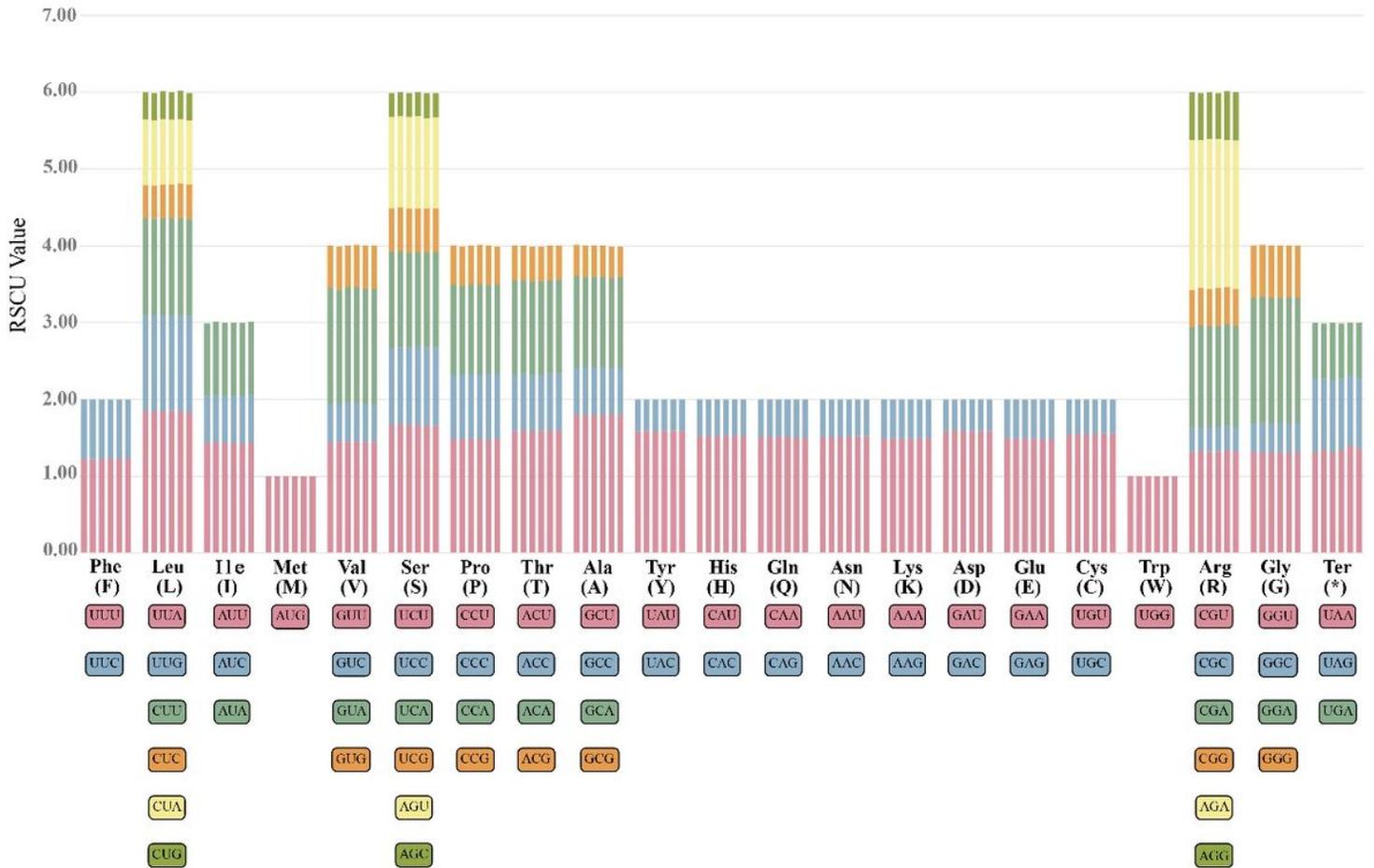


Figure 2

Relative synonymous codon usage (RSCU) value of 20 amino acids and stop codons of six chloroplast genomes of *Polygonatum* based on protein-coding sequences. The colors of the bar correspond to the colors of codons. Each amino acid corresponds to six histograms, and y-axis represents the RSCU value. The order of each six columns from left to right is *Polygonatum campanulatum*, *P. filipes*, *P. franchetii*, *P. zanscianense*, *P. cyrtoneura* and *P. sibiricum*.

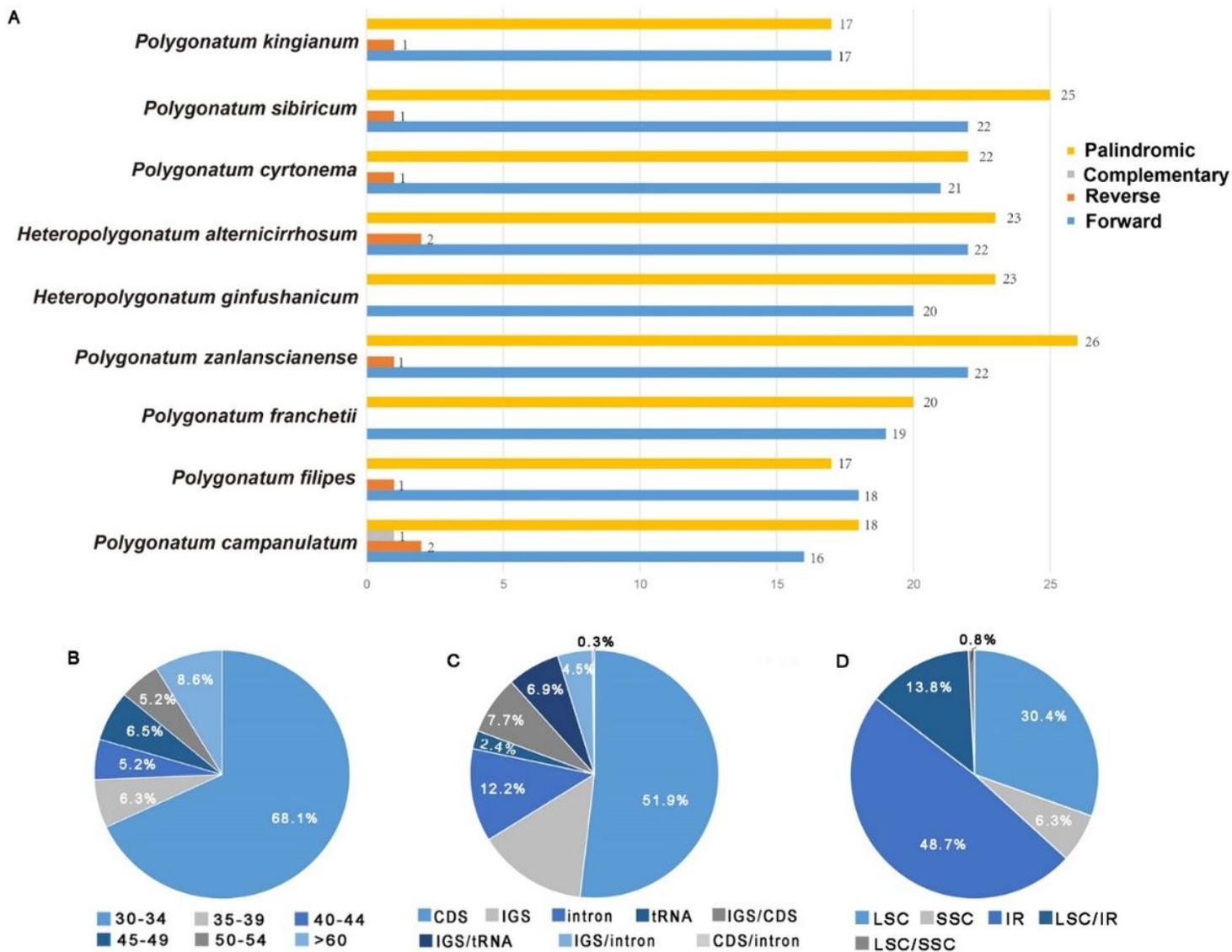


Figure 3

Analysis of long dispersed repeats in the cp genomes of seven *Polygonatum* and two *Heteropolygonatum* species. **A** The number of the four types of long repeats; **B** Distribution ratio of repeats in regions of the cp genome; **C** Distribution ratio of repetitive sequences in functional regions; **D** Proportion of repeats in different length intervals of the chloroplast genome.

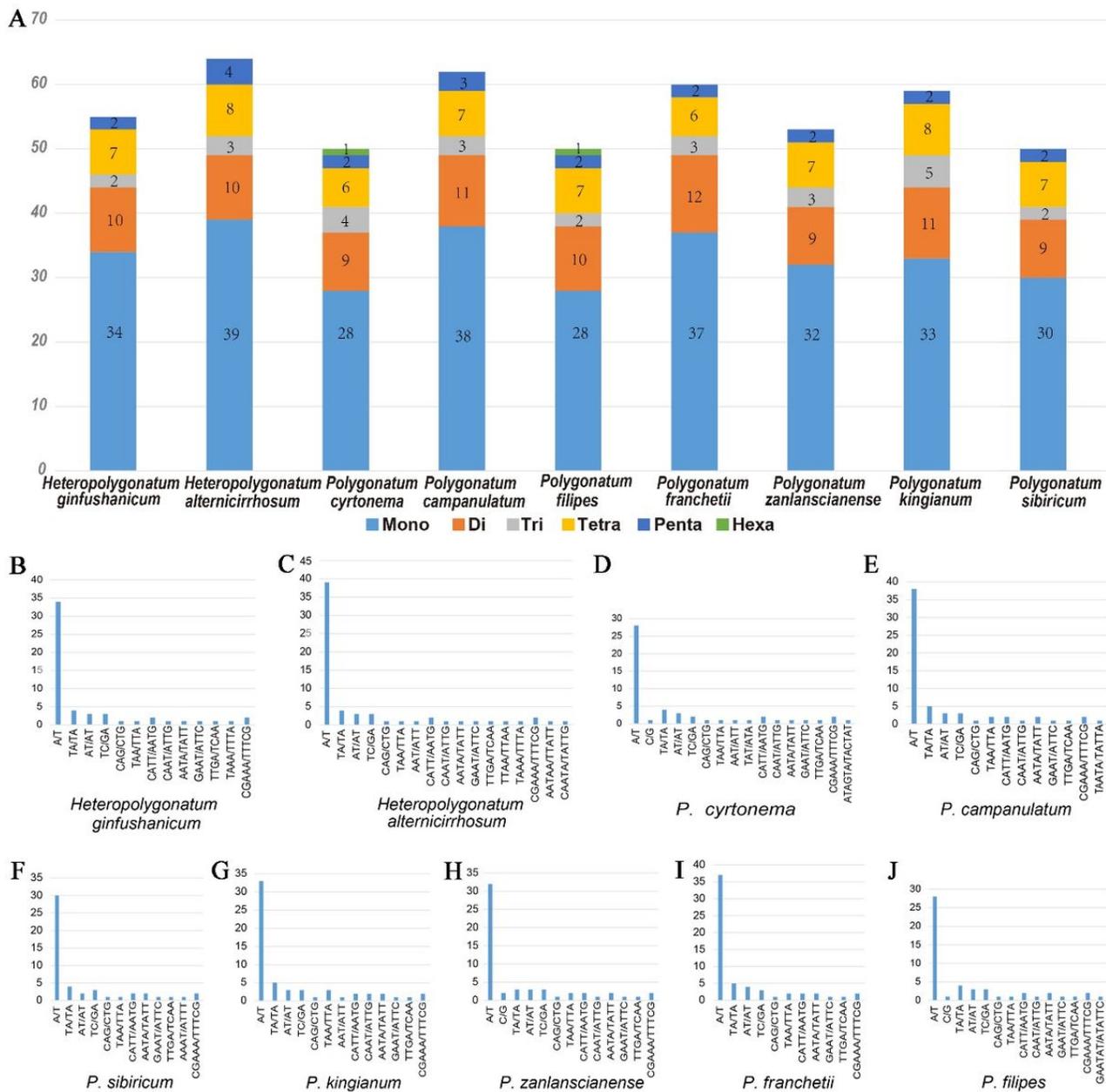


Figure 4

Simple sequence repeats (SSRs) analysis of the complete chloroplast genomes of the seven *Polygonatum* and two *Heteropolygonatum* species. (a) Number of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeats. (b-j). Frequencies of SSR motifs in different repeat class types.

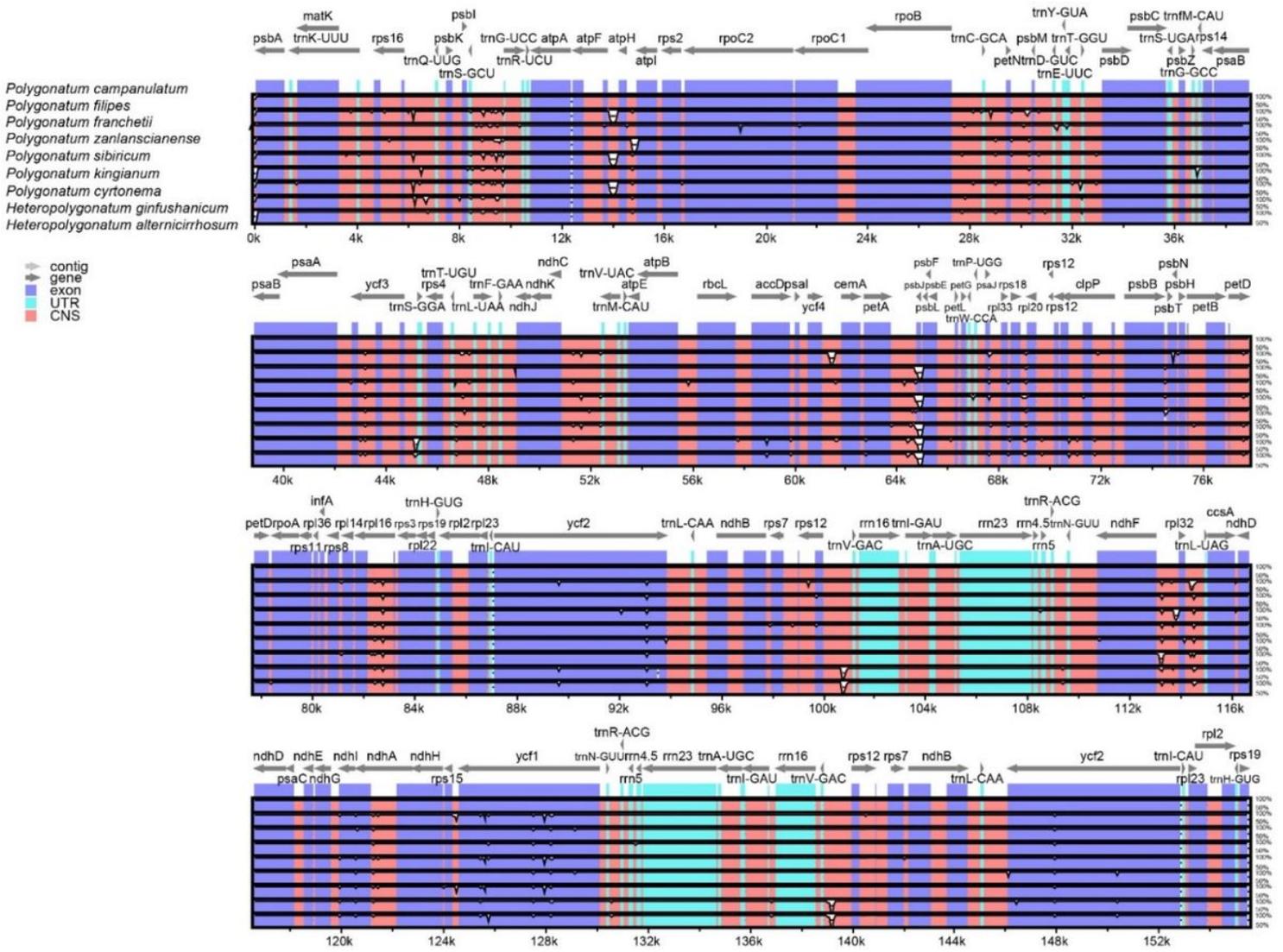


Figure 5

Alignment of chloroplast genomes of *Heteropolygonatum alternicirrhosum*, *H. ginfushanicum*, *Polygonatum campanulatum*, *P. filipes*, *P. franchetii*, *P. zanzlancianense*, *P. cyrtoneuma*, *P. sibiricum*, *P. kingianum*. The grey arrows at the top represent the direction of gene translation, the y-axis indicates the percentage identity between 50 and 100%. (exon: protein codes; UTR: tRNAs and rRNAs; CNS: conserved noncoding sequences)

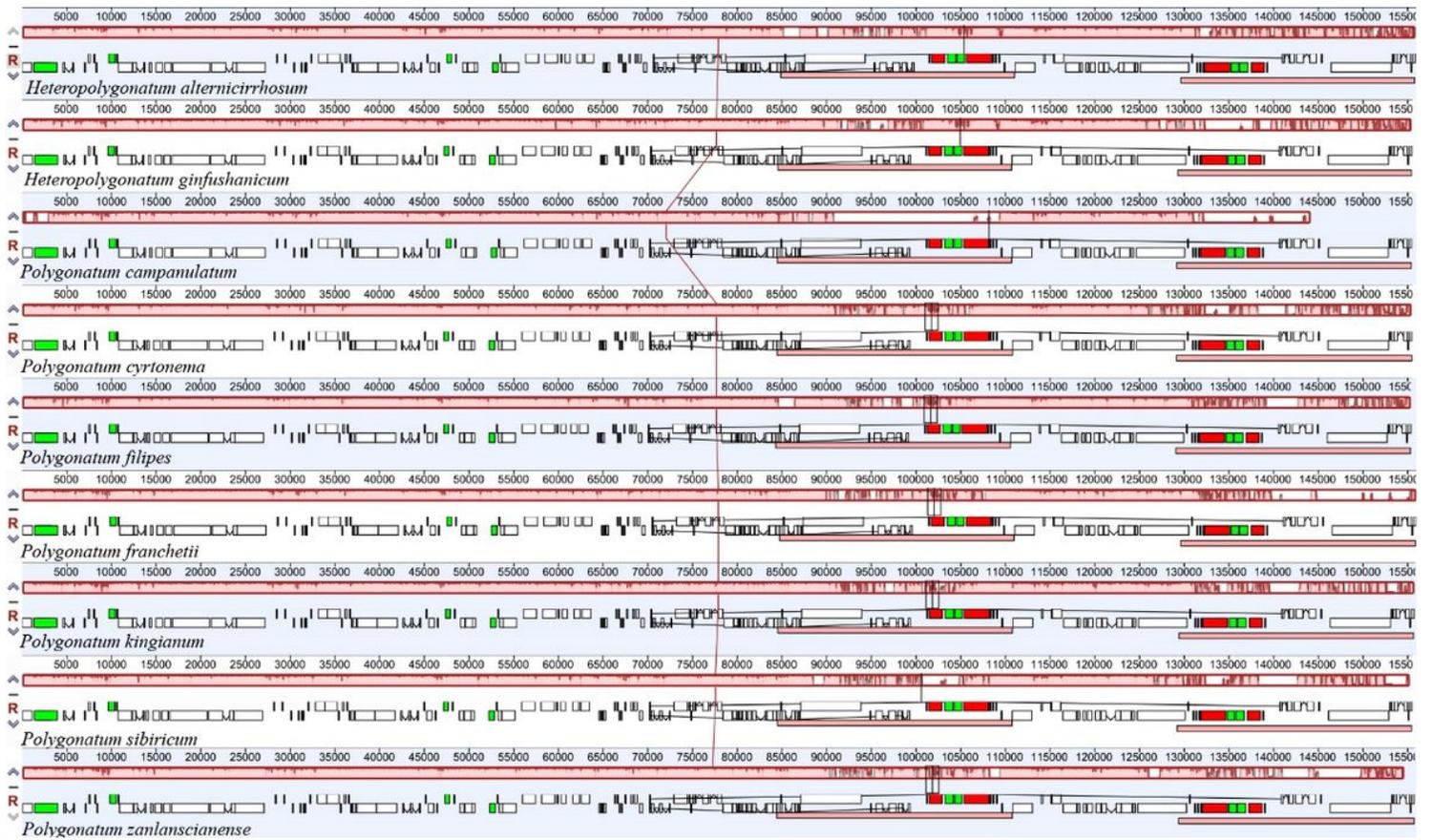


Figure 6

Genomic rearrangement of the seven *Polygonatum* and two *Heteropolygatum*. Blocks in different colors correspond to different gene types. Black: transfer RNA (tRNA); green: intron-containing tRNA; Red: ribosomal RNA; White: protein-coding genes (PCGs).

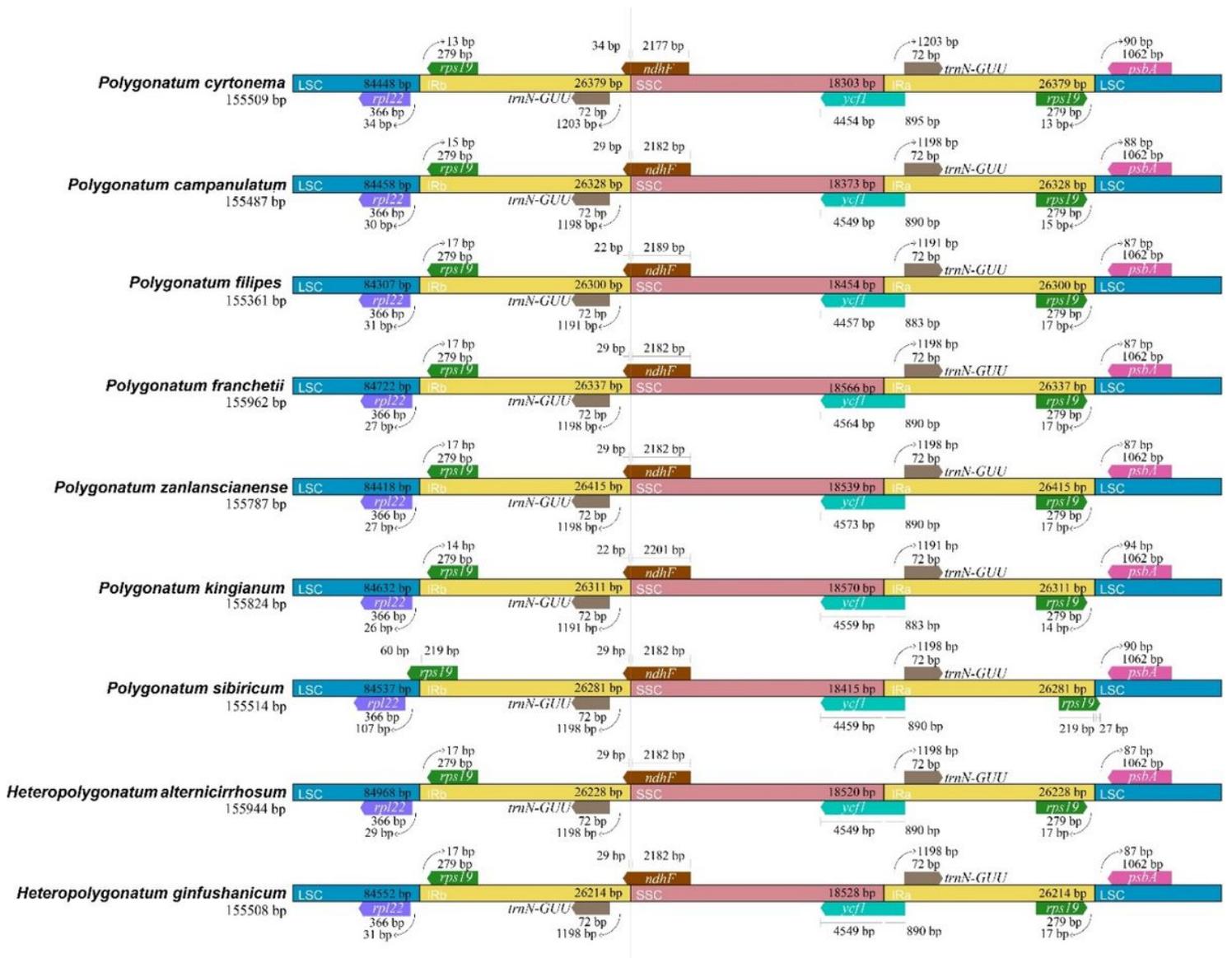


Figure 7

Comparative analysis of the LSC, IR and SSC boundary regions in the nine chloroplast genomes.

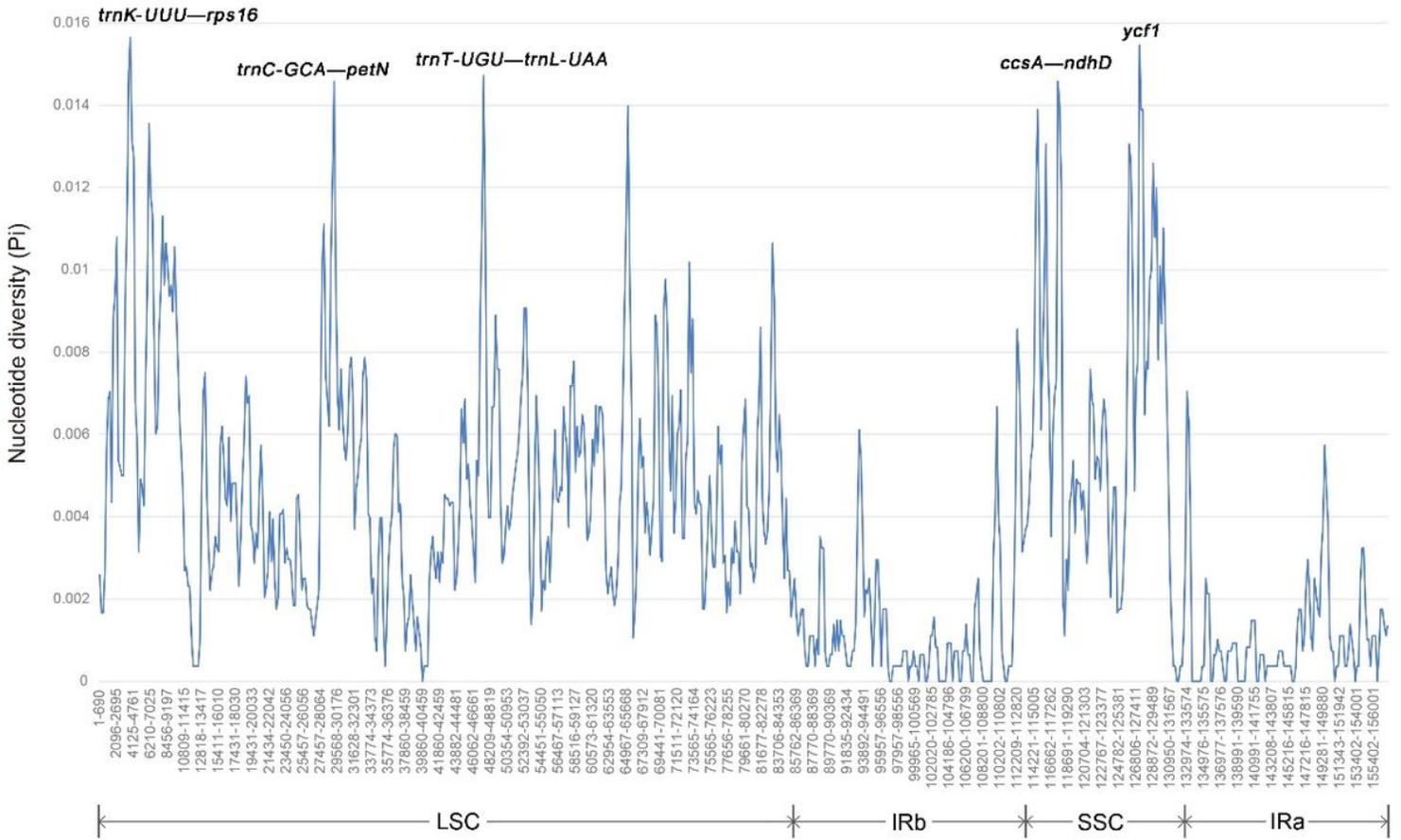


Figure 8

Nucleotide diversity analysis of the complete chloroplast genomes of the seven *Polygonatum* and two *Heteropolygatum* (window length: 600 bp; step size: 200 bp)

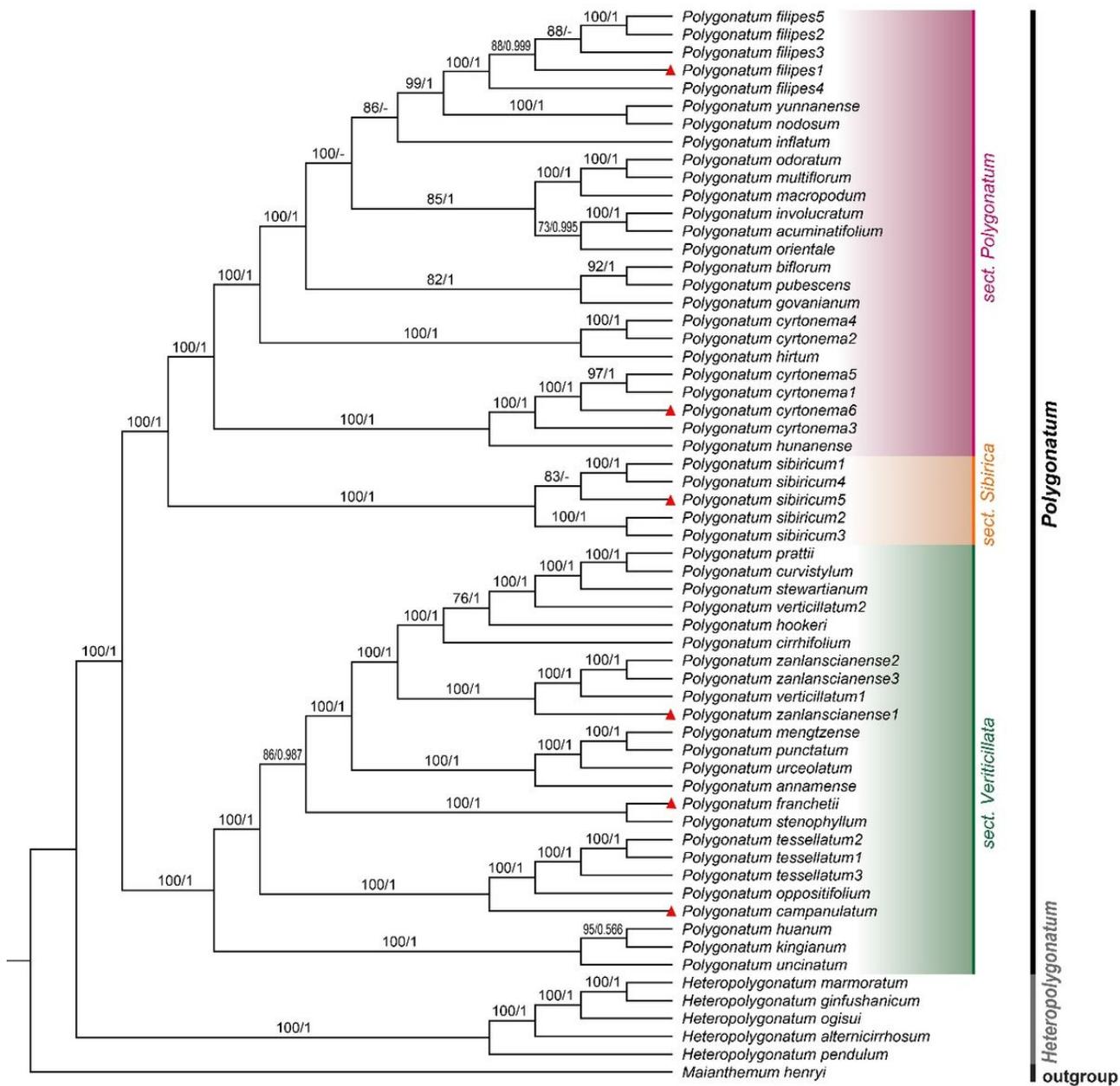


Figure 9

Phylogenetic relationships of the chloroplast genome from 60 *Polygonatum* and the related species. Maximum likelihood (ML) and Bayesian inference (BI) methods were used to reconstruct the tree, with *Maianthemum henryi* set as the outgroup. Only ML tree was shown, because of the highly identified topologies of ML tree and IQ tree. The value of ML supports and Bayesian posterior probabilities were shown above the branches. The cp genomes newly sequenced in this study is highlighted with red triangle mark.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryinformation.docx](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)

- [TableS4S7.docx](#)
- [TableS8.docx](#)
- [TableS9.docx](#)
- [TableS10.docx](#)
- [TableS11.docx](#)