



Comparative chloroplast genome analysis of *Polygala* species: insights into evolution, phylogeny, and DNA barcoding

İlayda Gülmez¹ · Ali A. Dönmez¹ · Zübeyde Uğurlu Aydın¹

Received: 30 October 2025 / Accepted: 26 February 2026 / Published online: 17 March 2026
© The Author(s) 2026

Abstract

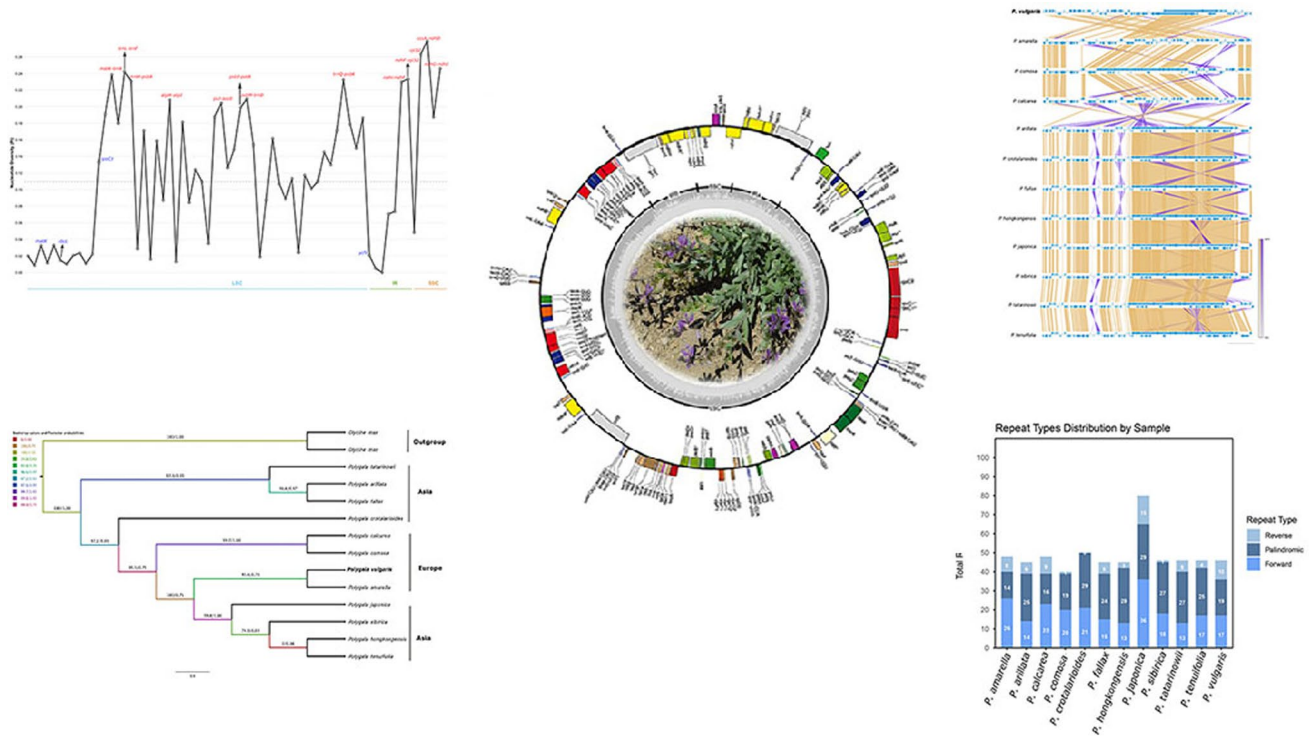
Polygala species are rich in saponin molecules and are traditionally used for medical purposes. Research on the chloroplast (cp) genome of *Polygala* is crucial to understanding its phylogenetics, biogeography, and species identification. In this study, the tetraploid *Polygala vulgaris* chloroplast genome was reported for the first time via various assembly and annotation tools. By combining these new data with the published cp genomes of 12 *Polygala* taxa from NCBI, we conducted a comparative analysis of nucleotide diversity, repeat sequences, genome synteny, phylogenetic relationships and DNA barcoding. The final assembled chloroplast genome sizes for the studied *Polygala* taxa ranged from 165,246 bp to 168,779 bp, with a total of 174 genes annotated, including 102 protein-coding genes, 63 transfer RNA genes, and 13 ribosomal RNA genes in *Polygala vulgaris*. A total of nine highly divergent genes were identified, including the CDS gene *rpl32* as well as the intergenic spacer regions *matK-trnK*, *trnL-trnF*, *trnH-psbA*, *trnQ-psbK*, *ndhI-ndhF*, *ndhF-rpl32*, *ccsA-ndhD*, and *ndhG-ndhI*. Phylogenetic relationships were reconstructed based on Bayesian inference and maximum likelihood analyses using ortholog-based and concatenated datasets derived from the identified hypervariable regions separately. The topological structure of the constructed phylogenetic tree showed biogeographic disjunction, and the most similar relative of *Polygala vulgaris* was one of the European taxa, *P. amerella*, based on high support values. The analysis of relative synonymous codon usage revealed that a codon usage pattern largely consistent with that observed across other *Polygala* species. A total of three types of repeat sequences, forward, palindromic, and reverse, were identified across the chloroplast genomes of the analyzed species. These findings provide a valuable reference for phylogenomic and evolution of *Polygala* taxa as well as to develop molecular markers for DNA barcoding.

Communicated by Stefan de Folter.

✉ İlayda Gülmez
ilaydagulmez@hacettepe.edu.tr

¹ Molecular Plant Systematic Laboratory (MOBIS),
Department of Biology, Faculty of Science, Hacettepe
University, 06800 Ankara, Türkiye

Graphical abstract



Keywords Comparative analysis · DNA barcoding · Plastome · Phylogeny · *Polygalaceae*

Introduction

The genus *Polygala*, commonly known as milkwort, is a polyphyletic assemblage and belongs to the *Polygalaceae* family (Pastore et al. 2019). *Polygala* species are native distributed worldwide, except in the Arctic and New Zealand (Eriksen and Persson 2007). Two well-supported clades of the genus are referred to as the New World Clade (NWC) and the Old World Clade (OWC) (Abbott 2009). The *Polygala* OWC is composed of more than 300 species of which 20 species are widespread in Türkiye including the recently described endemic species (Dönmez et al. 2024). Among them, *Polygala vulgaris* L. is widely distributed in Europe, and it also occurs in several locations in Türkiye. Cytogenetic data reveal that the number of *Polygala* chromosomes varies between species. However, *P. vulgaris* has been estimated to be diploid via flow cytometry (0.89 pg/2C, Horjales et al. 2003), and a tetraploid number of $2n=68$ has been reported in European and Turkish populations (Glendinning 1960; Şenova 2021).

The *Polygala* species includes several phytochemicals and is used in herbal medicines. Over the years, many phytochemicals have been successfully extracted from *Polygala* species (Çalış et al. 2022; 2023; Jing et al. 2024; Klein et al.

2012). The roots of *P. tenuifolia* (Willd.) Regel and *P. senega* L. are used for their expectorant, diuretic, and anti-inflammatory effects (Yoshikawa et al. 1995; 1996). Additionally, several studies have shown that *P. vulgaris* extracts have therapeutic benefits for neurodegenerative diseases (Çalış et al. 2023; Dall'Acqua et al. 2002).

The chloroplast (cp) is a dynamic photosynthetic organelle in green plant cells, and it is essential for metabolic activities such as photosynthesis and plant biosynthesis (Daniell et al. 2016; Neuhaus and Emes 2000; Serrano et al. 2016). The chloroplast genome represents relatively independent evolution, evolving slower than the nuclear genome but faster than the mitochondrial genome does, and it exhibits maternal inheritance. Compared with the nuclear genome, its genome is shorter and has highly conserved gene content and structure (Wicke et al. 2011). However, low rates of several chromosomal rearrangements, such as mutations, duplications, insertions/deletions, and inversions, have been reported (Amiryousef et al. 2018; Chung et al. 2006; Ghimiray and Sharma 2014). The typical angiosperm chloroplast genome is mostly circular in shape and has a quadripartite structure composed of two large copies of inverted repeats (IRa and IRb), which separate the small single-copy (SSC) and large single-copy (LSC) regions (Chung et al.

2006; Guzmán-Díaz et al. 2022; Liu et al. 2019). The chloroplast genome is between 72 and 220 kb in size and encodes 110–130 genes, with approximately eighty protein-coding genes, four rRNA genes, and thirty tRNA genes (Daniell et al. 2016).

The chloroplast genome plays an important role in plant molecular evolution and phylogenetic studies (Guzmán-Díaz et al. 2022), and it offers valuable genetic resources for complex phylogenetic relationships in angiosperms (Graham and Olmstead 2000) and polyploidization events in cultivated plants (Gornicki et al. 2014). Plastome analysis has been applied to obtain better results in the study of comparative genetic and phylogenetic relationships of polyploid plants, such as *Fragaria* L. Song et al. 2023a, b), and medicinal plants, such as *Nepeta* L. (Chen et al. 2023), *Thymus* L., and *Calendula* L. (Zhang et al. 2024), among others. Polyploid cytotypes are of particular evolutionary interest in plastid genome research because, although chloroplasts are typically maternally inherited in angiosperms, polyploidization can indirectly influence plastome evolution through cytonuclear interactions, genome dosage effects, and population-level genetic divergence (Gornicki et al. 2014; Soltis et al. 2015; 2014; Xiong et al. 2023). While the overall structure and gene content of chloroplast genomes are generally conserved across different ploidy levels, subtle differences in genome size, repeat content, mutation rates, and sequence variation between diploid and polyploid lineages have been reported in several plant groups (Li et al. 2025; Song et al. 2023a, b; 2025; Zhang et al. 2024). Moreover, polyploid lineages often exhibit broader ecological tolerance and geographic distributions, which may be reflected in plastid genome variation at the intraspecific level (Peer et al. 2017). Therefore, investigating the chloroplast genome of a tetraploid cytotype of *P. vulgaris* provides an opportunity to assess whether polyploidization is associated with detectable plastome-level variation and contributes to the evolutionary and phylogenetic interpretation of the species.

DNA barcoding, which uses short and standardized DNA fragments, was proposed as a quick and alternative identification method (Hebert et al. 2003). This genetic approach has been applied to many different taxa, and its ability to discriminate among species has been discussed (Bi et al. 2018; Chen et al. 2022; Hollingsworth et al. 2011; Uğurlu Aydın et al. 2021). Indeed, DNA barcoding have been widely used for medicinally important commercial plants to avoid misidentifications on the basis of morphological data (Chen et al. 2023). Therefore, Jeong et al. (2025) developed hot spot markers to accurate identification of two herbal medicine, *P. tenuifolia* and *P. sibirica* L. It is crucial to establish a rapid and accurate molecular method for other medical plants including *P. vulgaris*.

Currently, compared with the universal barcode region obtained from Sanger sequencing, the complete chloroplast

genome sequence provides a powerful tool for species identification (Chen et al. 2020; Wu 2021b).

Owing to high-throughput sequencing technologies, vast amounts of DNA data, including both nuclear and organelle genomes, have been generated in land plants, including *Polygala* taxa. Recently, comparative chloroplast (hereafter cp) genome analyses of several *Polygala* species have revealed nucleotide variations and genomic structures crucial for phylogenetic relationships and species identification (Jeong et al. 2025; Wang 2023).

In the present work, the cp genome of a tetraploid *P. vulgaris* individual's was analyzed for the first time via short-read DNA sequences. On the basis of cp genome sequences, we determined the genome structure, gene content, tandem repeats, and phylogenetic relationships among *Polygala* taxa. Moreover, we aimed to identify the most variable regions of the cp genome as potential DNA barcode markers and to test discrimination power of plastome-derived hot spot regions.

Materials and methods

Plant materials and DNA extraction

Fresh young leaves of *P. vulgaris* collected from Erzurum, Türkiye, were subsequently dried with silica gel. Voucher samples were stored at the HUB herbarium of Hacettepe University (A.A. Dönmez 20628). The samples' DNA was extracted with a DNeasy Plant Mini Kit (Qiagen Co., Germany). The quality and quantity of the genomic DNA were checked on a Qubit fluorometer. Three individuals of *P. vulgaris* were initially sequenced and all chloroplast genome assembly and downstream analyses were conducted using a representative individual (A.A. Dönmez 20628_1), which is therefore the only specimen reported in the Results.

Genome sequencing, assembly, and annotation

The whole-genome sequencing was carried out via the TruSeq Nano DNA Kit (Illumina, San Diego, CA, USA). Approximately 4 Gb of raw data were generated for *P. vulgaris* (A.A. Dönmez 20628_1) using 150 bp paired-end whole-genome sequencing. Based on an estimated nuclear genome size of ~1.82 Gb (Şenova 2021), this corresponds to an average whole-genome sequencing depth of approximately 2.2×.

According to the dataset, eleven *Polygala* chloroplast genomes were retrieved from NCBI as complete assembled plastomes, whereas chloroplast genomes of two additional species were assembled in this study from publicly available whole-genome sequencing (WGS) datasets (Table 1). GenBank accessions correspond to complete plastome sequences

Table 1 Comparison of the chloroplast genome data in this study

Sample	NCBI accession (GenBank/SRA)	LSC (bp)	LSC GC (%)	SSC (bp)	SSC GC (%)	IRA (bp)	IRB (bp)	IRs GC (%)	Sequence length (bp)	Total GC (%)	Protein coding gene (CDS)	tRNA	rRNA
<i>P. vulgaris</i>	SRR32576552	83,543	37,9	7994	29,5	36,736	36,736	31,9	165,009	36,6	102	63	13
<i>P. comosa</i> <i>Schkuhr</i>	ERX13411377	87,438	35,0	7863	29,8	34,525	34,525	39,8	164,351	36,8	107	66	14
<i>P. amarella</i> Crantz	ERR5554716	133,648	37,9	7840	29,9	11,568	11,568	32,1	164,624	36,7	77	39	8
<i>P. calcarea</i> F. W. <i>Schultz</i>	ERR14043099	133,368	38,0	7873	29,8	11,547	11,547	32,1	164,335	36,8	77	39	8
<i>P. tenuifolia</i>	NC_050829	83,696	34,7	8044	29,4	36,840	36,840	39,7	165,42	36,7	100	66	14
<i>P. arillata</i> Buch.- <i>Ham ex D. Dib</i>	SRX22362821	83,537	34,9	8210	30,1	36,600	36,600	39,8	164,947	36,8	103	66	10
<i>P. crotalari-</i> <i>oides</i> B. -Ham. <i>ex DC.</i>	NC_060367	86,873	35,1	8246	29,2	34,574	34,575	39,8	164,268	36,8	102	67	14
<i>P. fallax</i> Chodat	NC_052911	83,352	35,0	8283	29,7	36,526	3656	39,9	164,687	36,9	100	66	14
<i>P. hongkongen-</i> <i>sis</i> Hemsl. <i>ex</i> <i>F.B. Forbes &</i> <i>Hemsl.</i>	NC_066219	83,707	34,8	8035	29,6	36,777	36,777	39,7	165,296	36,7	101	66	14
<i>P. japonica</i> Houtt.	NC_052912	88,128	34,9	8084	29,5	34,517	34,517	39,8	165,246	36,7	101	66	14
<i>P. sibirica</i>	NC_056970	83,668	34,8	8095	29,5	36,742	36,742	39,7	165,247	36,7	101	66	14
<i>P. tata-</i> <i>rinowii</i> Regel	NC_086943	73,712	34,9	8213	29,0	43,427	43,427	39,1	168,779	36,8	111	67	14

The new sequence from this study is in boldface

downloaded directly from NCBI, while SRA accessions refer to raw WGS data used for de novo chloroplast genome assembly using GetOrganelle (Jin et al. 2020). Low-quality reads and adapters were filtered from the raw data via Trimmomatic (Bolger et al. 2014) with the default settings. The plastid assembly was performed from the clean reads via a reference-based method using the program GetOrganelle, with k -mers of 21, 45, 65, 85, and 105. Gene annotation of the cp genomes was performed via GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>). Protein-coding sequences (CDSs) were extracted manually from Geneious (Kearse et al. 2012) and aligned with MAFFT (Katoh et al. 2002). For orthology-based comparisons among taxa, annotated protein-coding genes were translated, and orthogroups were inferred using OrthoFinder (Emms and Kelly 2019). Where necessary, protein sets were standardized using AUGUSTUS-based gene prediction to ensure consistent input for orthogroup inference (Stanke et al. 2006). tRNA genes were identified via tRNAscan-SE 1.21 (Schattner et al. 2005). All *Polygala* taxa cp maps were obtained via OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>).

Analysis of repeats

We identified forward (F), reverse (R), and palindromic (P) repeat sequences via the online tool REPuter v1.0 (Kurtz et al. 2001). The analysis was conducted with conditions specifying a minimum repeat size of 8 base pairs and a maximum of 50 computed repeats. Dispersed repeat analysis performed using REPuter, tandem repeats and simple sequence repeats (SSRs) were automatically identified using

the cpGAVAS annotation pipeline, which integrates the MISA (MISatellite identification tool; Thiel et al. 2003) algorithm. Mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide SSRs were detected using the default parameters of cpGAVAS.

Comparison of the chloroplast genome

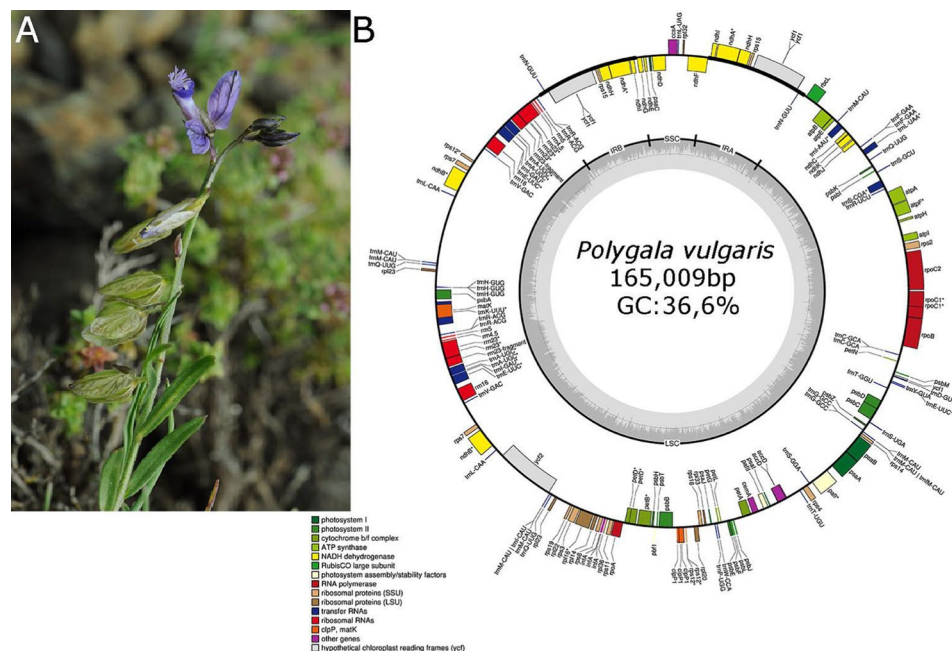
We extracted conserved CDSs and some gene regions via Geneious software. The nucleotide diversity (π) value between the *Polygala* cp genomes was subsequently calculated via DnaSP ver.6 (Librado and Rozas 2009) with a sliding window analysis. The step size was set to 50 bp, and the window length was set to 600 bp. Plots are generated with ggplot2 via R. Using pyGenomeViz (<https://github.com/moshi4/pyGenomeViz>), we created a synteny plot employing the MMseqs RBH mode and setting an identity threshold of 50%.

A comparative visual analysis of the boundaries of inverted repeat (IR) sequences was performed via IRscope Plus (<https://irsco.pe.shinyapps.io/IRplus/>). Additionally, relative synonymous codon usage (RSCU) analysis of the chloroplast genome of *Polygala* taxa was performed via the codonW (<https://galaxy.pasteur.fr/?form=codonw>) online software.

Phylogenetic reconstruction

A Bayesian inference (BI) phylogenetic tree was constructed via MrBayes v3.2.7 (Ronquist et al. 2012), which employs the Markov chain Monte Carlo (MCMC) method. Four

Fig. 1 **A** General view of *Polygala vulgaris* flower and fruit. **B** Physical map of the *P. vulgaris* chloroplast genome. Colored boxes represent genes grouped by functional categories. The innermost grey circle indicates GC content (%)



chains (three heated and one cold) were run for 10,000,000 generations, with a sampling frequency of 1000 and a printing frequency of 10,000. A burn-in of 2500 samples (corresponding to 25% of the total) was applied to ensure convergence. A 50% majority-rule consensus tree was generated to summarize the posterior distribution of trees.

Second, a maximum likelihood (ML) tree was generated via IQ-TREE v2.2.0 (Nguyen et al. 2015) under the best-fit model automatically selected via ModelFinder (cpREV + R2). Support for the nodes was assessed via 1000 ultrafast bootstrap replicates. Finally, both the BI and ML trees were compared, and the final consensus tree was annotated to include both posterior probabilities (from MrBayes) and bootstrap support values (from IQ-TREE). FigTree v1.4.4 (<https://tree.bio.ed.ac.uk/software/figtree/>) was used to visualize and annotate the combined tree topology. Nine highly variable chloroplast regions, (*ccsA-ndhD*, *matK-trnK*, *ndhF-rpl32*, *ndhG-ndhI*, *ndhI-ndhF*, *rpl32*, *trnH-psbA*, *trnL-trnF*, and *trnQ-psbK*), were extracted from chloroplast genomes and aligned individually using Geneious Prime. Sequence alignments were inspected manually to ensure positional homology, particularly within intergenic spacer regions. To infer species-level phylogenetic relationships, the nine aligned barcode regions were concatenated into a single supermatrix using AMAS v1.0. (Borowiec 2016). Maximum likelihood phylogenetic inference based on the concatenated dataset was performed using IQ-TREE. Partitioned analyses were conducted by applying the partition scheme produced by AMAS, allowing each gene region to evolve under its own substitution model. The

optimal substitution model for each partition was selected automatically using ModelFinder implemented in IQ-TREE. Branch support was assessed using 1000 ultrafast bootstrap replicates. The resulting phylogenetic tree was visualized using FigTree.

Results

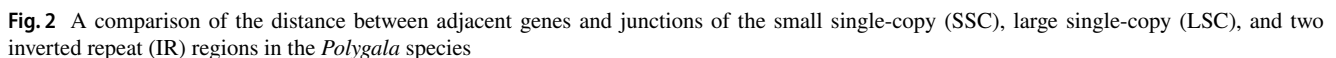
We obtained approximately 4 Gb of trimmed paired-end reads. The chloroplast genome was then generated through de novo assembly. The completed circular genome length was 165,009 bp (Fig. 1). The tetraploid *P. vulgaris* cp genome has a typical quadripartite structure similar to that of angiosperm plastid genomes, consisting of an LSC (83,543 bp) and SSC (7,994 bp), separated by a pair of identical IRs (11,636 bp and 36,736 bp). The overall guanine–cytosine (GC) content of the *P. vulgaris* cp genome is 36.6%, which is nearly identical to that of other *Polygala* cp genomes, although it is unevenly distributed within the cp genome. Moreover, the highest GC content was in the LSC region (37.9%), and the GC contents in the SSC and IR regions were 29.5% and 31.9%, respectively (Table 1).

The chloroplast genome of *P. vulgaris* is 165,009 bp in length and contains a total of 174 genes, including 102 protein-coding genes, 63 transfer RNA (tRNA) genes, and 13 ribosomal RNA (rRNA) genes. Among these genes, six protein-coding genes (*atpF*, *rpoC1*, *rpl16*, *rpl2*, *ndhA*, and *ndhB*) and seven tRNA genes (*trnK-UUU*, *trnL-UAA*, *trnV-UAC*, *trnI-GAU*, *trnA-UGC*, *trnE-UUC*, and *trnS-CGA*)

Table 2 List of genes annotated in the chloroplast genomes of twelve *Polygala* species

Group of genes	Name of genes
Photosystem I	<i>psaA</i> , <i>B</i> , <i>C</i> , <i>I</i> , <i>J</i> , <i>pafI</i> **, <i>pafII</i>
Photosystem II	<i>psbA</i> , <i>B</i> , <i>C</i> , <i>D</i> , <i>E</i> , <i>F</i> , <i>H</i> , <i>I</i> , <i>J</i> , <i>K</i> , <i>L</i> , <i>M</i> , <i>T</i> , <i>Z</i> , <i>pbfI</i>
Cytochrome b6/f	<i>petA</i> , <i>B</i> , <i>D</i> (× 2), <i>G</i> , <i>L</i> , <i>N</i>
ATP synthase	<i>atpA</i> , <i>B</i> , <i>E</i> , <i>F</i> *, <i>H</i> , <i>I</i>
RubisCO large subunit	<i>rbcL</i>
NADH oxidoreductase	<i>ndhA</i> * (× 2), <i>B</i> *, <i>C</i> , <i>D</i> , <i>E</i> , <i>F</i> , <i>G</i> , <i>H</i> , <i>I</i> , <i>J</i> , <i>K</i>
Large subunit ribosomal proteins	<i>rpl2</i> * (× 2), <i>14</i> , <i>16</i> *, <i>20</i> , <i>23</i> , <i>32</i> , <i>33</i> , <i>36</i>
Small subunit ribosomal proteins	<i>rps2</i> , <i>3</i> , <i>4</i> , <i>7</i> , <i>8</i> , <i>11</i> , <i>12</i> ** (× 2), <i>14</i> , <i>15</i> , <i>18</i> , <i>19</i>
RNA polymerase	<i>rpoA</i> , <i>B</i> , <i>C1</i> * (× 2), <i>C2</i>
Unknown function protein coding gene	<i>ycf1</i> (× 2), <i>ycf2</i>
Other genes	<i>accD</i> (× 2), <i>ccsA</i> , <i>cemA</i> , <i>clpP</i> , <i>infA</i> (× 2), <i>matK</i>
Ribosomal RNAs	<i>rrn4.5</i> , <i>rrn5</i> , <i>rrn16</i> , <i>rrn23</i> *
Transfer RNAs	<i>trnA-UGC</i> * (× 2), <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> *, <i>trnF-GAA</i> , <i>trnM-CAU</i> , <i>trnG-GCC</i> , <i>trnH-GUG</i> , <i>trnI-CAU</i> , <i>trnI-GAU</i> *, <i>trnK-UUU</i> *, <i>trnL-CAA</i> , <i>trnL-UAA</i> *, <i>trnL-UAG</i> , <i>trnM-CAU</i> (× 2), <i>trnN-GUU</i> (× 2), <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> , <i>trnR-UCU</i> , <i>trnS-CGA</i> *, <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> , <i>trnV-UAC</i> *, <i>trnW-CCA</i> , <i>trnY-GUA</i>

The genes marked with the sign are the genes with one (*) or two (**) introns and duplicated genes (× 2)



Contraction and expansion of inverted repeats

In *P. vulgaris*, the IRb/SSC (JSB) junction was located within the *ndhH* gene, extending 2,417 bp into the SSC region, whereas the IRa/SSC (JSA) junction was positioned within *ycfI*, overlapping by 7,994 bp. Compared with other *Polygala* species, *P. vulgaris* presented a relatively longer LSC region (83,743 bp), but the IR regions (11,838 bp) were shorter than those observed in most congeners. These differences suggest lineage-specific expansions and contractions of IR boundaries in *P. vulgaris*, whereas the overall IR/SC organization remained conserved across the genus.

The relative synonymous codon usage (RSCU) analysis conducted in this study revealed that *P. vulgaris* presented a codon usage pattern largely consistent with that of the

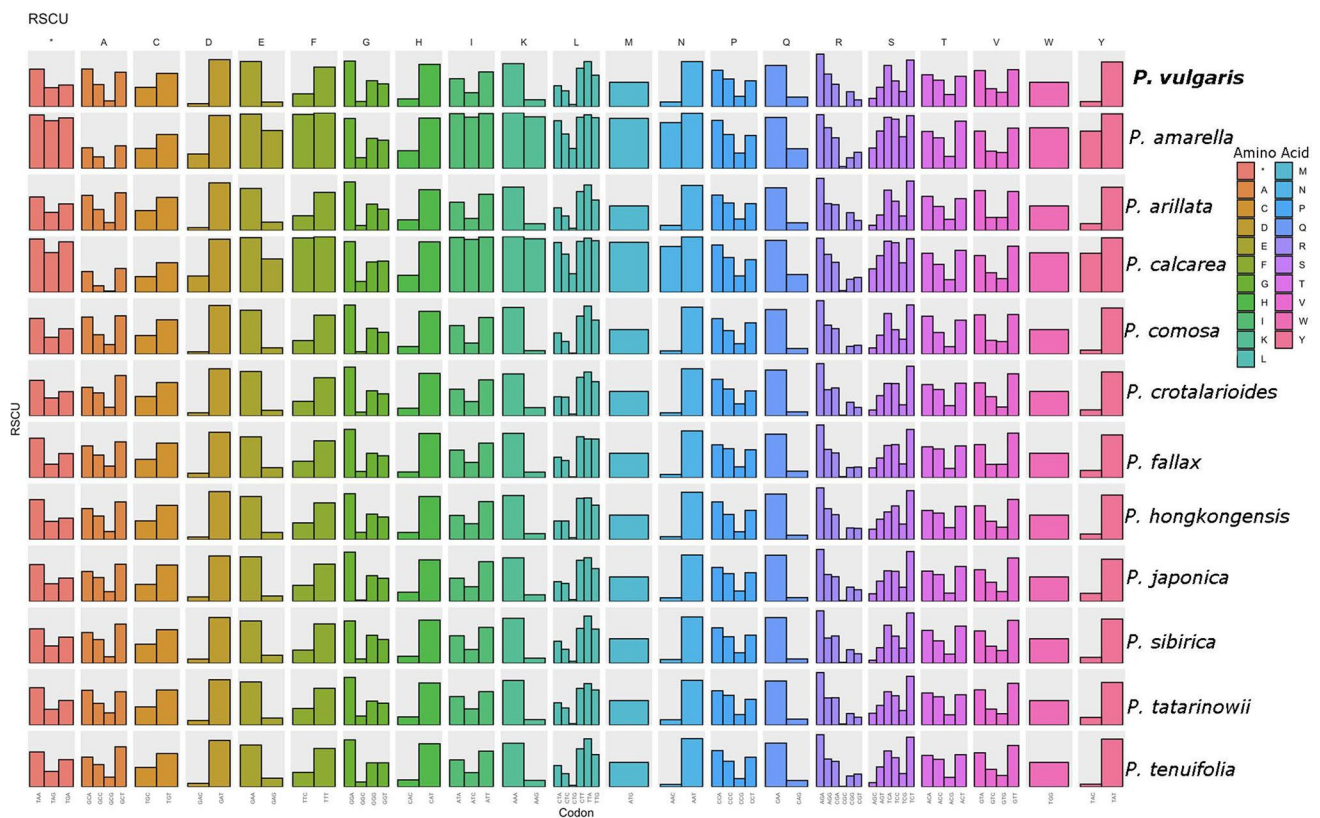


Fig. 3 Relative synonymous codon usage (RSCU) in all *Polygala* taxa for this study

other *Polygala* species examined (Fig. 3). While overall codon preferences are conserved across the genus, certain codons display species-specific variations in usage frequency. Our findings suggest that codon usage patterns in *Polygala* chloroplast genomes are shaped by both conserved evolutionary constraints and species-specific selective forces.

RSCU analysis revealed similar codon usage patterns among all the examined *Polygala* taxa (Fig. 3). Although overall trends were largely conserved, certain codons showed clear preferences, reflecting species-specific codon bias. Our own sample displayed codon usage patterns consistent with those of other *Polygala* species, suggesting a conserved evolutionary codon preference within the genus.

The nucleotide diversity (Pi value) within the sliding window of 600 bp was calculated to clarify hotspot regions in the genome (Fig. 4). The Pi value in the windows varied from 0.00509 to 0.27864, with a mean of 0.10970. Nine highly variable regions ($Pi > 0.22$) were identified in the *Polygala* genome. Among these regions, *matK-trnK*, *trnL-trnF*, *trnH-psbA*, and *trnQ-psbK* were located in the LSC region, and *ndhI-ndhF* and *ndhF-rpl32* were located in the IR region. The remaining three regions, *rpl32*, *ccsA-ndhD*, and *ndhG-ndhI*, were nested in the SSC region. Furthermore, the *ccsA-ndhD* region was the most highly divergent

region (Pi value = 0.27864). Some of the universal DNA barcode markers, such as *matK* and *rbcL*, presented lower Pi values, whereas *trnL-trnF* and *trnH-psbA* showed greater nucleotide variation.

The analysis revealed extensive collinearity across the plastomes, as indicated by the numerous yellow links, which represent conserved gene blocks (Fig. 5). Despite the overall conservation, several inverted regions (purple links) were detected, particularly in comparisons with *P. calcarea*, *P. arillata*, and *P. sibirica*, suggesting lineage-specific rearrangements. The gene features (blue) showed a largely consistent distribution across all species, reflecting the overall stability of plastid genome organization in the genus. Taken together, these results indicate that *Polygala* plastomes are highly syntenic, with only localized inversion events contributing to structural variation among species.

In order to create robust phylogeny of *Polygala* taxa, a phylogenetic tree was established using orthologous genes of the chloroplast genome (Fig. 6A). Moreover, discrimination power of hypervariable regions, suggested in this study, were tested based on ML and BI analyses (Fig. 6B). Two *Glycine* species were nested with high support values and served as outgroups. The tree topology of ML and BI analyses were identical, and the analyses were merged into a one tree.

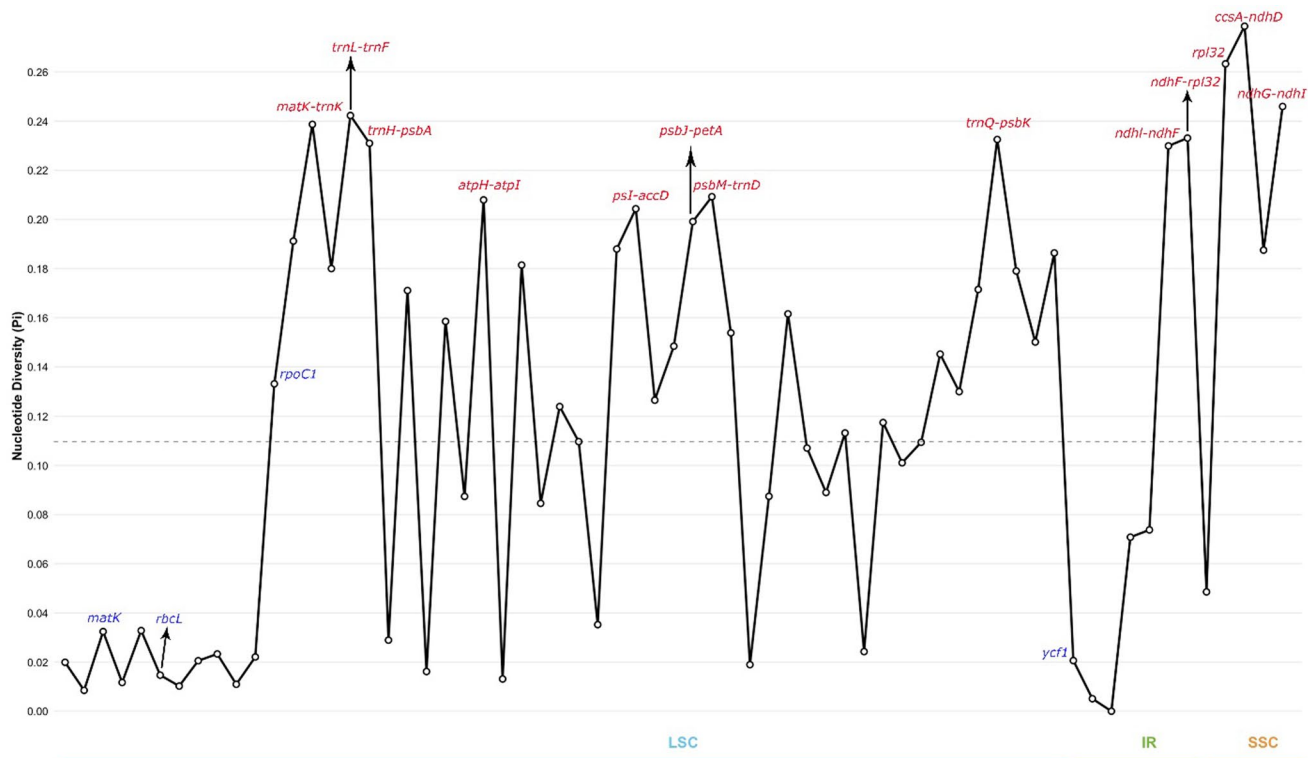


Fig. 4 Nucleotide diversity (Pi) analysis of chloroplast genomes in *Polygala* species. The dashed line represents the mean Pi value (0.10970328), Window length: 600 bp, step size 50 bp

Phylogenetic tree based on orthologues revealed that tetraploid *P. vulgaris* clustered with other morphologically similar diploid European species, *P. amarella* (PP=0.76, MLBS=93.4%). Clustering revealed that *P. vulgaris* was distinct from its sister groups, *P. japonica*, *P. sibirica*, *P. hongkongensis*, and *P. tenuifolia* (PP=0.75, MLBS=100%). Two other European taxa, *P. calcarea* and *P. comosa*, nested as sister clades to *P. vulgaris* and *P. amarella* (PP=0.75, MLBS=99.9%).

Based on the nucleotide diversity analysis, nine highly variable chloroplast regions were aligned and concatenated into a single dataset, which comprised all sampled taxa with consistent sequence representation across regions and showed no evidence of chimeric or duplicated sequences. Maximum likelihood and Bayesian analysis of the concatenated barcode dataset produced a well-resolved topology with strong support values for most internal nodes. The resulting tree topology was largely congruent with the phylogenomic tree inferred from whole chloroplast genes. Tetraploid *P. vulgaris* clustered with the European taxa *P. amarella*, *P. calcarea*, and *P. comosa*, forming a strongly supported clade. Asian taxa (*P. tenuifolia*, *P. sibirica*, *P. japonica*, and *P. hongkongensis*) were recovered as a separate lineage with high bootstrap support.

A total of three types of repeat sequences, forward, palindromic, and reverse, were identified across the chloroplast genomes of the analyzed species (Fig. 7). The total number of repeats varied among the species, ranging from 40 in *P. comosa* to a maximum of 80 in *P. japonica*. Among the repeat types, forward repeats were the most abundant in the majority of the species. The highest number of forward repeats (36) was observed in *P. japonica*, followed by *P. amarella* (26) and *P. calcarea* (23). Palindromic repeats were also commonly found, with the highest counts in *P. japonica* (29), *P. hongkongensis* (29) and *P. crotalariaoides* (29). Reverse repeats were relatively less common across all the species, with a maximum of 15 detected in *P. japonica*. Also, a total of 118 to 157 simple sequence repeats (SSRs) were identified across the analyzed *Polygala* plastomes using the cpGAVAS-MISA pipeline (Fig. 8). Mononucleotide repeats were the most abundant class in all species, ranging from 106 in *P. sibirica* to 143 in *P. calcarea*, and accounting for the vast majority of detected SSRs. Dinucleotide repeats were represented by 9 to 16 motifs depending on the species, whereas trinucleotide repeats were rare and varied between 0 and 3. Tetra-, penta-, and hexanucleotide SSRs were largely absent from most plastomes; only *P. tenuifolia* harbored a single tetranucleotide repeat. Overall, A/T mononucleotide

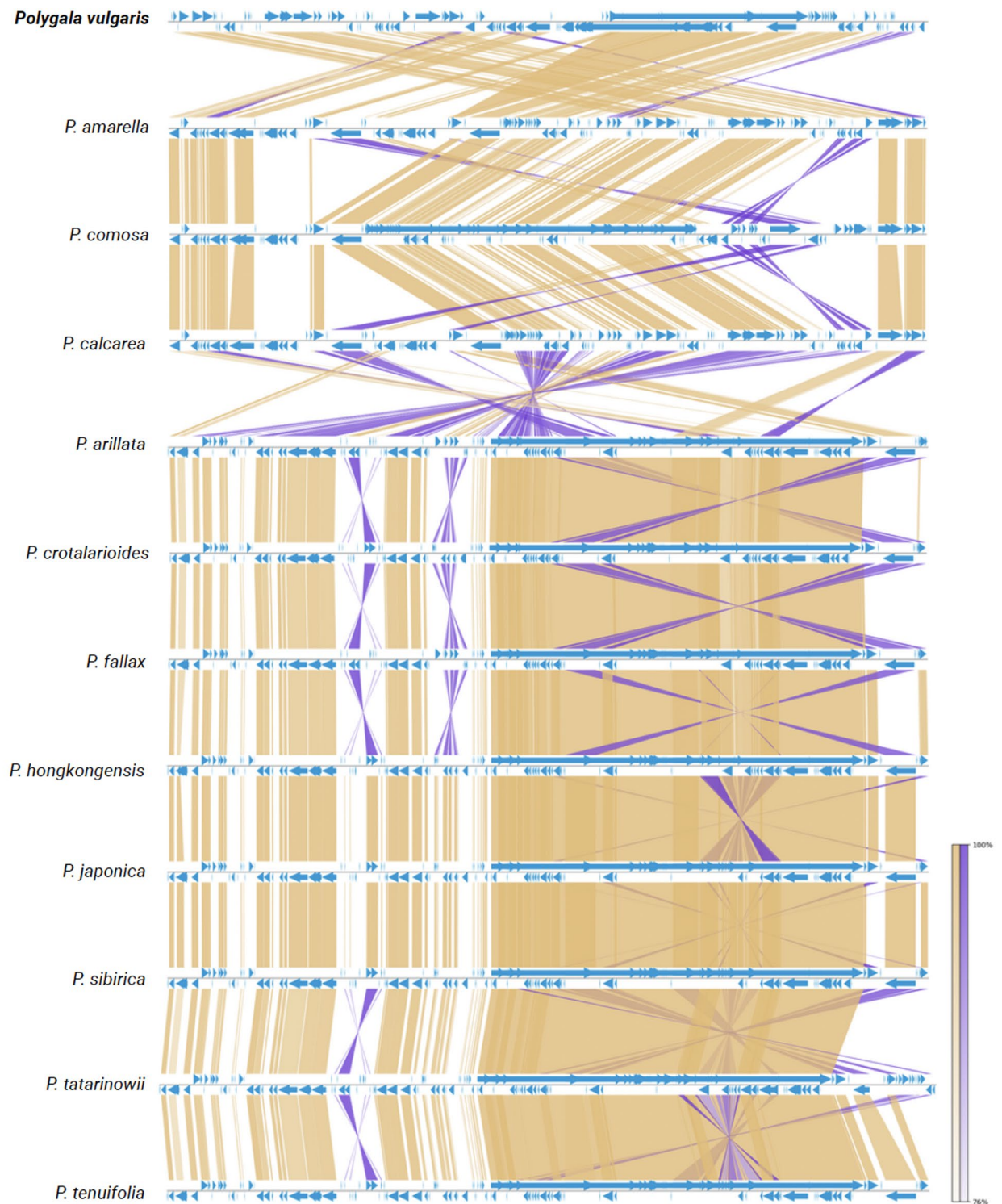


Fig. 5 Synteny plot of *Polygala vulgaris* and eleven other plastomes of some *Polygala* taxa. The synteny plot shows normal links in yellow, inverted links in purple, and gene features in blue

motifs dominated all plastomes, followed by a smaller proportion of AT/TA dinucleotide repeats.

Compared with the other taxa, *P. japonica* presented the highest repeat content overall, which may suggest a greater degree of genome rearrangement or structural variation in its

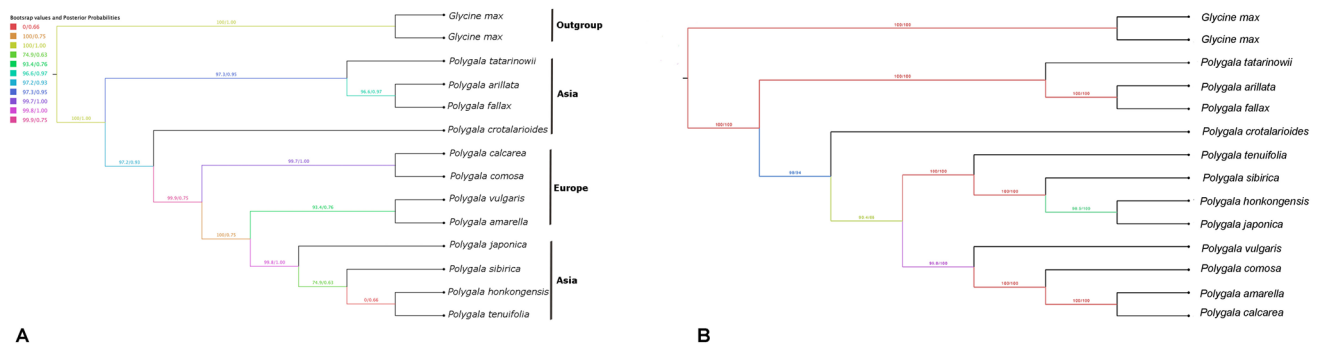


Fig. 6 Phylogenetic trees of studied *Polygala* taxa. **A** Phylogenetic tree based on orthologous genes of chloroplast genome. **B** Phylogenetic tree based on hypervariable regions of chloroplast genome. The

number above each node represents the bootstrap values (left) and posterior probabilities (right)

chloroplast genome. In contrast, species such as *P. comosa* presented the lowest number of repeats, indicating a more conserved chloroplast genome architecture.

Discussion

In this study, we utilized next-generation methods to sequence, assemble, and characterize the first complete plastid genome of *P. vulgaris*. The results revealed that the cp genome size of the studied *Polygala* species ranged from 164,268 in *P. crotarioides* to 168,779 bp in *P. tatarinowii* (Table 1). Compared with the cp genomes of other *Polygala* species (Jeong et al. 2025; Lee et al. 2020; Wang 2023), the cp genome of *P. vulgaris* shares typical features with a typical quadripartite structure, with the LSC and SSC regions separated by two IRs.

Genome size variation and gene order are relatively conserved within terrestrial plants. Significant plastome size variation and structural rearrangements have been reported in the closely related family *Fabaceae*, primarily driven by expansion, contraction, or loss of the inverted repeat (IR) region (Li et al. 2025; Perry and Wolfe 2002; 2025). Similar patterns of IR loss have also been documented in other angiosperm lineages, including genera within *Arecaceae* (e.g., *Phoenix* and *Cocos*; Barrett et al. 2016), *Fabaceae* (e.g., *Medicago*, *Pisum*, and *Trifolium*; Choi et al. 2019), and *Passifloraceae* (e.g., *Passiflora*; Cauz-Santos et al. 2020). In *Fabaceae*, repeated loss and secondary regain of the IR region have been linked to extensive plastome rearrangements, accelerated substitution rates, and instability of genes located near IR boundaries, particularly members of the *ndh* and *ycf* genes. In contrast, all *Polygala* plastomes analyzed in this study exhibit a typical quadripartite structure with two intact IR regions, and no evidence of IR contraction, expansion, or loss was detected. Genes that are frequently affected by IR boundary shifts in other angiosperms, such as *ndhF*,

ycf1, and *ycf2*, showed conserved positions and intact coding sequences across all examined *Polygala* species. The maintenance of IR structure in *Polygala* therefore indicates a high level of plastome stability. It suggests that the evolutionary dynamics observed in IR-labile families such as *Fabaceae* are not mirrored in *Polygalaceae*. Because gain and loss of the IR region are considered evolutionarily significant processes influencing plastome rearrangements, gene dosage effects, and substitution rates, the conserved IR architecture observed in *Polygala* provides further evidence for the structural stability of the plastid genome in this genus.

Compared with the ancestral organization of angiosperms, the *Polygalaceae* plastid genome has typical rearrangements (Ruhlman and Jansen 2014). The plastome of autotroph *Polygala* species is significantly larger than that of nonphotosynthetic species of *Epirixanthes* within the family (Petersen et al. 2019). *P. vulgaris*, an autotroph member of *Polygalaceae* included here, encoded a total of 174 genes, including 102 protein-coding genes (PCGs), 63 transfer RNA (tRNA) genes, and 13 ribosomal RNA (rRNA) genes (Table 1).

The guanine and cytosine (GC) contents are important for understanding genetic and species diversity (Liu et al. 2023; Singh et al. 2016; Smarda et al. 2014). The cpDNA GC content of *P. vulgaris* is highly similar to that of other studied *Polygala* taxa, in that it is approximately 36%; the GC content in the LSC, SSC, and IR regions varies between 29 and 39%. In addition, the GC content of IRs (between IRs 31 and 39%) is greater than that of LSCs (between 34 and 38%) and SSC regions (29–30%), which is in agreement with the results of previously reported studies of different genera (Hu et al. 2023; Ma et al. 2021; Yan et al. 2023). This phenomenon might be attributed to the rRNA sequences with high GC contents being located in the IR regions.

The Consortium for the Barcode of Life (CBOL) (2009) has recommended the use of *matK* + *rbcL* as the core barcode based on primer universality, PCR amplification

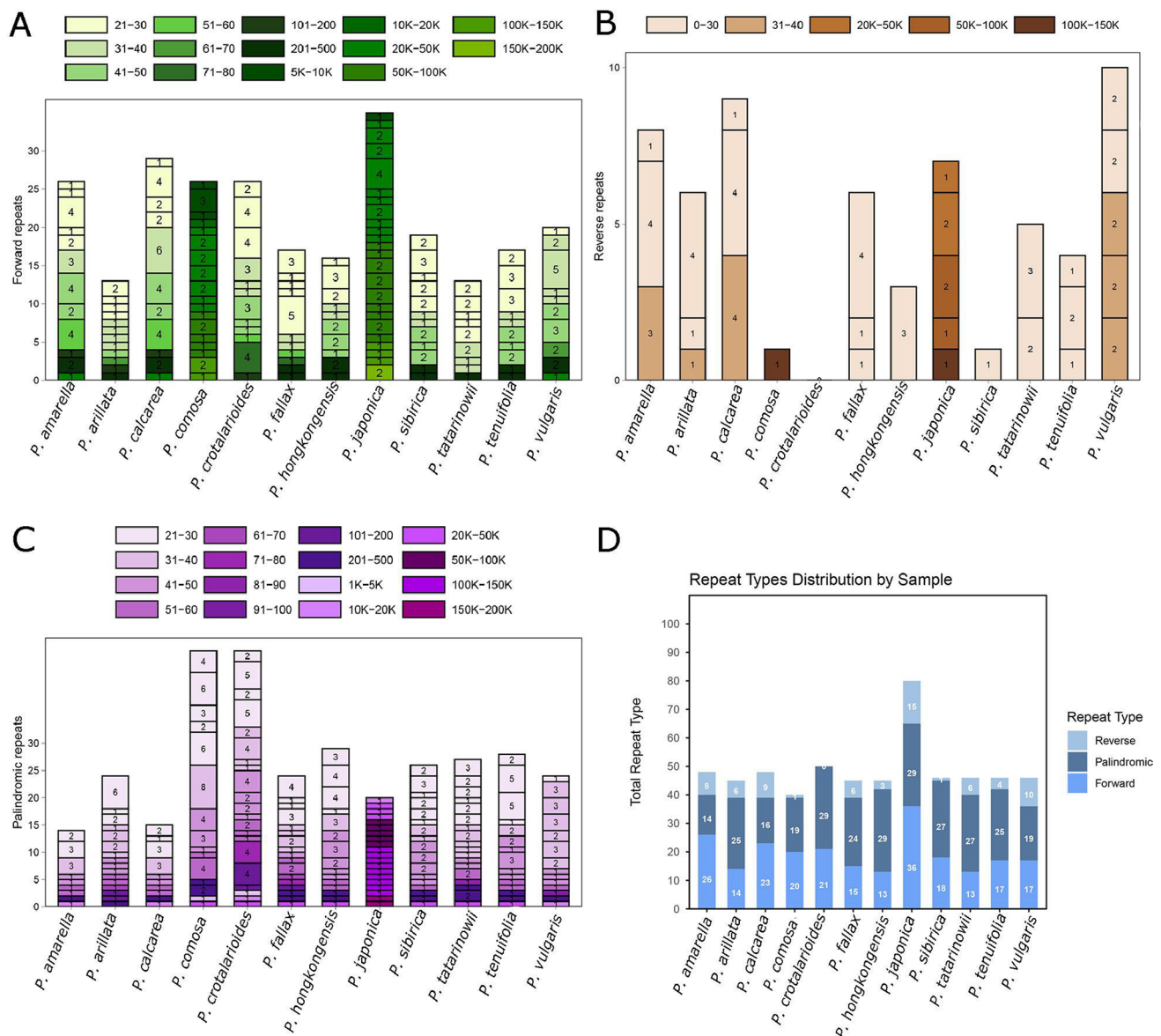


Fig. 7 Distribution of repeat sequences in the plastid genomes of *Polygala* species. **A** The number and size categories of forward repeats are shown for each species. **B** The number and size categories of reverse repeats detected in the plastomes. **C** The number and

size categories of palindromic repeats across species. **D** Comparative summary of the total number of repeat types (forward, reverse, palindromic) identified in each species

success and species discrimination power. Additionally, other barcode regions, such as *rpoC1*, *rpoB*, *psbA-trnH*, *psbK-psbI*, and *atpF-atpH*, *trnL-trnF*, and *ITS*, have been suggested to distinguish closely or distantly related taxonomic groups of land plants (Cowan et al. 2006; Hollingsworth et al. 2011; Kress et al. 2005; Uğurlu Aydın 2021). These hypervariable regions can serve as potential mini-barcode for biodiversity assessment and conservation especially in economical important or high-risk plant groups (Kress et al. 2015). Among chloroplast DNA barcoding markers, only the *matK-trnK* intergenic spacer region was

used to construct a phylogenetic tree (Pastore et al. 2019), and the remaining hotspot markers have not been used for efficient species identification in *Polygala*. Nucleotide variation analyses revealed that two potential barcode regions, *trnL-trnF* and *trnH-psbA*, can have enough discrimination power for DNA barcoding. Among them, only the *trnL-trnF* region was employed as a plant DNA barcode locus for *Polygala* taxa (Aygören Uluer 2021). Otherwise, *ycf1* was screened for species identification and suggested as one of the most promising DNA barcoding regions for land plants on the basis of its discrimination power (Dong et al.

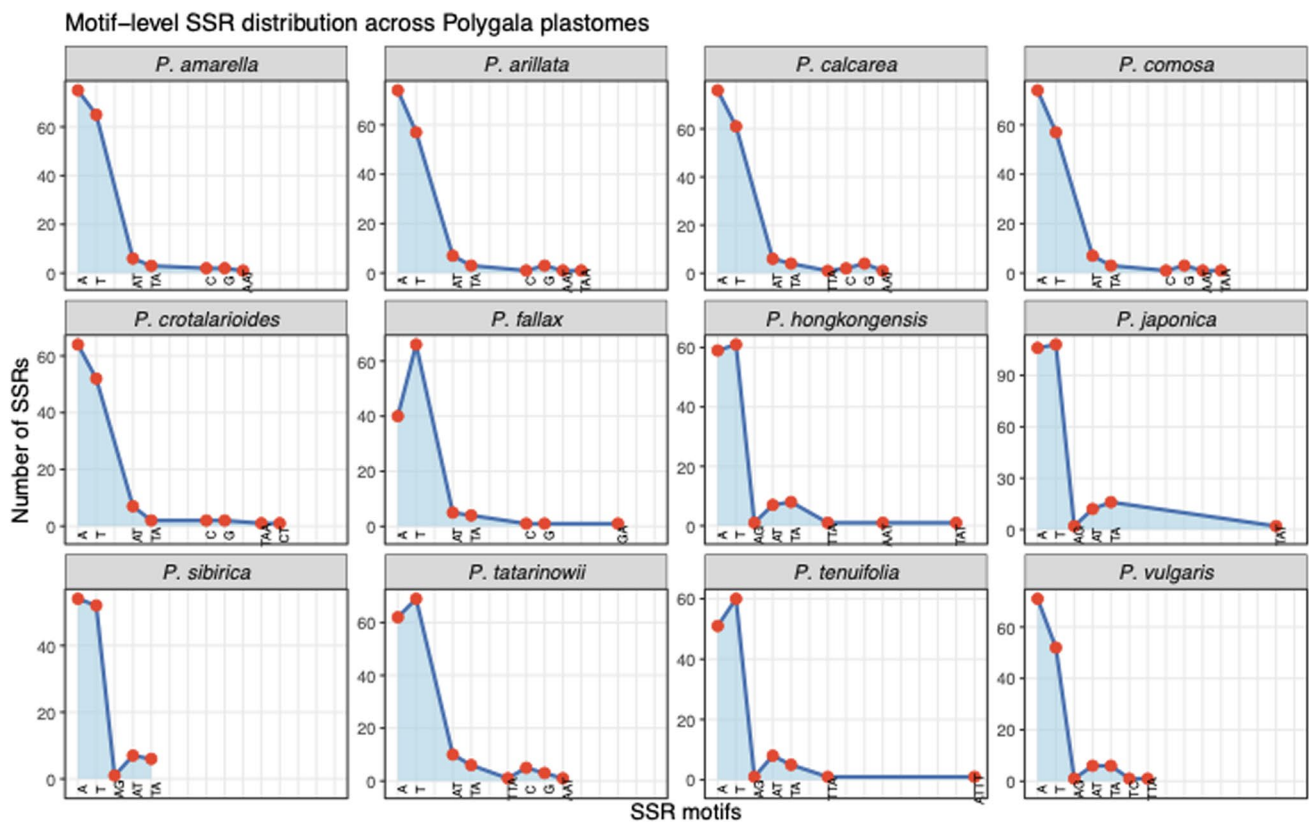


Fig. 8 Motif-level distribution of simple sequence repeats (SSRs) across *Polygala* plastomes.

2015), but the region had such low variation ($P_i > 0.2$) to be selected as a DNA barcoding marker for *Polygala*.

In this study, all hot spot regions were employed further supports the robustness of the phylogenetic relationships inferred from whole plastome data. The concatenated barcode-based phylogeny demonstrated that the selected hypervariable chloroplast regions provide sufficient phylogenetic signal to discriminate closely related *Polygala* species and to recover species-level relationships consistent with plastome-scale analyses. By concatenating nine chloroplast regions identified through nucleotide diversity analysis, we were able to reconstruct a species-level phylogeny that closely mirrored the topology obtained from complete chloroplast gene datasets. Our results indicate that these less commonly used regions may substantially contribute to phylogenetic resolution when combined in a multi-locus framework. Taken together, the integration of nucleotide diversity screening, targeted marker selection, and concatenated phylogenetic analysis offers a robust framework for DNA barcoding studies in *Polygala*. In this study, we tested their abilities of species identification but the other barcoding standards such as the universality of these markers, PCR success and sequence quality will be needed to test for rapid and accurate species identification in future.

In *P. vulgaris*, most of the nucleotide diversity values in the LSC, IR and SSC regions were greater than the mean values but IR regions had only two hot spot regions. Jeong et al. (2025) analyzed the nucleotide diversity in three *Polygala* chloroplast genomes and found similar results that the IR regions were more conserved than the LSC and SSC regions. Moreover, the P_i reported as 0.0096, is not in agreement with our finding. The overall P_i was found as 0.10970, and nine hypervariable gene regions were detected with a P_i value above 0.22 in this study.

The tetraploid species *P. vulgaris* is found in European countries, including Türkiye, and has important secondary metabolites for medicine (Çalış et al. 2025). Despite its pharmacological efficacy and cytological knowledge, the species has morphological variations, and it is necessary to clarify its phylogenetic position. The use of the cp genome has become increasingly common in the phylogeny of angiosperms (Hollingsworth et al. 2011; Xiong et al. 2023; Zhai et al. 2019) and offers advantages such as uniparental inheritance, small size, well-conserved structure, and moderate nucleotide variation (Mower and Vickrey 2018; Palmer 1985). Previous phylogenetic studies of *Polygala* have been based on several molecular markers sequenced via the Sanger method (Çeçen et al. 2023; Eriksen 1993; Lyskov et al. 2019; Martinez et al. 2022; Pastore et al. 2019; Persson

2001), whereas Lee et al. (2020), Ma et al. (2021), Wang (2023; 2025), Jeong et al. (2025), and our study produced relatively high phylogenetic resolution by utilizing plastome data. In this study, the phylogenomic tree yielded reliable results, confirming that *P. vulgaris* is a distinct species closely related to the European taxon *P. amerella*. Andrew J. Lack (1995) searched for putative hybrids between diploid *P. calcarea* and *P. vulgaris* and reported that both species hybridize with tetraploid *P. vulgaris* on the basis of isozyme enzymes. The phylogenetic tree inferred from the BI and ML analyses revealed that *P. vulgaris* separated from *P. amerella* and *P. calcarea*, a finding that corroborated the previous studies of Pastore et al. (2019). In the phylogenetic tree, *P. vulgaris* and *P. amerella* were clustered together with strong support values. However, Pastore et al. (2019) reported that *P. amerella* is more closely related to *P. calcarea* than to *P. vulgaris* on the basis of *trnL-F*, *rbcL*, and *trnK-matK* and the nuclear ribosomal ITS. Overall, the current taxon sampling methods are still limited, and the cp genome data of other European taxa should be evaluated by combining evidence from the chromosome evolution of *P. vulgaris*.

In addition, an important evolutionary aspect of the present study is the comparison between the tetraploid cytotype of *P. vulgaris* and the diploid cytotype of *P. tenuifolia* (Meng et al. 2023). Although polyploidization primarily affects the nuclear genome, cytonuclear interactions and genome dosage effects may indirectly influence plastid genome evolution (Soltis et al. 2015; Peer et al. 2017; Wicke et al. 2011). In this context, the largely conserved plastome structure observed between the tetraploid *P. vulgaris* and the diploid *P. tenuifolia* suggests a high level of chloroplast genome stability across different ploidy levels in *Polygala*. This conservation is reflected in multiple plastome features, including overall genome size and quadripartite organization, highly similar gene content and gene order, comparable distributions of repeat sequences and simple sequence repeats (SSRs), and the absence of extensive gene loss or pseudogenization. These patterns are consistent with previous plastome-based studies showing that chloroplast genomes generally remain structurally stable despite changes in nuclear ploidy, owing to their uniparental inheritance and independent evolutionary dynamics (Palmer 1985; Wicke et al. 2011; Soltis et al. 2015).

The analysis of relative synonymous codon usage (RSCU) in the chloroplast genome of *P. vulgaris* revealed a codon usage pattern largely consistent with that observed across other *Polygala* species. Like many angiosperm chloroplast genomes, codons ending with A or U are strongly favored, reflecting a general bias toward A/U at the third codon position (Dai et al. 2023; Wang et al. 2020; Zou et al. 2025). The universal initiation codon AUG (Met) likewise displayed high relative usage. Although the overall codon usage patterns were highly conserved among *Polygala* species, subtle

interspecific differences were detected, particularly within the sets of synonymous codons for leucine and serine. These minor deviations may reflect lineage-specific evolutionary pressures, such as mutation bias, translational selection, or genetic drift, that can shape codon preferences within chloroplast genomes (Chakraborty et al. 2020; Xiao et al. 2024; Yang et al. 2024). Importantly, the codon usage bias observed in *P. vulgaris* did not diverge markedly from that of its congeners, supporting the notion that codon usage in *Polygala* chloroplast genomes is relatively stable and conserved. Taken together, these results highlight that the chloroplast genomes of *Polygala* retain the characteristic A/U bias typical of higher plants while also exhibiting species-specific codon usage signatures that may provide insights into their evolution.

The distribution of repeat sequences showed considerable variation among the *Polygala* species analyzed. Forward repeats were generally the most abundant repeat type, followed by palindromic and reverse repeats, which is consistent with observations in other angiosperm plastomes (Guo et al. 2022; Tseng et al. 2025; Wu et al. 2021a). In *P. vulgaris*, the total number of repeats was moderate compared with that in the other species, with relatively balanced proportions of forward and palindromic repeats. The presence of a relatively high proportion of certain repeat classes may suggest potential hotspots for genome rearrangements or play a role in maintaining plastome stability and these studies have highlighted the role of repeats in plastome evolution, including structural variation and recombination events (Asaf et al. 2020; Wu et al. 2021a; Xu et al. 2023). Our findings for *P. vulgaris* provide further evidence that repeat content may contribute to lineage-specific patterns of plastid genome evolution within *Polygala*. The comparative SSR analysis revealed that mononucleotide repeats, particularly A/T motifs, dominated all *Polygala* plastomes in this study. This strong A/T bias is a common feature of angiosperm chloroplast genomes and has been widely attributed to the AT-rich nature of plastid DNA and replication mechanisms (Palmer 1985; Wicke et al. 2011; Daniell et al. 2016). Although the general SSR composition was highly conserved across species, moderate interspecific variation in total SSR number and repeat class distribution was detected. Such variation may reflect lineage-specific mutation dynamics and population-level divergence, and it provides a useful genomic resource for developing polymorphic chloroplast SSR markers. These markers have proven valuable tools for studies of species identification, phylogeography, and genetic diversity in closely related taxa.

Importantly, the limited presence of higher-order SSR motifs (tri- and tetranucleotide repeats) across *Polygala* plastomes suggests a relatively stable plastome architecture. The single tetranucleotide SSR detected in *P. tenuifolia* further supports the notion that long repeat motifs are rare but may

represent lineage-specific evolutionary signatures. Together, these findings highlight the potential of plastid SSRs as complementary molecular resources for future taxonomic, population genetic, and barcoding studies within *Polygalaceae*.

Conclusion

The comparative analysis of nucleotide diversity, repeat sequences, genome synteny, and phylogenetic relationships provides complementary insights into the evolutionary dynamics of *Polygala* chloroplast genomes. Our repeat analysis revealed that *P. vulgaris* has a distinctive distribution of forward, reverse, and palindromic repeats. Such repeats, particularly forward and palindromic repeats, are known to facilitate recombination and may underlie subtle structural polymorphisms, especially near IR boundaries and intergenic spacers. Despite this variability, the synteny comparison demonstrated that the overall gene order across *Polygala* species remains highly conserved, indicating a structurally stable plastome with no evidence of large-scale rearrangements. This structural conservation is consistent with the phylogenetic reconstruction, which strongly supported *P. vulgaris* as a sister to *P. amerella*. Nine hypervariable regions were detected and their discrimination ability were tested based on concatenated DNA dataset.

This research enhances our knowledge of chloroplast genome evolution in *Polygala*, ensuring that a plastome resource contributes to understanding evolution of organelle genomes, phylogenomics and DNA barcoding.

Acknowledgements We thank Xu Cai and Dr. Gökhan Karakülah for their valuable comments on the manuscript. This study is part of the MSc thesis of the first author titled “An Investigation of Ploidy Level in *Polygala vulgaris* L. (*Polygalaceae*) Using Next-Generation Sequencing.” The numerical calculations reported in this paper were fully/partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

Author contributions AAD and ZUA collected specimens. ZUA and İG conceived of and conducted the study. Laboratory work and bioinformatic analyses were performed by İG and ZUA. The first draft of the manuscript was written by ZUA and İG. AAD review and editing of the manuscript, obtained research funding. All authors read and approved the final manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). We are grateful to TÜBİTAK (Project no. 118 Z 708) for financial support.

Data availability The raw reads used in this study were deposited in the Sequence Read Archive (SRA) NCBI accession number PRJNA1139969.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval Research ethics committee’s approval was not required for this study, as the experimental organisms used were unregulated invertebrate species.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbot JR (2009) Revision of *Badiera* (*Polygalaceae*) and Phylogeny of the *Polygaleae* (Ph.D. dissertation) University of Florida, Gainesville
- Amiryousefi A, Hyvönen J, Poczar P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34(17):3030
- Asaf S, Khan AL, Lubna, et al (2020) Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of *Plantago ovata*. *Sci Rep* 10:3881
- Aygören Uluer D (2021) DNA barcoding of *Polygalaceae*. *Erzincan Univ J Sci Technol* 14(1):270–283
- Barrett CF, Bacon CD, Antonelli A, Cano Á, Hofmann T (2016) An introduction to plant phylogenomics with a focus on palms. *Bot J Linn Soc* 182:234–255
- Bi Y, Zhang M, Xue J, Dong R, Du Y, Zhang X (2018) Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. *Sci Rep* 8(1):1–12
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Borowiec ML (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660
- Çalış İ, Ünlü A, Aydın ZU, Dönmez AA, Yusufoglu HS, Jurt S, Linden A (2022) Xanthones and xanthone O-D-glucosides from the roots of *Polygala azizsancarii* Dönmez. *Chem Biodivers* 19(8):e202200499
- Çalış İ, Becer E, Ünlü A, Aydın ZU, Hanoğlu A, Vatansever HS, Dönmez AA (2023) Comparative phytochemical studies on the roots of *Polygala azizsancarii* and *P. peshmenii* and neuroprotective activities of the two xanthones. *Phytochemistry* 210:113650
- Çalış İ, Hanoğlu A, Uğurlu Aydın Z, Dönmez AA, Jurt S (2025) Phytochemical studies on the roots of *Polygala vulgaris*. *Nat Prod Res*, 1–10
- Cauz-Santos LA, da Costa ZP, Callot C, Cauet S, Zucchi MI, Bergès H, van den Berg C, Vieira MLC (2020) A repertoire of rearrangements and the loss of an inverted repeat region in *Passiflora* chloroplast genomes. *Genome Biol Evol* 12(10):1841–1857
- Çeçen Ö, Uğurlu Aydın Z, Dönmez AA (2023) Morphological and molecular data on a new species of *Polygala* (*Polygalaceae*) from Turkey. *Nord J Bot* 6:e03951
- Chakraborty S, Yengkhom S, Uddin A (2020) Analysis of codon usage bias of chloroplast genes in *Oryza* species. *Planta* 252:67
- Chen Q, Wu XB, Zhang DQ (2020) Comparison of the abilities of universal, super, and specific DNA barcodes to discriminate among

- the original species of *Fritillaria cirrhosae* bulbis and its adulterants. PLoS ONE 15:e0229181
- Chen Q, Hu H, Zhang D (2022) DNA barcoding and phylogenomic analysis of the genus *Fritillaria* in China based on complete chloroplast genomes. Front Plant Sci 13:764255
- Chen S, Yin X, Han J, Sun W, Yao H, Song J, Li X (2023) DNA barcoding in herbal medicine: retrospective and prospective. J Pharm Anal 13(5):431–441. <https://doi.org/10.1016/j.jpha.2023.03.008>
- Choi I-S, Jansen R, Ruhlman T (2019) Lost and found: return of the inverted repeat in the legume clade defined by its absence. Genome Biol Evol 11:1321–1333
- Chung HJ, Jung JD, Park HW, Kim JH, Cha HW, Min SR, Jeong WJ, Liu JR (2006) The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. Plant Cell Rep 2006(25):1369–1379
- Cowan RS, Case MW, Kress WJ, Savolainen V (2006) 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. Taxon 55:611–616
- Dai J, Liu Q, Xu X, Tan Z, Lin Y, Gao X, Zhu S (2023) Comparative and phylogenetic analysis of the complete chloroplast genomes of *Uncaria* (Rubiaceae) species. Front Plant Sci 14:1271689
- Dall'Acqua S, Innocenti G, Viola G, Piovan A, Caniato R, Cappelletti EM (2002) Cytotoxic compounds from *Polygala vulgaris*. Chem Pharm Bull 50(11):1499–1501
- Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17:1–29
- Dong W, Xu C, Li C et al (2015) *ycf1*, the most promising plastid DNA barcode of land plants. Sci Rep 5:8348
- Dönmez AA, Uğurlu Aydın Z, Yüzbaşıoğlu IS (2024) On the border of extinction and speciation, *Polygala seyfegoluensis* (Polygalaceae), a new species from marshy habitat, central türkiye: under the threat of climate change, water retreat, and subsequent degradation. Turk J Botany 48(1):44–56
- Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238
- Eriksen B (1993) Phylogeny of the Polygalaceae and its taxonomic implications. Plant Syst Evol 186(1):33–55
- Eriksen B, Persson C (2007) Polygalaceae. In: Kubitzki K (ed) The families and genera of vascular plants. Springer, Berlin
- Ghimiray D, Sharma B (2014) Comparative and bioinformatics analyses of the Solanaceae chloroplast genomes: plastome organization is more or less conserved at family level. J Appl Biol Biotechnol 2(03):021–026
- Glendinning DR (1960) Cytology of *Polygala*. Nature. <https://doi.org/10.1038/188604a0>
- Gornicki P, Zhu H, Wang J, Challa GS, Zhang Z, Gill BS, Li W (2014) The chloroplast view of the evolution of polyploid wheat. New Phytol 204(3):704–714
- Graham SW, Olmstead RG (2000) Utility of 17 chloroplast genes for inferring the phylogeny of basal angiosperms. Am J Bot 87(11):1712–1730
- Guo XX, Qu XJ, Zhang XJ, Fan SJ (2022) Comparative and phylogenetic analysis of complete plastomes among *Aristidoideae* species (Poaceae). Biology Basel 11:63
- Guzmán-Díaz S, Núñez FAA, Veltjen E, Asselman P, Larridon I, Samain MS (2022) Comparison of Magnoliaceae plastomes: adding Neotropical *Magnolia* to the discussion. Plants Basel 11:448
- Hebert PD, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270:313–321. <https://doi.org/10.1098/rspb.2002.2218>
- Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS ONE 6:e19254
- Horjales M (2003) Cantidades de DNA nuclear en árboles y arbustos. Nov. Acta Cient. Compostel., c. 13
- Hu Q, Qian R, Zhang Y et al (2023) Complete chloroplast genome molecular structure, comparative and phylogenetic analyses of *Sphaeropteris lepifera* (Cyatheaceae). Sci Rep 13:1356
- Jeong S, Han JW, Kim Y, Bak E, Ma KH, Lee JH et al (2025) Complete chloroplast genomes of three *Polygala* species and indel marker development for identification of authentic *Polygalae radix* (*Polygala tenuifolia*). Genes Genomics 47(1):99–112
- Jin JJ, Yu WB, Yang JB et al (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21:241
- Jing Y, Hu B, Ji H, Zhao F, Li B, Luo Y et al (2024) Unveiling the phytochemical profile and antioxidant activity of roots from six *Polygala* species. Arab J Chem 17(9):105915
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30(14):3059–3066
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649
- Klein Júnior LC, de Falmi Andra S, Filho VC (2012) A pharmacognostic approach to the *Polygala* genus: phytochemical and pharmacological aspects. Chem Biodivers 9:181–209
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102:8369–8374
- Kress WJ, García-Robledo C, Uriarte M, Erickson DL (2015) DNA barcodes for ecology, evolution, and conservation. Trends Ecol Evol 30(1):25–35
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucl Acids Res 29(22):4633–4642
- Lack AJ (1995) Relationships and hybridization between British species of *Polygala*—evidence from isozymes. New Phytol 130:217
- Lee DH, Cho WB, Park BJ, Kim JD, Byeon JG (2020) The complete chloroplast genome of *Polygala tenuifolia*, a critically endangered species in Korea. Mitochondrial DNA B 5:1919–1920
- Li QQ, Niu Y, Zhang ZP, Wen J, Liao CY (2025) Comparative chloroplast genome analyses of *Oxytropis* DC. species: new insights into genome evolution and phylogenomic implications. Front Plant Sci 16:1645582
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11):1451–1452
- Liu Y, Liang N, Xian Q et al (2023) GC heterogeneity reveals sequence-structures evolution of angiosperm ITS2. BMC Plant Biol 23:608
- Lyskov D, Pastore JF, Samigullin T (2019) One for all: molecular study of *Polygala major* complex (*Polygalaceae*) in Southwest Asia. Plant Syst Evol 305(10):975–984
- Ma J, Wang J, Li C, Liu J, Chen C, Hua Y (2021) The complete chloroplast genome characteristics of *Polygala crotalarioides* Buch.-Ham. ex DC. (*Polygalaceae*) from Yunnan, China. Mitochondrial DNA B 6(10):2838–2840
- Martínez A, Acosta JM, Ferrero MA, Pastore FB, Aagesen L (2022) Evolutionary patterns within the New World clade *Polygala* sections *Clinclinia* and *Monninopsis* (*Polygalaceae*). Perspect Plant Ecol Evol Syst 55:125673
- Meng F, Chu T, Feng P, Li N, Song C, Li C, Leng L, Song X, Chen W (2023) Genome assembly of *Polygala tenuifolia* provides insights into its karyotype evolution and triterpenoid saponin biosynthesis. Hortic Res 10(9):139
- Mower JP, Vickrey TL (2018) Structural diversity among plastid genomes of land plants. In: Chaw S-M, Jansen RK (eds) Advances

- in botanical research, vol 85. Academic Press, Cambridge, MA, USA, pp 263–292
- Neuhaus HE, Emes MJ (2020) Nonphotosynthetic metabolism in plastids. *Annu Rev Plant Physiol Plant Mol Biol* 51:111–140
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1):268–274
- Palmer JD (1985) Comparative organization of chloroplast genomes. *Annu Rev Genet* 19:325–354
- Pastore JFB, Abbott JR, Neubig KM, Van Den Berg C, Mota MCDA, Cabral A, Whitten WM (2019) Phylogeny and biogeography of *Polygala* (Polygalaceae). *Taxon* 68(4):673–691
- Perry AS, Wolfe KH (2002) Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J Mol Evol* 55(5):501–508
- Persson C (2001) Phylogenetic relationships in *Polygalaceae* based on plastid DNA sequences from the trnL-F region. *Taxon* 50(3):763–779
- Petersen G, Darby H, Lam VKY, Pedersen HÆ, Merckx VSFT, Zervas A et al (2019) Mycoheterotrophic *Epirixanthes* (Polygalaceae) has a typical angiosperm mitogenome but unorthodox plastid genomes. *Ann Bot* 124(5):791–807
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539–542
- Ruhlman TA, Jansen RK (2014) The plastid genomes of flowering plants. *Methods Mol Biol* 1132:3–38
- Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33:W686–W689
- Şenova MK (2021) Cytotaxonomy of Some Turkish *Polygala* L. (*Polygalaceae*) taxa with chromosome counting and flow cytometry (FCM) techniques. MSc Thesis, Hacettepe University, Ankara, Türkiye. <https://tez.yok.gov.tr/>
- Serrano I, Audran C, Rivas S (2016) Chloroplasts at work during plant innate immunity. *J Exp Bot* 67(13):3845–3854
- Singh R, Ming R, Yu Q (2016) Comparative analysis of GC content variations in plant genomes. *Trop Plant Biol* 9:136–149
- Smarda P, Bures P, Horova L, Leitch IJ, Mucina L, Pacini E et al (2014) Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc Natl Acad Sci U S A* 111:E4096–E4102
- Soltis PS, Marchant DB, Van de Peer Y, Soltis DE (2015) Polyploidy and genome evolution in plants. *Curr Opin Genet Dev* 35:119–125
- Song JH, Yang S, Choi G, Park I (2023a) Characterization of the complete chloroplast genome of the important herbal medicinal plant *Thymus quinquecostatus* (Lamiaceae). *J Asia Pac Biodivers* 16(1):140–143
- Song Y, Li C, Liu L, Hu P, Li G, Zhao X, Zhou H (2023b) The population genomic analyses of chloroplast genomes shed new insights on the complicated ploidy and evolutionary history in *Fragaria*. *Front Plant Sci* 13:1065218
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucl Acids Res* 34:435–439
- Thiel T et al (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *TAG Theor Appl Genet* 106:411–422
- Tseng YH, Chien HC, Zhu GX (2025) Comparative plastome analyses and phylogenetic insights of *Elatostema*. *BMC Plant Biol* 25:537
- Uğurlu Aydın Z, Kaya Y, Dönmez AA (2021) DNA barcoding evaluation of geophytes: comparative efficiency of three barcode loci for *Anemone* (*Ranunculaceae*) and *Gladiolus* (*Iridaceae*). *Plant Biosyst* 156(4):926–937
- Van de Peer Y, Mizrahi E, Marchal K (2017) The evolutionary significance of polyploidy. *Nat Rev Genet* 18:411–424
- Wang YH (2023) The complete chloroplast genome of *Polygala subopposita*, an endemic milkwort in China. *Mitochondrial DNA B* 8:1290–1293
- Wang YH (2025) The complete chloroplast genome of *Polygala tatarinowii* (*Polygalaceae*). *Mitochondrial DNA B* 10(5):337–341
- Wang Z, Xu B, Li B, Zhou Q, Wang G, Jiang X et al (2020) Comparative analysis of codon usage patterns in chloroplast genomes of six *Euphorbiaceae* species. *PeerJ* 8:e8251
- Wicke S, Schneeweiss GM, de Pamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 76:273–297
- Wu S, Chen J, Li Y, Liu A, Li A, Yin M et al (2021a) Extensive genomic rearrangements mediated by repetitive sequences in plastomes of *Medicago* and its relatives. *BMC Plant Biol* 21(1):421
- Wu L, Wu ML, Cui N, Xiang L, Li Y, Li XW et al (2021b) Plant superbarcode: a case study on genome-based identification for closely related species of *Fritillaria*. *Chin Med* 16:52
- Xiao M, Hu X, Li Y et al (2024) Comparative analysis of codon usage patterns in the chloroplast genomes of nine forage legumes. *Physiol Mol Biol Plants* 30:153–166
- Xiong C, Huang Y, Li Z et al (2023) Comparative chloroplast genomics reveals the phylogeny and adaptive evolution of *Begonia* in China. *BMC Genomics* 24:648
- Xu S, Teng K, Zhang H, Gao K, Wu J, Duan L et al (2023) Chloroplast genomes of four *Carex* species: long repetitive sequences trigger dramatic changes in chloroplast genome structure. *Front Plant Sci* 14:1100876
- Xue, Liu Er-Mei, Chang Jian-Feng, Liu Yue-Ning, Huang Ya, Wang Ning, Yao Ze-Ping, Jiang (2019) Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Quercus bawanglingensis* Huang Li et Xing a Vulnerable Oak Tree in China Forests 10(7) 587-10.3390/f10070587
- Yan M, Dong S, Gong Q et al (2023) Comparative chloroplast genome analysis of four *Polygonatum* species provides insights into DNA barcoding, evolution, and phylogeny. *Sci Rep* 13:16495
- Yang Y, Wang X, Shi Z (2024) Comparative study on codon usage patterns across chloroplast genomes of eighteen *Taraxacum* species. *Horticulturae* 10(5):492
- Yoshikawa M, Murakami T, Ueno T, Kadoya M, Matsuda H, Yamahara J, Murakami N (1995) E-senegasaponins A and B, Z-senegasaponins A and B, Z-seneginins II and III, new inhibitors of ethanol absorption from *Senegae Radix* (*Polygala senega*). *Chem Pharm Bull* 43(2):350–352
- Yoshikawa M, Murakami T, Matsuda H et al (1996) Bioactive saponins and glycosides from *Senegae Radix* (roots of *Polygala senega*). *Chem Pharm Bull* 44(10):1915–1922
- Zhai W, Duan X, Zhang R, Guo C, Li L, Xu G et al (2019) Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Mol Phylogenet Evol* 135:12–21
- Zhang N, Huang K, Xie P et al (2024) Chloroplast genome analysis and evolutionary insights in the versatile medicinal plant *Calendula officinalis* L. *Sci Rep* 14:9662
- Zou T, Li D, Zhao CY et al (2025) Chloroplast whole genome assembly and phylogenetic analysis of *Persicaria criopolitana* reveals its new taxonomic status. *Sci Rep* 15:19890

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