

The complete chloroplast genome of *Polygala tatarinowii* (Polygalaceae)

Yin-Huan Wang

School of Primary Education, Chongqing Normal University, Chongqing, China

ABSTRACT

Polygala tatarinowii (Polygalaceae) is an annual herb with ecological, medicinal, and agricultural potential. Here, we present the inaugural assembly of its chloroplast genome (plastome). The total plastome spans 168,779 bp in length, featuring two expanded inverted repeat (IR) regions measuring 43,427 bp each, flanking a large single-copy (LSC) region of 73,712 bp and a small single-copy (SSC) region of 8213 bp. The plastome exhibits a GC content of 36.8%. Annotation of the plastome revealed a total of 112 unique genes, comprising 78 protein-coding genes, 30 tRNAs and 4 rRNAs. Uncommonly, two copies of the *trnQ*^{UUG} gene were identified within the IRs, resulting in the presence of three *trnQ*^{UUG} copies in this plastome. Additionally, reductions were observed in the *ycf1* and *ycf2* genes. Phylogenetic analysis effectively resolves the relationships among sampled Polygalaceae species. Our findings underscore the instrumental role of plastomes in advancing research on the systematics, phylogenetics, and genetic diversity of Polygalaceae.

ARTICLE HISTORY

Received 30 May 2024
Accepted 14 March 2025

KEYWORDS

Comparative analysis;
phylogenetic analysis;
plastome; polygalaceae

Introduction

Polygala tatarinowii Regel 1862, a member of the family Polygalaceae, is an annual herb primarily found in temperate biome. Its native range spans NE. Pakistan to S. Russian Far East and Philippines, Papua New Guinea. It thrives in grasslands on slopes, thickets, roadsides, exhibiting a broad distribution across China (Chen et al. 2008). *Polygala tatarinowii* was considered the most important floristic regional indicator plant in the wall of Namhan Mountain Fortress (Cha et al. 2021). Its whole plant can be used medicinally to treat malaria and infirmity (Editorial Committee of Flora of China, Chinese Academy of Sciences 1997). Records dating back to the Northern Wei period in Chinese history indicate its cultivation for use as valuable green manure (Cui et al. 2014). Despite its ecological, medicinal, and agricultural potential, research on this species remains scarce.



The Polygalaceae family comprises approximately 29 genera and over 1200 species, distributed worldwide (Pastore et al. 2019). The chloroplast genomes (plastomes) have been proven invaluable in phylogenetic and evolutionary studies. However, research on the chloroplast genomes of Polygalaceae plants is still limited. This article presents, for the first time, the plastome of the *P. tatarinowii*, highlighting its structural features and phylogenetic analysis. Through the construction of a phylogenetic tree alongside published plastomes from the Polygalaceae family, it aims to assess their


evolutionary relationships, laying the groundwork for future research on utilization, systematics, and evolution within the family.

Materials and methods

Plant material was collected from wild grass ground in Xinjie Town of Wenshan Prefecture, Yunnan Province (23°11'15.49"N, 104°0'52.03"E). The identifier was Chunyan Han from Kunming Caizhi Biotechnology Co., Ltd. The voucher specimen (no. WYH2020007) was deposited in the Chinese National Herbarium with barcode number of 2329598 (PE; <http://pe.ibcas.ac.cn/>; Curator: Qiang Wang, wangqiang@ibcas.ac.cn) (Figure 1).

DNA extraction was performed on silica gel-dried leaves to obtain total genomic DNA. Subsequently, a 150-bp paired-end library was constructed and subjected to sequencing on the Illumina HiSeq 4000 platform at Beijing Novogene Bioinformatics Technology Co., Ltd. (Nanjing, China). The raw sequencing data were then assembled to the complete plastome using the GetOrganelle (Jin et al. 2020). To affirm the assembly, paired-end reads were aligned to the final plastome using Bowtie2 v.2.4.5 (Langmead and Salzberg 2012) within the Geneious Prime 2022.2.2 software platform (Biomatters Ltd.). Annotation was conducted using the PGA software (Qu et al. 2019), with manual adjustments performed using Geneious Prime 2022.2.2. Both the raw

CONTACT Yin-Huan Wang  wangyinahuan@outlook.com  School of Primary Education, Chongqing Normal University, No. 1 Tuanshan Town, Beibei District, Chongqing 400700, China

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23802359.2025.2482279>.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

sequencing reads and the annotated plastome sequence were deposited in the Sequence Read Archive (SRA) and GenBank, respectively. Finally, the physical genome map was generated using the CPGView tool (<http://www.1kmppg.cn/cpgview/>) (Liu et al. 2023).

We conducted phylogenetic analysis to examine the relationships among *P. tatarinowii* and 25 other species of Polygalaceae with available sequence data. Four species representing the other three families of Fabales were utilized as outgroups for comparison. Sequences, including both genes and noncoding regions (including introns and intergenic sequences), were extracted and then aligned using the MAFFT Multiple Alignment v.1.5.0 plugin within Geneious Prime 2022.2.2. Matrices containing fewer than fifteen species or sequences shorter than 22 base pairs (bp) were excluded from subsequent analyses. Ultimately, we obtained 236 matrices, comprising 110 genes and 126 noncoding regions. Details regarding the genes, noncoding regions, and their respective lengths for each species are provided in [Supplementary Table 1](#). To improve alignment accuracy, ambiguously aligned sites were removed from all alignments using GBLOCKS v.0.91b (Castresana 2000; Talavera and Castresana 2007), with default parameters except for allowing all gap positions. Subsequently, the 236 GBLOCKS-edited matrices were concatenated to construct a supermatrix. In cases where a species lacked a specific gene or noncoding region, it was treated as missing data in the supermatrix. The constructed supermatrix was subjected to maximum likelihood (ML) phylogenetic reconstruction using the RAxML V.4.0 plugin within Geneious Prime 2022.2.2, employing the GTRGAMMA nucleotide substitution model and conducting 1,000 rapid bootstrap replicates for robustness assessment.

Results and discussion

The plastome of *Polygala tatarinowii* spans 168,779 bp in length, with a GC content of 36.8%. The read alignment to the plastome showed coverage ranging from $7\times$ to $3590\times$, with an average of $2481.7\times$ ([Supplementary Figure 1](#)). It exhibits the characteristic quadripartite structure, comprising two inverted repeat (IR) regions flanking a large single-copy (LSC) region and a small single-copy (SSC) region ([Figure 2](#)). Comparative analysis with the plastome of the outgroup *Adenolobus garipensis* (E. Mey.) Torre & Hillc., representing a typical angiosperm plastome, reveals an expansion of the IR region by over 8500 bp into the LSC region and approximately 8000 bp into the SSC region. Consequently, the longer IR region measures 43,427 bp, while the LSC and SSC regions span 73,712 bp and 8213 bp, respectively. Notably, one inverted and translocated region, along with two inversions, are identified within this plastome.

Polygala tatarinowii's plastome harbors 112 unique genes, encompassing 78 protein-coding genes, 30 tRNAs and 4 rRNAs. Among these, 33 genes are duplicated within the IRs, including an additional *trnQ*^{UUG} copy, resulting in the presence of three *trnQ*^{UUG} copies in this plastome. A similar duplication of this gene has been observed in another Polygalaceae species *Polygala subopposita* S. K. Chen (Wang 2023). Furthermore, the *ycf1* and *ycf2* genes exhibit reductions to 1,452 bp and 2,580 bp, respectively, in this plastome. These two genes are found to be functional and essential for cell survival in dicot plastomes (Drescher et al. 2000). While the reduction or pseudogenization have been documented in various lineages, such as *Cynodon dactylon* (L.) Persoon (Huang et al. 2017), legumes (Oyebanji et al. 2020) and *Pelargonium* (Weng et al. 2019), etc, potentially indicating



Figure 1. *Polygala tatarinowii*. (A–C) Images of the plant, flowers, fruits and leaves (photo by the Caizhi Biotechnology Co., LTD); (D) Specimen (made by the Caizhi Biotechnology Co., LTD). Herbs annual, erect, 5–15 cm tall. Stems branched, longitudinally angular, glabrous. Petals 3, connate in lower 2/3, red to purple-red; keel slightly shorter than lateral petals, without appendage.

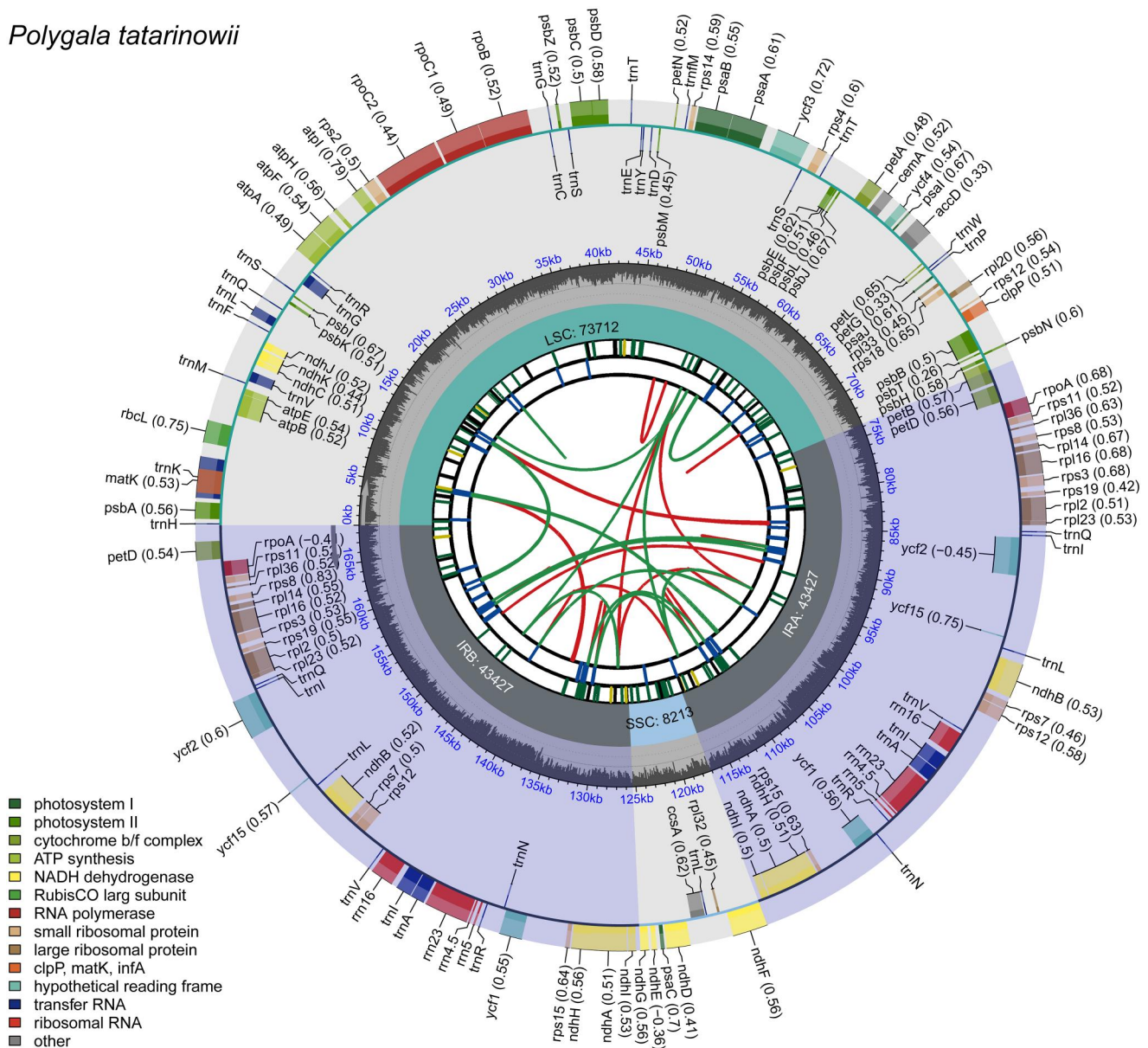
Polygala tatarinowii

Figure 2. Schematic map of overall features of the chloroplast genome of *Polygala tatarinowii*. The map contains six tracks in default. From the Centre outward, the first track shows the dispersed repeats. The dispersed repeats consist of direct (D) and palindromic (P) repeats, connected with red and green arcs. The second track shows the long tandem repeats as short blue bars. The third track shows the short tandem repeats or microsatellite sequences as short bars with different colors. The small single-copy (SSC), inverted repeat (IRa and IRb), and large single-copy (LSC) regions are shown on the fourth track. The GC content along the genome is plotted on the fifth track. The genes are shown on the sixth track. The optional codon usage bias is displayed in the parenthesis after the gene name. Genes are color-coded by their functional classification which is shown in the bottom left corner. The transcription directions for the inner and outer genes are clockwise and anticlockwise, respectively.

functional transfer to the nucleus. Moreover, the *clpP* gene in this plastome lacks a portion of intron 1 and the entire intron 2, a common feature shared with the majority of Polygalaceae plastomes and certain other lineages, including select legumes (Wang, Qu, et al. 2017). Consequently, *atpF*, *clpP*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl16*, *rpl2*, and *rpoC1* contain one intron, while *rps12* and *ycf3* contain two introns (Supplementary Figure 2).

The ML phylogenetic tree effectively resolve the relationships among sampled Polygalaceae species, largely aligning with previous findings (Pastore et al. 2019; Wang 2023) (Figure 3). *Polygala tatarinowii* emerges as sister to the clade comprising *P. karensium*, *P. caudata*, and *P. wattersii*, with robust bootstrap support (value 100%). *Xanthophyllum* is

identified as the earliest diverged genus within Polygalaceae. Additionally, the grouping of *Salomonina* and *Epirixanthes* within *Polygala* suggests the polyphyletic nature of the latter genus.

Conclusion

In conclusion, our study elucidates the plastome of *Polygala tatarinowii*, unveiling several novel structural variations. Furthermore, our construction of a robust phylogenetic tree based on plastome data sheds light on the evolutionary relationships within the family Polygalaceae. These findings underscore the utility of plastomes in advancing research on the systematics, phylogenetics, and genetic diversity of

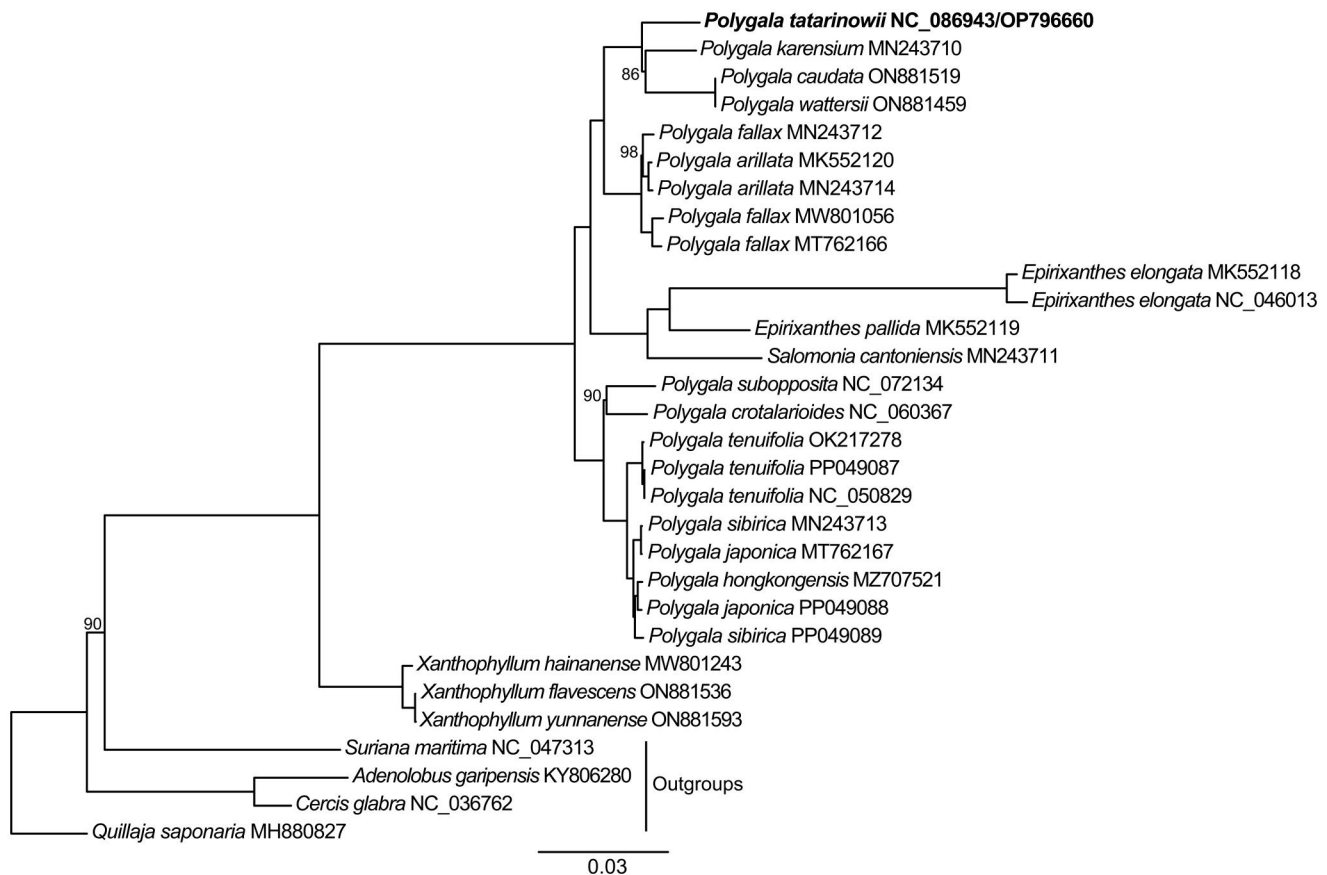


Figure 3. The maximum likelihood phylogenetic tree of sampled species of polygalaceae and outgroups based on genes and noncoding regions of plastomes. Maximum likelihood bootstrap percentages of less than 100% are given at nodes. Bold type marks species sequenced in this study. Numbers following the species names represent GenBank accession numbers and the corresponding publication are as follows: NC_036762 and KY806280 (Wang, Wang, et al. 2017); MK552118, MK552119, MK552120, NC_046013 (Petersen et al. 2019), MN243710, MN243711, MN243712, MN243713, and MN243714, NC_047313 (Zhang et al. 2020); MW801056, and MW801243 (Jin et al. 2021); NC_050829 (Lee et al. 2020); NC_060367 (Ma et al. 2021); NC072134 (Wang 2023); ON881459, ON881519, ON881536, and ON881593 (Jin et al. 2024).

Polygalaceae. Moving forward, plastomic analyses hold significant promise for furthering our understanding of the intricate evolutionary dynamics within this plant family.

Acknowledgment

I would like to express my sincere gratitude to ChatGPT (GPT-4 v2) for providing invaluable assistance in manuscript language editing.

Author contributions

Yin-Huan Wang designed this study, performed data analysis, wrote and revised the manuscript.

Ethical approval

No specific permission is required to collect the samples described in this study as the planting area of the sample is not in the natural reserve or any private domain. The field collection does not involve endangered or protected species. The use of the plant materials does not pose any risk to other species in nature.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by the Natural Science Foundation of Chongqing, China under Grant [number CSTB2022NSCQ-MSX0462]; the Science and Technology Research Programs of Chongqing Municipal Education Commission under Grant [number KJQN202200552 and KJQN202300531]; and the University-level fund project of Chongqing Normal University under Grant [number 18XLB009].

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at [<https://www.ncbi.nlm.nih.gov>] under the accession no. OP796660 and NC_086943 (<https://www.ncbi.nlm.nih.gov/nuccore/OP796660/>; https://www.ncbi.nlm.nih.gov/nuccore/NC_086943). The associated BioProject, SRA, and Bio-Sample numbers are PRJNA694335, SRR28825985, and SAMN41108749, respectively.

References

- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17(4):540–552. doi:10.1093/oxfordjournals.molbev.a026334.
- Cha D-W, Choi D-S, Kim J-S. 2021. A study on the characteristics of vascular plants in the wall of Namhan mountain fortress. *J Korean Inst Trad Landscape Arch.* 39:20–30.
- Chen S-K, Ma H-Y, Parnell JAN. 2008. Polygalaceae. In *Flora of China*, Z.-Y. Wu, P.H. Raven, and D.-Y. Hong, editors. Beijing & St. Louis: Science Press & Missouri Botanical Garden Press, p. 139–159.

- Cui Z-L, Dou Z-X, Chen X-P, Ju X-T, Zhang F-S. 2014. Managing agricultural nutrients for food security in China: past, present, and future. *Agron J.* 106(1):191–198. doi:10.2134/agronj2013.0381.
- Drescher A, Ruf S, Calsa T, Jr, Carrer H, Bock R. 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* 22(2):97–104. doi:10.1046/j.1365-3113x.2000.00722.x.
- Editorial Committee of Flora of China, Chinese Academy of Sciences. 1997. *Flora of China*. Beijing: Science Press, vol. 43, p. 170–171.
- Huang Y-Y, Cho S-T, Haryono M, Kuo C-H. 2017. Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. *PLoS One.* 12(6):e0179055. doi:10.1371/journal.pone.0179055.
- Jin L, Liu J-J, Li Q-M, Lin L-X, Shao X-N, Xiao T-W, Li B-H, Mi X-C, Ren H-B, Zhu Y, et al. 2024. Stronger latitudinal phylogenetic patterns in woody angiosperm assemblages with higher dispersal abilities in China. *J Biogeogr.* 51(2):269–279. doi:10.1111/jbi.14746.
- Jin L, Liu J-J, Xiao T-W, Li Q-M, Lin L-X, Shao X-N, Ma C-X, Li B-H, Mi X-C, Qiao X-J, et al. 2021. Community phylogenetics require phylogenies reconstructed from plastid genomes. *Authorea.* 2021:13.
- Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21(1):241. doi:10.1186/s13059-020-02154-5.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359. doi:10.1038/nmeth.1923.
- Lee D-H, Cho W-B, Park BJ, Kim JD, Byeon JG. 2020. The complete chloroplast genome of *Polygala tenuifolia*, a critically endangered species in Korea. *Mitochondrial DNA Part B.* 5(2):1919–1920. doi:10.1080/23802359.2020.1754947.
- Liu S-Y, Ni Y, Li J-L, Zhang X-Y, Yang H-Y, Chen H-M, Liu C. 2023. CPGView: a package for visualizing detailed chloroplast genome structures. *Mol Ecol Resour.* 23(3):694–704. doi:10.1111/1755-0998.13729.
- Ma J-Y, Wang J-M, Li C-Y, 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 Resour.* 6(10):2838–2840. doi:10.1080/23802359.2021.1964396.
- Oyebanji O, Zhang R, Chen S-Y, Yi T-S. 2020. New insights into the plastome evolution of the millettoid/phaseoloid clade (Papilionoideae, Leguminosae). *Front Plant Sci.* 11:151. doi:10.3389/fpls.2020.00151.
- Pastore JF, Abbott J, Neubig K, van den Berg C, Mota M, Cabral A, Whitten W. 2019. Phylogeny and biogeography of *Polygala* (Polygalaceae). *Taxon.* 68(4):673–691. doi:10.1002/tax.12119.
- Petersen G, Darby H, Lam VKY, Pedersen HAE, Merckx V, Zervas A, Seberg O, Graham SW. 2019. Mycoheterotrophic *Epirixanthes* (Polygalaceae) has a typical angiosperm mitogenome but unorthodox plastid genomes. *Ann Bot.* 124(5):791–807. doi:10.1093/aob/mcz114.
- Qu X-J, Moore MJ, Li D-Z, Yi T-S. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods.* 15(1):50. doi:10.1186/s13007-019-0435-7.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56(4):564–577. doi:10.1080/10635150701472164.
- Wang Y-H. 2023. The complete chloroplast genome of *Polygala subopposita*, an endemic milkwort in China. *Mitochondrial DNA B Resour.* 8(11):1290–1293. doi:10.1080/23802359.2023.2284410.
- Wang Y-H, Qu X-J, Chen S-Y, Li D-Z, Yi T-S. 2017. Plastomes of mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. *Tree Genetics & Genomes.* 13(2):41. doi:10.1007/s11295-017-1124-1.
- Wang Y-H, Wang H, Yi T-S, Wang Y-H. 2017. The complete chloroplast genomes of *Adenolobus garipensis* and *Cercis glabra* (Cercidoideae, Fabaceae). *Conservation Genet Resour.* 9(4):635–638. doi:10.1007/s12686-017-0744-y.
- Weng M-L, Ruhlman TA, Jansen RK. 2019. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytol.* 214(2):842–851. doi:10.1111/nph.14375.
- Zhang R, Wang Y-H, Jin J-J, Stull GW, Bruneau A, Cardoso D, de Queiroz LP, Moore MJ, Zhang S-D, Chen S-Y, et al. 2020. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Syst Biol.* 69(4):613–622. doi:10.1093/sysbio/syaa013.