CompCytogen 20: 127–139 (2026)
doi: 10.3897/compcytogen.20.184176
https://compcytogen.pensoft.net

RESEARCH ARTICLE

COMPARATIVE Cytogenetics
A peer-reviewed open-access journal
International Journal of Plant & Animal Cytogenetics, Karyosystematics, and Molecular Systematics

Genome survey and chromosome number determination of *Polygala fallax* (Polygalaceae), an endemic medicinal plant from southern China

Lin Fu[1,2*], Ziyue Liu[1,2*], Lina Han[3*], Ting Zhang[1,2,4], Yan Jiang[1,2], Denis A. Krivenko[5], Jin Hu[6], Hongfeng Chen[1,2], Lei Duan[1,2]

1 *State Key Laboratory of Plant Diversity and Specialty Crops/Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China* **2** *South China National Botanical Garden, Guangzhou, 510650, China* **3** *Guangdong Liangtian Agriculture & Forestry Technology Co., Ltd., Qingyuan, 511899, China* **4** *University of Chinese Academy of Sciences, Beijing, 100049, China* **5** *Siberian Institute of Plant Physiology and Biochemistry, Siberian Branch, Russian Academy of Sciences, Irkutsk 664033, Russia* **6** *Guangdong Eco-Engineering Polytechnic, Guangzhou, 510520, China*

Corresponding authors: Jin Hu (hujin0204@163.com); Hongfeng Chen (h.f.chen@scbg.ac.cn); Lei Duan (duanlei@scbg.ac.cn)

Academic editor: Desislava Stoianova | Received 4 January 2026 | Accepted 6 March 2026 | Published 15 April 2026

https://zoobank.org/D2F55FBC-F361-4308-BAAA-D1EE3FCCC7E9

Citation: Fu L, Liu Z, Han L, Zhang T, Jiang Y, Krivenko DA, Hu J, Chen H, Duan L (2026) Genome survey and chromosome number determination of *Polygala fallax* (Polygalaceae), an endemic medicinal plant from southern China. Comparative Cytogenetics 20: 127–139. https://doi.org/10.3897/compcytogen.20.184176

Abstract

Polygala fallax Hemsley, 1886 is a perennial herb of the family Polygalaceae. Because of its antioxidant, anti-aging, anti-inflammatory, and antibacterial properties, it is regarded as a traditional medicinal plant in China with high economic value. Although this species has significant medicinal and health benefits, cytological research on it remains very limited. In the present study, the genome size and chromosome number of *P. fallax* were determined through k-mer analysis, flow cytometry, and cytogenetics. Firstly, a genome survey was conducted using Illumina NovaSeq 6000 DNA sequencing. K-mer analysis revealed that the genome size of *P. fallax* was approximately 2199.64 Mb (ca. 2.20 Gb), the heterozygosity rate was about 1.11%, and the proportion of repeated sequences was about 73.61%. As for flow cytometry analysis, *Solanum lycopersicum* Linnaeus, 1753 was used as internal standard species, and the 2C-DNA value of *P. fallax* is 2.49 Gb. We counted the chromosome number of *P. fallax* using a common cytogenetic method, and the result showed a chromosome

* These authors contributed equally to this work.

Copyright *Lin Fu et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

number of 2n = 40. This comprehensive study offers a foundation for subsequent whole-genome sequencing of *P. fallax*, contributing valuable insights into its genetic characteristics and providing genomic information for further research on the synthesis mechanisms of its medicinal secondary metabolites, species conservation, and sustainable utilization.

Keywords
Chromosome number, genome size, genome survey, *Polygala fallax*

Introduction

Polygala fallax Hemsley, 1886, a member of the family Polygalaceae, is an endemic species in southern China, mainly distributed in Fujian, Guangdong, Guangxi, Guizhou, Hunan, Jiangxi and Yunnan Provinces. It is a shrub or small tree with stout roots, gray-green branchlets, lanceolate to elliptic-lanceolate leaves and yellow flowers in racemes (Fig. 1), and growing in forests, streamsides and moist habitats in valleys at altitudes of 400–1700 m (Chen et al. 2008). It is a perennial plant that has been widely used in traditional Chinese medicine for treating lumbar and leg pain resulted from strain, as well as acute and chronic hepatitis. Its main active components, including flavonoids, polysaccharides, volatile oils, and saponins, exhibit pharmacological activities such as antioxidant, anti-aging, anti-inflammatory, and antibacterial effects (Xu et al. 2006; Yao et al. 2020). Recent research found it has significant pharmacological effects in enhancing stress resistance, lowering lipid level, exhibiting anti-hepatocellular carcinoma activity, suppressing inflammation, improving myocardial ischemia, and facilitating recovery from the liver tissue (Klein et al. 2012; Wang et al. 2022). Ning and Dela Cruz (2023) reported that *P. fallax* might delay aging in rat through down-regulating the phosphorylation of *PI3K*, *AKT*, *mTOR* and *S6K1*, key proteins involved in PI3K/AKT/mTOR signaling pathway and has a promising application in anti-aging effect.

Due to the wide utilization as traditional medicine in southern China, natural resources of *P. fallax* have been extensively exploited, making its wild individuals over-collected. Genomic study is an important strategy to provide comprehensive data on individuals' genetic variation, offering the most complete information for researches on plant growth, development, evolution, and adaptation. However, most of current studies on *P. fallax* focus on its pharmacological components and activities, genomic study is still limited until now.

Before conducting whole-genome sequencing, it is necessary to understand the characteristics of the genome, such as heterozygosity, repetitive sequences, and GC content, in order to determine an efficient and accurate assembly strategy for subsequent research. Genome survey can provide preliminary information on genome structure, such as genome size, heterozygosity, and repeat contents. Genome size estimated by flow cytometry and chromosome number counting are also necessary analyses for studying genomic characterization. These methods can evaluate genome size from

Figure 1. The plant and flowers of *Polygala fallax*. **A** plant in the primary habitat **B** raceme with yellow flowers.

different perspectives, ensuring the sequencing progress and providing reliable support for subsequent genome assembly.

Polygala Linnaeus, 1753 is the largest genus within the family Polygalaceae, including about 600 taxa around the world (Chen et al. 2008). Genome size is one of the most extensively studied topics in plant systematics and angiosperm evolution (Jang et al. 2018), while within the family, only the genome of *P. tenuifolia* Willdenow, 1802 has been reported till now, of which the assembled genome is 769.62 Mb and the estimated genome size is 798.58 Mb (Meng et al. 2023). Only 11 species' genome sizes in *Polygala* have been recorded in the "Plant DNA C-value database" (Leitch et al. 2019; https://cvalues.science.kew.org/; release 7.1, April 2019), ranging from 0.84 pg (0.82 Gb, 2C-DNA value) in *P. amara* subsp. *brachyptera* (Chodat, 1889) Hayek, 1906 to 3.10 pg (3.03 Gb, 2C-DNA value) in *P. paucifolia* Willdenow, 1802, generally larger than *P. tenuifolia*.

According to the combined data in the "Chromosome Counts Database", CCDB: https://ccdb.tau.ac.il/ (Rice and Mayrose 2023) and new cytological studies (e.g. Shi and Niu 2017), nearly 100 *Polygala* species' chromosome numbers have been reported, with different ploidy status and primary basic chromosome number. Variations in chromosome numbers are considered as a major source of evolution and diversification in plants (Guerra 2008; Weiss-Schneeweiss and Schneeweiss 2013; Choi et al. 2019; Kang et al. 2021; Kim et al. 2024). Lewis and Davis (1962) found four primary basic chromosome numbers in *Polygala*: x = 6(12), 7(14), 8 and 10 and x = 17, 19 and 23 as secondary base. A perusal of the chromosomal data for the various species of Polygalaceae reveals a greater diversity in the chromosome numbers with 2n = 14, 16, 24, 28, 30, 32, 34, 36, 38, 40, 42, 46, 52, 56, 68, 72 and 96 (Bolkhovskikh et al. 1969; Moore 1973).

At present, comprehensive cytogenetic data, including chromosome number and genome size of *P. fallax* has not been reported. Therefore, the objectives of this study

were to estimate the genome size estimation and chromosome numbers of *P. fallax* through flow cytometry, k-mer analysis, and chromosome counts. The study revealed some basic genomic features of the genome, offering a foundation for subsequent whole-genome sequencing of *P. fallax*, and providing genomic information for further research on the synthesis mechanisms of its medicinal components, species conservation, and sustainable utilization.

Material and methods

Plant materials

The plant material of *Polygala fallax* were collected from natural populations in Ruyuan County, Guangdong Province, China (24.91255589°N, 113.03523594°E, alt. 890 m), then cultivated in the South China Botanical Garden, Chinese Academy of Sciences. In total, five individual plants were used for chromosome study. The voucher specimens were deposited in the Herbarium of South China Botanical Garden (IBSC).

Methods

Genome size estimation

Flow cytometry analysis is a characterization method in cytology, which determines the amount of nuclear DNA content in cells by measuring the fluorescence intensity proportional to DNA content, thereby offering a macroscopic viewpoint for observing genomic material (Mgwatyu et al. 2020). Three fresh plants of *P. fallax* were collected for flow cytometry analysis, and we used *Solanum lycopersicum* Linnaeus, 1753 (2C-DNA value = 0.90 Gb) as the internal reference standard (The Tomato Genome Consortium 2012). Seeds of tomato were provided by the Kunming institute of botany, Chinese Academy of Sciences, and one plant individual was used as material sources in the analysis.

Fresh leaf material of each studied individual was chopped with a razor blade, mixing with leaf material of an internal standard in LB01 buffer, and standing for 10 min. The resulting liquid with chopped plant material was filtered and the nuclei suspension was stained with propidium iodide (50 μg/mL, Sigma-Aldrich) and 50 μg/mL of ribonuclease A (RNase A, Boehringer), kept on ice for 0.5–1 h in dark environment (Doležel et al. 2007). The concentrations of the sample suspension is 2×10^5 particles/mL. Measurements were carried out with a BD FACSCalibur (BD Biosciences, San Jose, CA, USA). The fluorescence intensity of propidium iodide was detected by a standard 488 nm blue light excitation. One run can be done per preparation, in which 10000 particles were measured (CV% was controlled within 5%). The generated data was analyzed using Modifit 3.0 software (Becton Dickinson). The peak values corresponding to the G0/G1 phase of both the determination samples and internal standard were obtained and recorded. The DNA content of our samples was calculated according to the following formula: Sample DNA content = Reference DNA content ×

(sample 2C mean peak position/reference 2C mean peak position). Three independent replicates (runs) of each individual were analyzed.

Genome survey

Fresh leaves (over 50 g) were collected, cleaned with 75% ethanol, then quick-frozen with liquid nitrogen for genome survey. The analysis was conducted on Illumina NovaSeq 6000 platform. The genomic DNA was extracted according to the CTAB protocol (Doyle and Doyle 1987), and the purity and concentration of DNAs were tested. A paired-end library was constructed and sequenced using the Illumina NovaSeq 6000 platform to generate approximately 50× coverage of the estimated genome. Raw reads were trimmed using FastP v0.21.0 with the key parameters "–detect_adapter_for_pe" (Chen et al. 2018), and quality control of raw sequence data was launched using FastQC v0.11.9 with the key parameters "-o ." (Lo et al. 2014). The reads were filtered before assembly to ensure that a pair of paired-end reads had more than 90% of bases with quality ≥ Q30. Genome size, heterozygosity, and repeat content were estimated by analyzing the 19-mer frequency distribution of the clean reads with Jellyfish and GenomeScope (Marçais and Kingsford 2011; Vurture et al. 2017). K-mer depth distribution was counted and the peak value of the depth distribution was identified. The mean k-mer depth equals the peak value of the k-mer depth distribution as the depth of the k-mer coverage following a Poisson distribution. The k-mer count distribution from Jellyfish was used to produce a report and several informative plots describing the genome properties by GenomeScope 2.0 (Vurture et al. 2017). The heterozygous ratio and repeat sequence content were obtained with GCE software (Liu et al. 2013).

Chromosome number

Young root tips (1–2 cm long) from five individual plants were collected between 10:00–11:00 a.m. Root tip meristems were pretreated with 0.01% aqueous colchicine solution for 6 h at room temperature, fixed in 3:1 (v/v) absolute ethanol/glacial acetic acid for 24 h, then transferred to 70% ethanol and stored at –4 °C. For chromosome counting, fixed root tips were hydrolysed in 1M HCL at 60 °C for 50 min in a water bath, rinsed three times in distilled water and stained with carbol fuchsin for 5 min. Meristems were subsequently excised and squashed for microscope observations. The mitotic metaphase chromosomes were observed and recorded in an Olympus BX-43 microscope at 100× magnification with an Olympus DP26 camera, and the best cell images were selected for processing with Adobe Photoshop.

Data availability

The newly generated Illumina NovaSeq 6000 paired-end short-read sequencing data used for the *Polygala fallax* genome survey (i.e., the Illumina short-read dataset underpinning the k-mer–based genome profiling described in this study) have been deposited in CNGB database and are publicly available at https://ngdc.cncb.ac.cn/gsa/browse/CRA039476.

Results

Genome size estimation

For flow cytometry analysis, *Solanum lycopersicum* (2C-DNA value = 0.90 Gb) was used as the internal reference standard (The Tomato Genome Consortium 2012). Flow cytometry results of *Polygala fallax* is shown in Table 1 and Fig. 2. The fluorescence intensity of *S. lycopersicum* was 1563 (left peak of the histogram in Fig. 2) with the CV 3.9%, and that of *P. fallax* was 4323 (right peak of the histogram in Fig. 2) with the CV 3.6%. The fluorescence intensity ratio of *P. fallax* to the *S. lycopersicum* is 2.77, which suggests that the nuclear DNA content (2C-DNA value) of *P. fallax* is approximately 2.49 Gb.

Table 1. Genome size of *Polygala fallax* and *Solanum lycopersicum* determined by flow cytometry.

Species	Events	Fluorescence intensity	CV (%)	Genome size (Gb)	Mean value ± SD of genome size (Gb)
Polygala fallax	1465	4323	3.6	2.49	2.49 ± 0.97
Solanum lycopersicum	1311	1563	3.9	0.90	

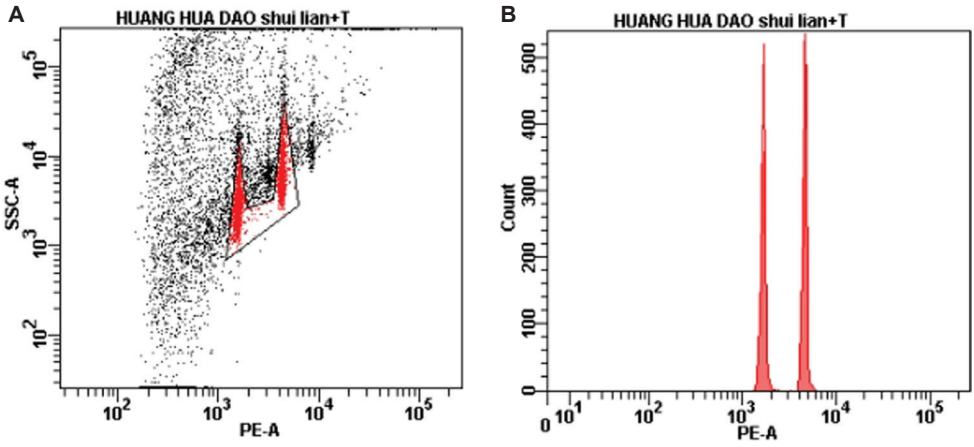

Figure 2. Flow cytometry analysis results of *Polygala fallax* and *Solanum lycopersicum* **A** Scatter plots of mixed samples of *P. fallax* and *S. lycopersicum* (after PI staining). **B** Histogram of mixed samples of *P. fallax* and *S. lycopersicum*.

Genome survey

Using paired-end sequencing of the genome with the Illumina NovaSeq 6000 platform, 110.40 Gb raw data were obtained, then we applied FastP filtering to remove the adapters and 109.79 Gb clean data was generated. The Q20 ratio of the sequencing data was above 99%, and the Q30 ratio was over 97% (Table 2), indicating a high quality of the sequencing data. The genome's GC content was 38.3%, which is moderate and it is unlikely to cause GC bias.

Table 2. Statistics of genome sequencing data of *Polygala fallax*.

Type	Number of read	Number of base	GC content (%)	Q20 (%)	Q30 (%)
raw data	736,021,030	110,403,154,500	38.36	99.44	97.30
clean data	732,956,988	109,792,276,089	38.30	99.58	97.57

In Totall 50,000 sequences were selected randomly from the clean data and compared for homology in the NT library using BLASTn (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.11.0/). The top 8 species belong to the family Polygalaceae and Fabaceae, and *P. fallax* has the highest sequence alignment rate of 0.27% (Table 3). Due to the unknown genomic information of *P. fallax* and the scarcity of gene annotations in the NT database, the comparison percentage is very low. According to the results, the sample library was aligned to DNA of closely related species, indicating that there was no significant exogenous contamination in the library, and the library construction and sequencing were successful.

Table 3. Genome sequence data NT library comparison of *Polygala fallax*.

Species	Family	Kingdom	Reads	Percentage (%)
Polygala fallax	Polygalaceae	Viridiplantae	135	0.27
Glycine soja	Fabaceae	Viridiplantae	128	0.26
Cicer arietinum	Fabaceae	Viridiplantae	110	0.22
Trifolium occidentale	Fabaceae	Viridiplantae	92	0.18
Trifolium striatum	Fabaceae	Viridiplantae	90	0.18
Trifolium repens	Fabaceae	Viridiplantae	90	0.18
Lotus japonicus	Fabaceae	Viridiplantae	79	0.16
Polygala arillata	Polygalaceae	Viridiplantae	72	0.14

Using 109.79 Gb of effective sequencing data, K = 19 was selected for analysis. After processing and analyzing the sequencing reads using Jellyfish, the total number of k-mers was 85,526,950,325. The k-mer frequency distribution was plotted using GenomeScope v.2, and the k-mer frequency distribution curve is shown in Fig. 3, with the horizontal axis representing k-mer depth and the vertical axis representing k-mer frequency. The figure shows that the k-mer distribution curve has a main peak around the depth of 39.2, a heterozygosity peak near 19–20 (approximately half coverage of the main peak) and a duplication peak near 80 (approximately two times coverage of the main peak), consistent with a diploid genome. The tailing phenomenon of the duplication peak indicates the presence of repetitive sequences and a relatively big genome. Although the heterozygosity peak is not prominent as the duplication peak, GenomeScope v.2 further estimated the genome heterozygosity by quantifying the proportion of the "heterozygous peak", and repetitive sequence content of *P. fallax*, finding a genome heterozygosity of 1.11% (> 1%) and a repetitive sequence content of 73.61%, a high repeat sequence ratio (> 50%).

Based on the total number of k-mers and the main peak k-mer depth of 39.2, the preliminary calculated genome size of *P. fallax* was 2199.64 Mb (ca. 2.20 Gb), and the sequencing depth was estimated to be 49.91×. Therefore, the genome size obtained

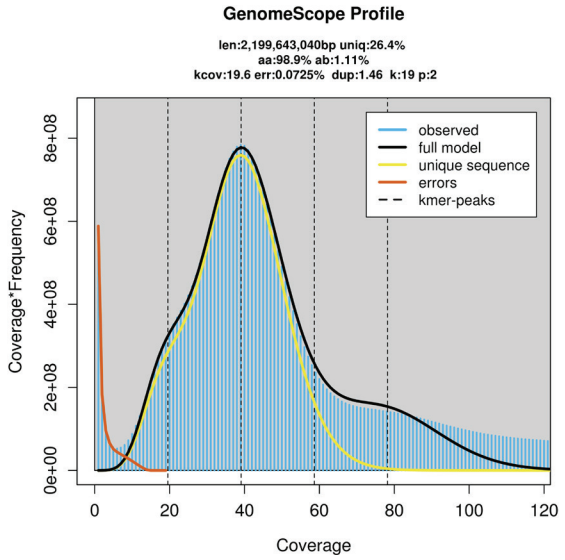

Figure 3. The 19 k-mer distribution of *Polygala fallax*.

through k-mer analysis is slightly smaller but close to the result estimated by flow cytometry (2.49 Gb), suggesting a relatively high reliability of the measurement results.

Chromosome number

For cytological study, young root tips (1–2 cm long) from five individual plants were collected and 20 metaphase plates were counted totally. All examined metaphase plates consistently showed 40 chromosomes. So *P. fallax* has the somatic chromosome number of 2n = 2x = 40 (Fig. 4) and the primary basic chromosome number is n = x = 20.

Figure 4. The chromosomes at metaphase in a root tip cell of *Polygala fallax*, 2n = 40. Scale bar: 10 μm.

Discussion

A genome encompasses all the information required for an individual's development and functioning. Chromosome number, ploidy, and genome size estimates in angiosperms have been widely used to better understand their genome information (Choi et al. 2019; Kim et al. 2024). Chromosome number of *Polygala fallax*, 2n = 2x = 40, was reported for the first time in this study. Previous chromosomal data of Polygalaceae reveals a great diversity in the chromosome numbers and the genus *Polygala* does not have a fixed primary basic chromosome number (Bolkhovskikh et al. 1969; Moore 1973; Table 4). After checking the reported chromosome number data in the CCDB (Rice and Mayrose 2023), only one species, *P. curtissii* A. Gray, 1867, has the same chromosome number as *P. fallax*. This indicates that the chromosome number of *P. fallax* is special and different from most of the species within the genus (Lewis and Davis 1962; Bolkhovskikh et al. 1969; Moore 1973; Rice and Mayrose 2023).

Genome size plays a crucial role in genome sequencing projects, as it determines the sequencing depth budget and offers an initial estimate for assessing the completeness of genome assembly (Mgwatyu et al. 2020). The genome size data of *Polygala* in the "Plant DNA C-values Database" (Leitch et al. 2019) and the chromosome numbers in the CCDB (Rice and Mayrose 2023) are shown in Table 4, together with the information of *P. fallax* and *P. tenuifolia* (Shi and Niu 2017). The comparative framework shows that *P. tenuifolia* has the smallest genome size and genome size of *P. fallax* is the third largest in the genus. The genome size of *P. fallax* is about three times larger than that of *P. tenuifolia* (2199.64–2490 Mb vs. 798.58 Mb), but it has only two more chromosomes than *P. tenuifolia* (Shi and Niu 2017; Meng et al. 2023). Data of other species also show that there is no significant correlation between chromosome number and genome size. For example, *P. paucifolia* has the largest genome size in the genus, while it possesses the same chromosome number (2n = 34) with *P. amara* subsp. *brachyptera*, which has a second smaller genome size (Temsch et al. 2010; Bai et al. 2012).

Table 4. Comparison of genome size and chromosome number of 13 species in *Polygala*.

No.	Species	2C-DNA value (Gb/pg)	Chromosome number, 2n
1	*Polygala tenuifolia*	0.80/0.82	38
2	*Polygala amara* subsp. *brachyptera*	0.82/0.84	28, 34
3	*Polygala angustifolia*	0.83/0.85	–
4	*Polygala calcarea* cv. *Lillet*	0.84/0.86	34
5	*Polygala vulgaris*	0.87/0.89	34
6	*Polygala sanguinea*	0.88/0.90	–
7	*Polygala microphylla*	1.03/1.05	32
8	*Polygala major*	1.05/1.07	34
9	*Polygala polygama*	1.28/1.31	28
10	*Polygala supina*	1.75/1.79	–
11	*Polygala fallax*	2.25-2.49/2.30-2.61	40
12	*Polygala vayredae*	2.65/2.71	28
13	*Polygala paucifolia*	3.03/3.10	34

–: No data were reported.

There are similar cases in other families, for instance, *Silene latifolia* Poiret, 1789 and *S. vulgaris* (Moench, 1794) Garcke, 1869 (Caryophyllaceae) are both diploid species and possess the same chromosome number (2n = 24), but differ in their genome size and mode of reproduction (Cegan et al. 2010). The lack of a direct correlation between chromosome number and genome size indicates that genome size is not strictly determined by chromosome count, and may also rely on the content of repetitive sequences and transposable element (Gregory 2005; Bennetzen and Wang 2014).

Genome size evolution in plants is a complex process shaped by various molecular mechanisms and evolutionary forces, leading to considerable variation across species. Formation of a large genome is primarily driven by whole-genome duplication events and the accumulation of repeated sequences, such as tandem repeats and transposable elements (Bennett and Leitch 2005; Niu et al. 2022; Duan et al. 2025). For instance, transposable elements, which can replicate and insert themselves into other genomic locations, can significantly contribute to genome expansion without changing the number of chromosomes (Lee and Kim 2014).

Besides the above reasons, the large genome size may also originate from ancient polyploidy followed by diploidization, which is strongly supported by previous genomic research (e.g. Vision et al. 2000; Conant and Wolfe 2008; Soltis and Soltis 2016). However, this phenomenon has not been reported in Polygalaceae. Until recently, there was no high-quality, chromosome-level reference genome for any Polygalaceae species. Except for the only one, that of *P. tenuifolia*, was published in 2023. This lack of data become a major obstacle for comparative genomics, gene family analysis, and evolutionary studies.

Genomic information has significant value on studying genetic resources of plant species, and it is especially crucial for elucidating synthesis mechanisms of various secondary metabolites. Meng et al. (2023) demonstrated that whole-genome duplications and tandem duplications play critical roles in the expansion of P450 and UGT gene families, greatly contributing to the synthesis of triterpenoid saponins, which were also considered as the main medicinal component of *P. fallax*. The complexities of how the large genome of *P. fallax* was formed and its precise biological function remain to be determined through future analyses. In this study, we present the genomic and cytological data of *P. fallax*, in order to provide an important reference for whole-genome sequencing and the selection of assembly strategies in subsequent steps.

Acknowledgments

This work was financially supported by National Natural Science Foundation of China (No. 32470221), Science and Technology Projects in Guangzhou (No. E33309), Department of Science and Technology of Guangdong Province (No. 2024A1515011508), Siberian Institute of Plant Physiology and Biochemistry SB RAS (No. 125021902461-9) and International Partnership Program of Chinese Academy of Sciences (Grant No. 068GJHZ2022055FN).

References

Bai C, Alverson WS, Follansbee A, Waller DM (2012) New reports of nuclear DNA content for 407 U.S. plant species. Annals of Botany 110(8): 1623–1629. https://doi.org/10.1093/aob/mcs222

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms-progress, problems and prospects. Annals of Botany 95(1): 45–90. https://doi.org/10.1093/aob/mci003

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annual Review of Plant Biology 65: 505–530. https://doi.org/10.1146/annurev-arplant-050213-035811

Bolkhovskikh ZV, Grif VG, Zakharieva OI, Matveeva TS (1969) Chromosome numbers of flowering plants. Leningrad: Nauka, 927 pp.

Cegan R, Marais GAB, Kubekova H, Blavet N, Widmer A, Vyskot B, Dolezel J, Safár J, Hobza R (2010) Structure and evolution of Apetala3, a sex-linked gene in *Silene latifolia*. BMC Plant Biology 10: 180. https://doi.org/10.1186/1471-2229-10-180

Chen SF, Zhou YQ, Chen YR, Gu J (2018) Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17): i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen SK, Ma HY, Parnell JAN (2008) Polygalaceae. In: Wu ZY, Raven PH (Eds) Flora of China 11. St. Louis: Missouri Botanical Garden Press and Beijing: Science Press, 145 pp.

Choi B, Yang S, Song JH, Jang TS (2019) Karyotype and genome size variation in *Ajuga* L. (Ajugoideae-Lamiaceae). Nordic Journal of Botany 37(5): e02337. https://doi.org/10.1111/njb.02337

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nature Reviews Genetics 9(12): 938–950. https://doi.org/10.1038/nrg2482

Doležel J, Greilhuber J, Suda J (2007) Estimation of nuclear DNA content in plants using flow cytometry. Nature Protocols 2: 2233–2244. https://doi.org/10.1038/nprot.2007.310

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19(1): 11–15.

Duan L, Xia ZQ, Ji YW, Han LN, Mattapha S, Wang Y, Li YQ, Kruglikov NA, Hu J, Wang FG, Chen HF (2025) Chromosome-level genome assembly and annotation of the endangered species *Ormosia henryi*. Scientific Data 12: 1846. https://doi.org/10.1038/s41597-025-06004-6

Gregory TR (2005) The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. Annals of Botany 95(1): 133–146. https://doi.org/10.1093/aob/mci009

Guerra M (2008) Chromosome numbers in plant cytotaxonomy: Concepts and implications. Cytogenetetic and Genome Research 120(3–4): 339–350. https://doi.org/10.1159/000121083

Jang TS, Parker JS, Emadzade K, Temsch EM, Leitch AR, Weiss-Schneeweiss H (2018) Multiple origins and nested cycles of hybridization result in high tetraploid diversity in the monocot Prospero. Frontiers in Plant Science 9: 433. https://doi.org/10.3389/fpls.2018.00433

Kang S, Choi B, Jang TS (2021) Chromosome evolution in Korean *Carduus–Cirsium* taxa (Asteraceae). Cytologia 86(4): 375–384. https://doi.org/10.1508/cytologia.86.375

Kim H, Choi B, So S, Myeong HH, Jang TS (2024) Cytological and pollen fertility analysis of diploids and triploids in Korean *Disporum* species. Plant Systematics and Evolution 310: 11. https://doi.org/10.1007/s00606-024-01894-2

Klein LC, de Andrade SF, Cechinel Filho V (2012) A pharmacognostic approach to the *Polygala* genus: phytochemical and pharmacological aspects. Chemistry and Biodiversity 9(2): 181–209. https://doi.org/10.1002/cbdv.201000319

Lee SI, Kim NS (2014) Transposable elements and genome size variations in plants. Genomics and Informatics 12(3): 87–97. https://doi.org/10.5808/GI.2014.12.3.87

Leitch IJ, Johnston E, Pellicer J, Hidalgo O, Bennett MD (2019) Plant DNA C-values Database (Release 7.1). Royal Botanic Gardens, Kew. https://cvalues.science.kew.org/

Lewis WH, Davis SA (1962) Cytological observations of *Polygala* in Eastern North America. Rhodora 64: 102–113.

Liu BH, Shi YJ, Yuan JY, Hu XS, Zhang H, Li N, Li ZY, Chen YX, Mu DSH, Fan W (2013) Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. https://arxiv.org/abs/1308.2012. https://doi.org/10.48550/arXiv.1308.2012v1

Lo CC, Chain PSG (2014) Rapid evaluation and quality control of next generation sequencing data with FaQCs. BMC Bioinformatics 15(1): 366. https://doi.org/10.1186/s12859-014-0366-2

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6): 764–770. https://doi.org/10.1093/bioinformatics/btr011

Meng FB, Chu TZ, Feng PM, Li N, Song C, Li CJ, Leng L, Song XM, Chen W (2023) Genome assembly of *Polygala tenuifolia* provides insights into its karyotype evolution and triterpenoid saponin biosynthesis. Horticulture Research 10(9): uhad139. https://doi.org/10.1093/hr/uhad139

Mgwatyu Y, Stander AA, Ferreira S, Williams W, Hesse U (2020) Rooibos (*Aspalathus linearis*) genome size estimation using flow cytometry and k-mer analyses. Plants 9(2): 270. https://doi.org/10.3390/plants9020270

Moore RJ (1973) Index to Plant Chromosome Numbers 1967–1971. Utrecht: Oosthoek's Uitgeversmaatschappij B.V., 539 pp.

Ning L, Dela Cruz RY (2023) Prediction of interactions between chemical compounds of *Polygala fallax* Hemsl. and key proteins involved in PI3K/AKT/mTOR signaling pathway via molecular docking technology and mechanism of this plant contributing to anti-aging. 17th International Conference on Complex Medical Engineering (CME), Suzhou, China, 37–42.

Niu SH, Li J, Bo W, Yang WF, Zuccolo A, Giacomello S, Chen X, Han FX, Yang JH, Song YT, Nie YM, Zhou B, Wang PY, Zuo Q, Zhang H, Ma JJ, Wang J, Wang LJ, Zhu QY, Zhao HH, Li ZHM, Zhang XM, Liu T, Pei SR, Li ZHM, Hu Y, Yang YH, Li WZH, Zan YJ, Zhou LH, Lin JX, Yuan TQ, Li W, Li Y, Wei HR, Wu HX (2022) The Chinese pine genome and methylome unveil key features of conifer evolution. Cell 185(1): 204–217. https://doi.org/10.1016/j.cell.2021.12.006

Rice A, Mayrose I (2023) The Chromosome Counts Database (CCDB). In: Garcia S, Nualart N (Eds) Plant genomic and cytogenetic databases methods in molecular biology. Vol. 2703. New York: Humana, 123–129. https://doi.org/10.1007/978-1-0716-3389-2_10

Shi LR, Niu YL (2017) Karyotype Analysis of *Polygala tenuifolia* WilId. Lishizhen Medicine and Materia Medica Research 28: 767–768. [In Chinese]

Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. Current Opinion in Plant Biology 30: 159–165. https://doi.org/10.1016/j.pbi.2016.03.015

Temsch EM, Temsch W, Ehrendorfer-Schratt L, Greilhuber J (2010) Heavy metal pollution, selection, and genome size: The species of the Žerjav study revisited with flow cytometry. Journal of Botany 1: 596542. https://doi.org/10.1155/2010/596542

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485(7400): 635–641. https://doi.org/10.1038/nature11119

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in Arabidopsis. Science 290(5499): 2114–2117. https://doi.org/10.1126/science.290.5499.2114

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Han F, Gurtowski J, Schatz MC (2017) GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33(14): 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Wang MH, Liu XY, Wang ZL, Xu Q (2022) The extract of *Polygala fallax* Hemsl. slows the progression of diabetic nephropathy by targeting TLR4 anti-inflammation and MMP-2/9-mediated anti-fibrosis in vitro. Phytomedicine 104: 154251. https://doi.org/10.1016/j.phymed.2022.154251

Weiss-Schneeweiss H, Schneeweiss GM (2013) Karyotype diversity and evolutionary trends in angiosperms. In: Leitch IJ, Greilhuber J, Doležel J, Wendel JF (Eds) Plant Genome Diversity: Physical Structure, Behavior and Evolution of Plant Genomes, Vol. 2. Springer Verlag, Vienna, 209–230.

Xu KP, Huang W, Tan JB, Zhou YJ, Li FS, Huang ZH, Liu BC, Tan GS (2006) Study on the antihyperlipidemia effective constituent of *Polygala fallax* Hemsl. Zhong Yao Cai 29(1): 16–19. [In Chinese]

Yao ZR, Li Y, Wang ZW, Lan YY, Zeng TX, Gong HF, Zhu KM, Tang H, Gu SJ (2020) Research on anti-hepatocellular carcinoma activity and mechanism of *Polygala fallax* Hemsl. Journal of Ethnopharmacology 260: 113062. https://doi.org/10.1016/j.jep.2020.113062

ORCID

Lin Fu https://orcid.org/0000-0003-4933-221X
Denis A. Krivenko https://orcid.org/0000-0003-2658-1723
Lei Duan https://orcid.org/0000-0001-6152-5458

Supplementary material 1

Additional figure

Authors: Lin Fu, Ziyue Liu, Lina Han, Ting Zhang, Yan Jiang, Denis A. Krivenko, Jin Hu, Hongfeng Chen, Lei Duan

Data type: docx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/compcytogen.20.184176.suppl1