Postprint: Whole-Genome Size and Characteristics Analysis of Cliff Plants *Opisthopappus taihangensis* and *Opisthopappus longilobus*

Authors: Wang Yiling, Zang En, Zhang Hao, Liu Zhixia, Lan Yafei, He Shan, Hao Weili, Yanling Cao

Date: 2022-02-14T21:21:02Z

Abstract

Opisthopappus taihangensis and Opisthopappus longilobus are endemic perennial cliff herbs of the Taihang Mountains, important wild resources of the family Compositae with high economic and ecological value. To determine suitable whole-genome sequencing strategies for the two species, this study employed flow cytometry and high-throughput sequencing technologies to analyze genome size, heterozygosity, repetitive sequence content, and GC content. The results showed: (1) Flow cytometry estimated the genome size of O. taihangensis at 2.1 Gb and O. longilobus at 2.4 Gb; (2) After correction by high-throughput sequencing, the genome size of O. taihangensis was 3.13 Gb, with repetitive sequences accounting for 84.35%, heterozygosity of 0.99%, and GC content of 36.56%; the genome of O. longilobus was 3.18 Gb, with repetitive sequences accounting for 83.83%, heterozygosity of 1.17%, and GC content of 36.62%; (3) Following preliminary assembly, abnormal GC content distribution and average depth were observed, exhibiting stratification, which may be attributed to the high heterozygosity of the two species' genomes. From the perspective of genome structure, both O. taihangensis and O. longilobus belong to complex genomes characterized by high repetition, high heterozygosity, and large genome size; it is recommended to use an Illumina + PacBio sequencing and assembly strategy for whole-genome sequencing analysis.

Full Text

Genome Sizes and Characteristics of Cliff Plants *Opisthopappus taihangensis* and *O. longilobus* on the Taihang Mountains

WANG Yiling*, ZANG En, ZHANG Hao, LIU Zhixia, LAN Yafei, HE Shan, HAO Weili, CAO Yanling
(College of Life Science, Shanxi Normal University, Taiyuan 030031, China)

Abstract

Opisthopappus taihangensis and *O. longilobus* are perennial cliff herbs endemic to the Taihang Mountains and represent important wild germplasm resources in the Compositae family, possessing high economic and ecological value. To determine appropriate whole-genome sequencing strategies for these two species, this study employed flow cytometry and high-throughput sequencing technologies to analyze genome size, heterozygosity rate, repetitive sequences, and GC content. The results showed: (1) Flow cytometry estimated the genome size of *O. taihangensis* at approximately 2.1 Gb and *O. longilobus* at 2.4 Gb; (2) After correction via high-throughput sequencing, the genome size of *O. taihangensis* was 3.13 Gb, with repetitive sequences accounting for 84.35%, heterozygosity at 0.99%, and GC content at 36.56%; for *O. longilobus*, the genome size was 3.18 Gb, with repetitive sequences at 83.83%, heterozygosity at 1.17%, and GC content at 36.62%; (3) Following initial assembly, abnormal GC content distribution and average depth with stratification phenomena were observed, likely caused by the relatively high heterozygosity rates in both species. Based on these genomic structural characteristics, both *Opisthopappus* species possess large, complex genomes with high repetition and heterozygosity, suggesting that an Illumina + PacBio sequencing and assembly strategy should be employed for whole-genome sequencing analysis.

Key words: *Opisthopappus taihangensis*, *O. longilobus*, genome survey, genome size

Introduction

With the development and maturation of next-generation sequencing technology and third-generation single-molecule sequencing technology, the time cost of sequencing has continuously decreased, providing conditions and convenience for genome sequencing of different species (Aird et al., 2011; Shi et al., 2012; Li et al., 2019). Whole-genome determination can provide reference clues for genomic and evolutionary biology research, and lay the foundation for molecular biology, transcriptomics, and bioinformatics studies. Analyzing plant growth, adaptation, and evolution from the genomic level can greatly promote understanding of plants and accelerate the discovery, mining, and utilization of new genes (Bi et al., 2019; Li et al., 2020; Zhao et al., 2021). However, before large-scale deep

sequencing, it is necessary to conduct genome surveys to estimate plant genome size and complexity, understand basic genome characteristics in advance, reduce sequencing blindness, and select appropriate sequencing strategies and assembly software accordingly (Tang et al., 2015; Huo et al., 2018; Bi et al., 2019; Li et al., 2019; Zheng et al., 2020).

Opisthopappus taihangensis and *O. longilobus* are both perennial herbs of the genus *Opisthopappus* that grow exclusively in cliff crevices of the Taihang Mountains, representing typical cliff plants with excellent drought and cold resistance (Chai et al., 2018, 2020). As diploid plants (2n=18), they are relatively primitive species in the subtribe Chrysantheminae (Ye et al., 2021) and important wild germplasm resources in Compositae, likely containing numerous superior genes for cold and drought tolerance, making them excellent gene sources for Compositae germplasm innovation. However, information on whole-genome size and characteristics of these two species is lacking, hindering further genome sequencing work and related evolutionary biology research (Huo et al., 2018). Therefore, conducting whole-genome sequencing of *O. taihangensis* and *O. longilobus* is essential to reveal their adaptation, evolution, and resistance mechanisms at the molecular level and provide scientific theoretical support for the comprehensive and rational utilization of their economic value (Song et al., 2018).

Since the publication of the model plant *Arabidopsis thaliana* genome, more than 400 plant genomes have been sequenced (Chen et al., 2018, 2019; https://www.plabipd.de/index.ep), with many more species currently being sequenced, providing abundant reference information for plant whole-genome sequencing. Particularly, the completion of whole genomes for several Compositae species in the subfamily Carduoideae, including *Artemisia annua* (Shen et al., 2018), *Chrysanthemum nankingense* (Song et al., 2018), *Helianthus annuus* (Badouin et al., 2017), *Conyza canadensis* (Peng et al., 2014), and *Cynara cardunculus* (Scaglione et al., 2016), offers important insights for genome analysis of *Opisthopappus* species.

Phylogenetically, the genus *Opisthopappus* together with most groups of *Dendranthema* and *Ajania* constitute the Chrysanthemum group, with its position being closer to the subtribe Artemiisinae (Zhao et al., 2010; Zhao et al., 2010). Both *C. nankingense* (Artemiisinae) (Shen et al., 2018; Song et al., 2018) and *A. annua* (Artemiisinae) have complex genomes characterized by high repetition, high heterozygosity, and large size. What are the genomic characteristics of the two *Opisthopappus* species with close phylogenetic relationships? Do they exhibit similar genome features?

Therefore, this study employed flow cytometry (Arumuganathan & Earle, 1991; Doležel et al., 2007) and high-throughput sequencing technology to address the following questions: (1) estimate the genome sizes of *O. taihangensis* and *O. longilobus*; (2) determine and evaluate the whole-genome size and characteristics of both species. The results aim to comprehensively understand the genomic features of these two species, provide a basis for subsequent de novo sequencing and assembly strategies, and offer clues for mining drought- and cold-resistant

genes and utilizing their potential genetic resources (Huo et al., 2018).

Materials and Methods

Sample Collection In October 2019, seeds of *O. taihangensis* and *O. longilobus* were collected from Wangmangling (the growth site of *O. taihangensis*) and Huguan (the growth site of *O. longilobus*) in Shanxi Province and brought back to the laboratory. Germination was conducted in January 2020, followed by pot cultivation. In June, healthy individuals with good growth (three individuals per species) were selected, and intact leaves were collected, snap-frozen in liquid nitrogen, and stored at -80°C for later use (Song et al., 2018).

Flow Cytometry Detection For each species, 0.5 cm leaf samples were placed in a flat-bottom culture dish, and 400 μl of OTTO extraction buffer was added. The leaves were chopped vertically with a blade for 30-60 seconds, incubated at room temperature for 30-90 seconds, filtered, and then 1.6 ml of staining solution (staining buffer + PI + RNase stock solution) was added. After incubating in the dark at room temperature for 30-60 minutes, samples were analyzed using a Sysmex CyFlow® Cube6 flow cytometer.

Maize with a known genome size (approximately 2.3 Gb) was used as a control. Maize, *O. taihangensis*, and *O. longilobus* were first analyzed separately to detect the relative fluorescence intensity of each sample. Then, mixed samples of maize with *O. taihangensis* and maize with *O. longilobus* were analyzed. Based on the relative fluorescence intensity peaks of different samples and referencing the control genome, the genome sizes of *O. taihangensis* and *O. longilobus* were estimated.

DNA Extraction and Sequencing Genomic DNA was extracted from leaves of both species using a modified CTAB method. The purity, concentration, and integrity of the extracted DNA were assessed using a spectrophotometer and agarose gel electrophoresis (Zhao et al., 2020). DNA samples from both species (three samples each) were sent to Hangzhou Lianchuan Biotechnology Company for sequencing. DNA was randomly fragmented using a Covaris ultrasonic crusher, and the entire library was prepared through end repair, A-tailing, sequencing adapter ligation, purification, and PCR amplification. The constructed libraries were subjected to paired-end (PE) sequencing using Illumina Hiseq. Raw reads containing low-quality sequences with adapters can affect subsequent analysis, so raw reads were carefully filtered to obtain clean reads for subsequent analysis of genome size, heterozygosity, GC content, etc. Q20 and Q30 were used as quality metrics; when Q20 90% and Q30 80%, the sequencing data quality was considered good.

Contamination Assessment Contamination in genomic DNA samples not only reduces effective data volume but also affects the accuracy of genome sur-

vey analysis results, leading to errors in genome assessment and deviations in assembly strategies that impact subsequent genome assembly effectiveness (Zhao et al., 2020). To determine whether the extracted genomic DNA of *O. taihangensis* and *O. longilobus* was contaminated, 10,000 clean reads were randomly extracted from the filtered high-quality data and compared against the NCBI nucleotide database (NT) using Blast software. If homologous alignment was found, the sample was considered free of exogenous contamination.

Genome Characterization Estimation The K-mer method was used to estimate genome size (Liu et al., 2013; Chen et al., 2015). A K-mer depth-frequency distribution curve was plotted with K-mer depth on the x-axis and frequency on the y-axis to estimate K-mer depth values and predict genome size. Oligonucleotide sequences of length K extracted from sequencing data were analyzed with K=17 for both species. Genome size was calculated using the formula: Genome size = Total base pairs / Average sequencing depth = Total K-mer number / Average K-mer depth (Huo et al., 2018; Zhao et al., 2020).

A Bayesian model was used to iteratively correct heterozygosity and repetitive sequences based on K-mer frequency and depth values. Heterozygosity was calculated based on the percentage of heterozygous types, homozygous types, and total types. The percentage of repetitive sequences was calculated by comparing the area difference between the standard Poisson distribution and the actual data curve after the peak.

Genome Assembly Soapdenovo software (Vurture et al., 2017) was used to assemble clean reads into contigs and scaffolds with K=41. Assembled genome sequences were compared with raw reads to analyze GC content, contig coverage depth, length, and quantity distribution. Based on the GC depth distribution map of genome sequencing sequences, GC bias was analyzed. Generally, high or low GC regions show significant differences in sequencing depth compared to normal regions, with lower coverage. In this study, 10 kb non-overlapping windows were used to calculate GC content for both species.

Results

Flow Cytometry Estimation of Genome Size Individual analysis of maize, *O. taihangensis*, and *O. longilobus* showed relative fluorescence intensity peaks at 58, 56, and 41, respectively (Figure 1: A-C). Mixed samples of maize with *O. taihangensis* and maize with *O. longilobus* both showed peaks around 40 (Figure 1: D, E). Based on these results, the genome size of *O. taihangensis* was estimated at 2.1 Gb and *O. longilobus* at 2.4 Gb.

Seencing and Analysis Results Sequencing Yield Statistics Library construction yielded 99.94 Mb raw data for *O. taihangensis*, with 22.67 Mb high-quality data after filtering; for *O. longilobus*, raw data was 109.74 Mb,

with 80.49 Mb high-quality data after filtering. For both species, Q20 was above 97.42% and Q30 above 92.53%, with a sequencing error rate of 0.04% (normal range <0.05%), indicating good sequencing quality suitable for further analysis.

Sample Contamination Assessment Homology comparison of 10,000 randomly extracted clean reads against the NT database showed that for *O. taihangensis*, alignments to *Artemisia frigida*, *Chrysanthemum indicum*, *Helianthus maximiliani*, and *Chrysanthemum* x accounted for 1.56%, 0.72%, 0.54%, and 0.27% of reads aligned to NT, respectively. For *O. longilobus*, alignments to these four species accounted for 1.09%, 0.48%, 0.31%, and 0.09%, respectively. *A. frigida* and *C. indicum* belong to the same subtribe as the *Opisthopappus* species, with *A. frigida* being more closely related, resulting in higher alignment proportions. Due to the unknown genome information and minimal gene annotation for *Opisthopappus* species in the NT database, alignment proportions to other species were low. No abnormal alignments to animals or microorganisms were detected, indicating that the sequencing data were not contaminated and could be used for survey analysis.

Genome Size Estimation The K=17 curves for both species showed severe tailing, suggesting high proportions of repetitive sequences (Figure 2). A main peak appeared near depth=28 for *O. taihangensis*, yielding an estimated genome size of approximately 3.15 Gb using the formula Kmer-number/depth. The corrected genome size was 3.13 Gb, with heterozygosity at 0.99% and repetitive sequences at 84.35% (Table 1). For *O. longilobus*, the main peak appeared at depth=26, with an estimated genome size of 3.20 Gb, corrected to 3.18 Gb, heterozygosity at 1.17%, and repetitive sequences at 83.83% (Table 1). Both species thus possess genomes characterized by high repetition and heterozygosity.

Preliminary Assembly Results For *O. taihangensis*, assembly yielded 4,148,869 contigs with a total length of 1.19 Mb, Contig N50 of 445 bp, N90 of 114 bp, and maximum sequence length of 24,674 bp. Further assembly produced 3,885,802 scaffolds with a total length of 1.22 Mb, maximum length of 24,674 bp, Scaffold N50 of 510 bp, and N90 of 118 bp (Table 2). For *O. longilobus*, there were 4,776,945 contigs totaling 1.30 Mb, with Contig N50 of 408 bp, N90 of 113 bp, and maximum length of 24,198 bp. Scaffold assembly produced 4,453,317 sequences totaling 1.34 Mb, with maximum length of 24,198 bp, Scaffold N50 of 477 bp, and N90 of 116 bp (Table 2). The short Contig N50 and Scaffold N50 lengths may be due to heterozygosity rates above 0.99% in both species. The distribution plots showed clear peaks, with pre-main peaks representing heterozygous peaks and post-main peaks representing repetitive peaks. Both species showed homozygous peaks at approximately 20× depth, indicating complex genomes.

GC Content and Distribution GC content in both species fell almost entirely within 20-60%, concentrated around 30%. Specifically, GC content was 36.56% for *O. taihangensis* and 36.63% for *O. longilobus* (Table 1). No obvious abnormalities were observed, and GC content showed no significant bias. The

GC depth distribution could be divided into three layers: high, medium, and low depth regions (Figure 4). The medium depth region was approximately 50% of the high depth region, possibly related to heterozygosity in both species. During assembly, heterozygosity may cause single-assembly of homologous chromosome heterozygous regions, leading to stratification in GC content.

Discussion

This study preliminarily surveyed the genome size, heterozygosity rate, and GC content of the Compositae cliff plants *O. taihangensis* and *O. longilobus* using flow cytometry and high-throughput sequencing K-mer analysis (Song et al., 2018). The preliminary estimated genome sizes were 2.1 Gb and 2.4 Gb, respectively, which were corrected to 3.13 Gb and 3.18 Gb after sequencing analysis.

Among currently published Compositae genomes, the smallest is *Conyza canadensis* (Astereae) at only 335 Mb (Peng et al., 2014), while the largest is an Anthemideae species at approximately 138.88 Gb (Garcia et al., 2013). The closely related *A. annua* and *C. nankingense* have genome sizes of 1.74 Gb (Shen et al., 2018) and 3.07 Gb (Song et al., 2018), respectively. The genome sizes of *Opisthopappus* species align with Compositae genome characteristics (Garcia et al., 2013). The relationship between evolution and DNA content is complex, with larger genomes associated with higher endangerment levels (Vinogradov, 2003). Compared with *C. nankingense* and *A. annua*, *Opisthopappus* species grow in harsh cliff environments and are listed as national second-class endangered species.

Flow cytometry estimates were approximately 1 Gb smaller than K-mer analysis results, possibly due to using maize with a relatively small genome (2.3 Gb) as the control, whereas K-mer analysis based on mathematical calculations may be more comprehensive and accurate (Doležel et al., 2007; Wang et al., 2018). In *C. nankingense* genome analysis, flow cytometry yielded larger estimates than K-mer analysis (Song et al., 2018), while other plant genome surveys have also shown inconsistent results between the two methods, such as in *Ipomoea pes-caprae* (Huo et al., 2019) and *Hydrangea macrophylla* (Chen et al., 2021).

GC content in most published plant genomes ranges from 30-47% (Deng et al., 2013; Song et al., 2019; Yu et al., 2019). The GC contents of *O. taihangensis* (36.56%) and *O. longilobus* (36.63%) fall within this range. The GC content of *Cynara cardunculus* (same subfamily) is 32% (Scaglione et al., 2016), *A. annua* is 31.5% (Shen et al., 2018), and *C. nankingense* is 37.2% (Song et al., 2018).

Based on heterozygosity levels, genomes are classified as micro-heterozygous (0.5% heterozygosity < 0.8%), highly heterozygous (heterozygosity 0.8%), and highly repetitive (repetitive sequence proportion 50%) (Wu et al., 2014; Zhou et al., 2017; Wang et al., 2018). *O. taihangensis* and *O. longilobus* showed heterozygosity rates of 0.99% and 1.17%, respectively, with repetitive sequence proportions of 84.35% and 83.83%. Plant heterozygosity is affected by breed-

ing systems, with self-pollinating species generally showing lower heterozygosity than outcrossing plants (Wang et al., 2018; Du et al., 2019). *Opisthopappus* species are self-incompatible (Hu and Zhao, 2008) and can reproduce both sexually through seeds and asexually through new branches at stem nodes, resulting in certain heterozygosity levels. *A. annua* has heterozygosity of 1.0-1.5% and repetitive sequences of 61.57% (Shen et al., 2018), while *C. nankingense* also shows high heterozygosity with 69.6% repetitive sequences (Song et al., 2018). High proportions of repetitive sequences are a common feature of large-genome Compositae species (3 Gb) (Garcia et al., 2013), which contributes to the large genome sizes of *Opisthopappus* species.

In summary, both *O. taihangensis* and *O. longilobus* possess complex genomes characterized by high repetition, high heterozygosity, and large size.

Preliminary assembly using K-mer=41 yielded Contig N50 of 445 bp and Scaffold N50 of 510 bp for *O. taihangensis*, with maximum scaffold length of 24,674 bp. For *O. longilobus*, Contig N50 was 408 bp and Scaffold N50 477 bp, with maximum scaffold length of 24,198 bp. Based on these survey results, we recommend that future studies employ combined second- and third-generation sequencing technologies for genome sequencing and assembly, supplemented with Hi-C for chromosome-level assembly to obtain high-quality whole-genome maps for both species.

The genome size and characteristic information obtained in this study for *Opisthopappus* species lays the foundation for future fine genome mapping and provides reference for research and utilization of wild Compositae germplasm resources (Zhao et al., 2020).

References

AIRD D, ROSS MG, CHEN WS, et al., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries [J]. Genome Biol, 12(2): R18.

ARUMUGANATHAN K, EARLE ED, 1991. Estimation of nuclear DNA contents of plants by flow cytometry [J]. Plant Mol Biol Rep, 9(3): 229-241.

BADOUIN H, GOUZY J, GRASSA CJ, et al., 2017. The sunflower genome provides insights into oil metabolism, flowering and asterid evolution [J]. Nature, 546: 148–152.

BI QX, ZHAO Y, CUI YF, et al., 2019. Genome survey sequencing and genetic background characterization of yellow horn based on next-generation sequencing [J]. Mol Biol Rep, 46(4): 4303-4312.

CHAI M, WANG S, HE J, et al., 2018. De novo assembly and transcriptome characterization of *Opisthopappus* (Asteraceae) for population differentiation and adaption [J]. Front Genet 9:371.

CHAI M, YE H, WANG Z, et al., 2020. Genetic divergence and relationship among *Opisthopappus* species identified by development of EST-SSR markers

[J]. Front Genet, 11:177.

CHEN F, DONG W, ZHANG J, et al., 2018. The sequenced angiosperm genomes and genome databases [J]. Front Plant Sci, 9: 418.

CHEN F, SONG Y, LI X, et al., 2019. Genome sequences of horticultural plants: past, present, and future [J]. Hortic Res, 6:112.

CHEN SS, LI XY, FENG J, et al., 2021. Genome size and characterization of *Hydrangea macrophylla* based on flow cytometry and genome survey [J]. Jiangsu Agr Sci, 49(12): 39-44.

CHEN W, HASEGAWA D, ARUMUGANATHAN K, et al. 2015. Estimation of the whitefly *bemisia tabaci* genome size based on k-mer and flow cytometric analyses [J]. Insects, 6: 704-715.

DENG GT, LIU QB, JIANG JX, et al., 2013. Estimation of genome size of *Miscanthus floridulus* [J]. J Plant Genet Resour, 14(2): 339-341.

DOLEŽEL J, GREIHUBER J, SUDA J. 2007. Estimation of nuclear DNA content in plants using flow cytometry [J]. Nat Protoc, 2(9): 2233-2244.

DU ML, XU J, ZHU CR, et al., 2019. Chromosome number determination and genome size estimation of three *Bupleurum* species [J]. Jiangsu Agr Sci, 47(11): 191-193.

GARCIA S, LEITCH IJ, ANADON-ROSELL A, et al., 2013. Recent updates and developments to plant genome size databases [J]. Nucleic Acids Res, 42: D1159–D1166.

HU X, ZHAO EH, 2008. A preliminary study on distant hybridization of *Opisthopappus* with *Chrysanthemum* and *Ajania* [J]. Modern Agr Sci, 6: 13-14.

HUO KS, ZHAO DL, CHEN YL, et al., 2019. Analysis of genome size and characteristics of salt-tolerant plant *Ipomoea pes-caprae* (L.) R. Brown [J]. J Plant Genet Resourc, 20(3): 728-735.

LI GQ, SONG LX, JIN CQ, et al., 2019. Genome survey and SSR analysis of *Apocynum venetum*[J]. Biosci Rep, 39(6): BSR20190146.

LI JY, LU TQ, YANG JB, et al., 2020. Genome survey and analysis of SSR molecular markers on *Macaranga indica* and *M. denticulate* [J]. Guihaia, DOI:10.11931/guihaia.gxzw202001024.

LIU B, SHI Y, YUAN J, et al. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects [J]. arXiv: Genomics, 1308.

OBERPRIELER C, HIMMELREICH S, VOGT R. 2007. A new subtribal classification of the tribe Anthemideae (Compositae) [J]. Willdenowia, 37: 89-114.

PENG Y, LAI Z, LANE T, et al., 2014. De novo genome assembly of the economically important weed horseweed using integrated data from multiple

sequencing platforms [J]. Plant Physiol, 166: 1241–1254.

SCAGLIONE D, REYES-CHIN-WO S, ACQUADRO A, et al., 2016. The genome sequence of the outbreeding globe artichoke constructed de novo incorporating a phase-aware low-pass sequencing strategy of F1 progeny [J]. Sci Rep, 6: 19427.

SHEN Q, ZHANG L, LIAO Z, et al., 2018. The Genome of *Artemisia annua* provides insight into the evolution of Asteraceae family and artemisinin biosynthesis [J]. Mol Plant, 11: 776-788.

SONG C, LIU Y, SONG A, et al., 2018. The *Chrysanthemum nankingense* genome provides insights into the evolution and diversification of Chrysanthemum flowers and medicinal traits [J]. Mol Plant, 11:1482–1491.

SONG LX, LI GQ, JIN CQ, et al., 2019. Whole genome sequencing and development of SSR markers in *Apocynum cannabinum* [J]. J Plant Genet Resour, 20(5): 1309-1316.

TANG Q, MA XJ, MO CM, et al., 2015. Genome survey analysis in *Siraitia grosvenorii* [J]. Guihaia, 35(6): 786-791.

VINOGRADOY AE. 2003. Selfish DNA is maladaptive: evidence from the plant Red List [J]. Trends Genet, 19(11): 609-614.

VURTURE GW, SEDLAZECK FJ, NATTESTAD M, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads [J]. Bioinformatics, 33(14): 2202-2204.

WANG CR, YAN HD, LI J, et al., 2018. Genome survey sequencing of purple elephant grass (*Pennisetum purpureum* Schum 'Zise') and identification of its SSR markers [J]. Mol Breed, 38(7): 94-104.

WANG X, ZHOU JY, SUN HG, et al., 2018. Genomic survey sequencing and estimation of genome size of *Ammopiptanthus mongolicus* [J]. J Plant Genet Resourc, 19(1): 143-149.

WU YF, XIAO FM, XU HN, et al., 2014. Genome survey in *Cinnamomum camphora* L. Presl [J]. J Plant Genet Resourc, 15(1): 149-152.

YE H, WANG Z, HOU H, et al., 2021. Localized environmental heterogeneity drives the population differentiation of two endangered and endemic *Opisthopappus* Shih species [J]. BMC Ecol Evol, 2021, 56.

YU FL, HUANG M, ZHANG HB, et al., 2019. Genome survey and characteristic analysis of SSR in *Callicarpa nudiflora* [J]. China J Chin Mat Med, 44(18): 3974-3978.

ZHAO HB, CHEN FD, CHEN SM, et al., 2010. Molecular phylogeny of *Chrysanthemum*, *Ajania* and its allies (Anthemideae, Asteraceae) as inferred from nuclear ribosomal ITS and chloroplast trn L-F IGS sequences [J]. Plant Syst Evol, 284(3): 153-169.

ZHAO HB, MIAO HB, WU GS, et al., 2010. Intergeneric phylogenetic relationship of *Dendranthema* (DC.) Des Moul., *Ajania* Poljakov and their allies based on amplified fragment length polymorphism [J]. Sci Agr Sin, 43(6): 1302-1313.

ZHAO L, ZHU YH, WANG M, et al., 2021. Estimation of *Rehmannia glutinosa* genome size based on flow cytometry and genome survey analysis [J]. Chin Trad Herb Drugs, 52(3): 821-826.

ZHENG Y, HE ZK, YAO MJ, et al., 2020. Genome survey study of *Alpinia katsumadai* based on Illumina high throughput sequencing [J]. Chin Trad Herb Drugs, 51(13): 3530-3534.

ZHOU JY, WANG X, GAO F, et al., 2017. Genome Survey and SSR Analysis of *Ammopiptanthus mongolicus* [J]. Genom Appl Biol, 36(10):4334-4338.

Note: Figure translations are in progress. See original paper for figures.