Comparative Analysis of Chloroplast Genome Characteristics of Three Medicinal Panax Species from the Qinba Mountains (Postprint)

Authors: Peng Yidan, Ma Nan, Ye Yihan, Yixin Liu, Tan Qiyi, Chen Ying, Zhang Yuqu, Xinjie Yang

Date: 2024-02-07T23:05:22+00:00

Abstract

Using three medicinal plant species of the Panax genus from the Qinba Mountains as research subjects, we employed bioinformatics techniques to analyze their chloroplast genome characteristics and codon usage bias, aiming to clarify the chloroplast genome features and phylogenetic relationships of Panax japonicus var. major (Zhuzishen), Panax bipinnatifidus (Yuye Sanqi), and Panax elegans (Xiuli Jia Renshen), thereby providing a basis for the molecular identification of Zhuzishen medicinal material. The results showed that: (1) The chloroplast genomes of the three Panax medicinal plants exhibited a typical quadripartite structure, with total sequence lengths of 156,071~156,104 bp, an overall GC content of 38.10%, and high similarity in genome size. (2) A total of 133 genes were annotated, including 88 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. (3) The chloroplast codon usage bias was similar among the three Panax medicinal plants, with the third codon position predominantly ending in A/U nucleotides; the codon usage pattern was mainly influenced by natural selection, while also being affected by mutational pressure. (4) Phylogenetic analysis revealed close genetic relationships among the three Panax medicinal plants, with Panax elegans being more closely related to Panax bipinnatifidus. This study demonstrates a close phylogenetic relationship between Panax elegans and the source plant of Zhuzishen, a finding that holds significant importance for the development and utilization of Zhuzishen medicinal resources, and provides an important foundation for further research on the classification, phylogeny, and evolutionary mechanisms of Panax species.

Full Text

Analysis of Chloroplast Genome Characteristics of Three Medicinal Panax Species from the Qinba Mountains

Peng Yidan[12], Ma Nan[12], Ye Yihan[12], Liu Yixin[1], Tan Qiyi[1], Chen Ying[12], Zhang Yuqu[12], Yang Xinjie[12]*

[1]College of Pharmacy, Shaanxi University of Chinese Medicine, Xi'an 712046, China
[2]Shaanxi Qinling Application Development and Engineering Center of Chinese Herbal Medicine, Key Laboratory for Research of "Qin Medicine" of Shaanxi Administration of Traditional Chinese Medicine, Xi'an 712046, China

Abstract

This study investigates three medicinal Panax species from the Qinba mountainous region through bioinformatic analysis of their chloroplast genome characteristics and codon usage preferences. We characterized the chloroplast genomes of *Panax japonicus* var. *major*, *P. japonicus* var. *bipinnatifidus*, and *P. pseudoginseng* var. *elegantior* to clarify their genomic features and phylogenetic relationships, providing a foundation for molecular identification of *Panacis majoris rhizoma*. The results demonstrate: (1) All three species possess typical quadripartite chloroplast genomes spanning 156,071–156,104 bp with a total GC content of 38.10%, indicating high similarity in genome size. (2) Each genome contains 133 annotated genes, including 88 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. (3) Codon usage preferences are similar across the three species, with the third codon position predominantly ending in A/U, and usage patterns primarily influenced by natural selection rather than mutation pressure. (4) Phylogenetic analysis reveals close relationships among the three Panax species, with *P. pseudoginseng* var. *elegantior* showing a closer affinity to *P. japonicus* var. *bipinnatifidus*. These findings establish a close genetic relationship between *P. pseudoginseng* var. *elegantior* and the source plants of *Panacis majoris rhizoma*, which has significant implications for the development and utilization of medicinal resources and provides an important basis for further studies on Panax taxonomy, phylogeny, and evolutionary mechanisms.

Keywords: *Panax japonicus* var. *major*, *P. japonicus* var. *bipinnatifidus*, *P. pseudoginseng* var. *elegantior*, chloroplast genome, structural characteristics, codon preference

---

Introduction

The genus *Panax* represents a branch of the Araliaceae family, comprising approximately eight species distributed across eastern Asia, the Himalayan region,

Indochina, and North America. China hosts seven species (including one introduced cultivar) according to the *Flora of China* (Editorial Board of Flora of China, Chinese Academy of Sciences, 2013). Most *Panax* species serve as traditional Chinese medicinal materials, representing valuable pharmaceutical resources. In the Qinba mountainous region, three *Panax* species are primarily distributed: *Panax japonicus* var. *major* (bead ginseng), *P. japonicus* var. *bipinnatifidus* (pinnate-leaf ginseng), and *P. pseudoginseng* var. *elegantior* (elegant false ginseng). All three are documented in *Flora of China* as varieties of *P. pseudoginseng* (Editorial Board of Flora of China, Chinese Academy of Sciences, 1978). The rhizomes of these three medicinal plants share extremely similar morphological characteristics, appearing as beaded structures, and all can be commercially sold as bead ginseng products. Both *P. japonicus* var. *major* and *P. japonicus* var. *bipinnatifidus* are officially listed as source plants for the traditional Chinese medicine "bead ginseng" in the 2020 Chinese Pharmacopoeia. Due to morphological and distributional similarities, *P. pseudoginseng* var. *elegantior* is frequently confused and misused with the other two species in local practice. Therefore, comprehensive research is urgently needed to determine whether *P. pseudoginseng* var. *elegantior* could serve as a substitute or expanded source for bead ginseng.

Chloroplasts are ubiquitous plastids in plant cells that serve as the site of photosynthesis. As organelles unique to green plants, they possess a complete genetic system—the chloroplast genome (cpDNA)—capable of independently directing replication and other functions within the plant (Camiolo et al., 2015). Variation in chloroplast genomes among plant species plays a crucial role in studying population genetics and species identification. Recent advances in sequencing technology and bioinformatics tools have significantly propelled chloroplast genomics research. Scholars have proposed that chloroplast genome studies can provide fundamental insights into species identification, origin, evolution, genetic diversity, and resource conservation and utilization in *Panax* (Liu et al., 2022).

While the chloroplast genome of medicinal bead ginseng has been previously investigated (Sun, 2021), the comparative chloroplast genome characteristics and phylogenetic relationships among these three *Panax* species from the Qinba region remain unclear. This study employs high-throughput sequencing technology (Wang et al., 2018) to analyze the chloroplast genome features, codon usage preferences, and phylogenetic relationships of these three *Panax* medicinal plants, aiming to provide a scientific basis for the rational development and utilization of bead ginseng resources and molecular identification of these three medicinal species.

Materials and Methods

1.1 Experimental Materials Fresh, healthy leaves of *P. japonicus* var. *major*, *P. japonicus* var. *bipinnatifidus*, and *P. pseudoginseng* var. *elegantior* were collected from Honghe Valley, Meixian County, Baoji City, Shaanxi Province

(107°47 29.3928 E, 34°0 54.2844 N, altitude 2,276.65 m). Leaves were washed with pure water, dried, wrapped in tin foil, flash-frozen in liquid nitrogen, and stored at -80°C. Voucher specimens were deposited at Shaanxi University of Chinese Medicine with accession numbers 202020203, 202020204, and 202020205.

1.2 Genomic DNA Extraction and Sequencing Total DNA was extracted using a plant DNA extraction kit (TIANGEN, Beijing). DNA quality was assessed via agarose gel electrophoresis and Qubit 3.0 fluorometry. Qualified DNA samples were sequenced on the NovaSeq6000 Illumina high-throughput sequencing platform. Raw data were filtered using CLC Genomics Workbench 22.0 to obtain clean data for assembly.

1.3 Chloroplast Genome Assembly and Annotation Using the chloroplast genome of a closely related species, *Panax japonicus* (GenBank ID: KP036469.1), as a reference, clean data were assembled with GetOrganelle (v1.6.4). Annotation was performed using CPGAVAS2, and the three annotated sequences were submitted to NCBI via BankIt. Physical maps of the chloroplast genomes were generated using the online organelle genome drawing tool Chloroplot (https://irscope.shinyapps.io/Chloroplot/).

1.4 Repeat Sequence Analysis SSR loci were identified using the online tool MISA (https://webblast.ipk-gatersleben.de/misa/) with minimum repeat thresholds of 10, 5, 4, 3, 3, and 3 for mono- to hexanucleotide repeats, respectively, and a minimum distance of $ $100 bp between two SSRs.

1.5 Comparative Chloroplast Genome Analysis The mVISTA online tool (https://genome.lbl.gov/vista/mvista/submit.shtml) was used to compare chloroplast genome sequences of six *Panax* species, using original FASTA format nucleotide sequences and annotation files as input.

1.6 Codon Usage Bias and Optimal Codon Analysis Protein-coding sequences (CDS) $ $300 bp were selected after removing duplicate genes, yielding 53 qualified CDS sequences for each
//sourceforge.net/projects/codonw/) and CUSP online software (https :
//www.bioinformatics.nl/cgi−bin/emboss/cusp) were used to analyze GC content at different codon positions,
(Yu et al., 2023).

1.7.1 Neutral Plot Analysis A scatter plot was generated with GC3 values on the X-axis and GC12 values (average of GC1 and GC2) on the Y-axis. Regression coefficients approaching 1 indicate mutation pressure as the primary driver of codon bias, whereas lower coefficients suggest natural selection as the dominant force (Mao et al., 2022).

1.7.2 ENC Analysis A two-dimensional scatter plot was created with GC3s content on the X-axis and observed ENC values on the Y-axis. The standard

curve representing expected ENC values was added, calculated as: ENC (expected) = 2 + GC3s + 29/[GC3s^2 + (1-GC3s)^2]. The ENC ratio was computed as (ENC expected - ENC observed)/ENC expected, representing the distance from the standard curve. Smaller ratios indicate weaker natural selection effects (Zhang et al., 2022).

1.7.3 PR2-Plot Analysis Base composition at the third codon position was analyzed by plotting A3/(A3+T3) on the Y-axis against G3/(G3+C3) on the X-axis. Even distribution across four quadrants would indicate equal usage of purines (A, G) and pyrimidines (C, T) at the third position, suggesting mutation-driven codon bias (Wang et al., 2022).

1.8 Phylogenetic Analysis Complete chloroplast genome sequences of 16 Araliaceae species were downloaded from NCBI. Together with the three newly assembled genomes (*P. japonicus* var. *major*: GenBank ID OL543604.1; *P. japonicus* var. *bipinnatifidus*: OL543605.1; *P. pseudoginseng* var. *elegantior*: OL543606.1), a total of 19 species were analyzed. Multiple sequence alignment was performed using MEGA 11, and a phylogenetic tree was constructed using the neighbor-joining method with the two-parameter model. Branch support was evaluated via 1,000 bootstrap replicates.

Results

2.1 Chloroplast Genome Structure and Characteristics The chloroplast genomes of all three species are double-stranded circular molecules with typical quadripartite structures. Genome sizes range from 156,071 to 156,104 bp, comprising a pair of inverted repeats (IRs: 25,989–25,994 bp), a large single-copy region (LSC: 86,097–86,114 bp), and a small single-copy region (SSC: 17,986–18,012 bp)

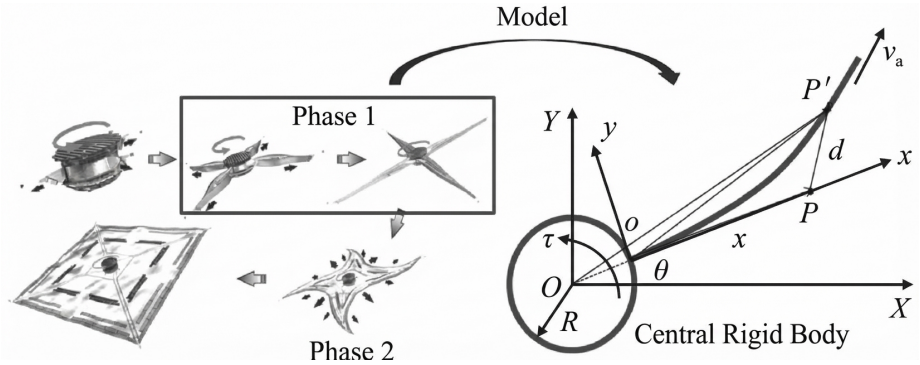

Figure 1: Figure 1

. GC content serves as an important indicator of phylogenetic relationships; all three species exhibit 38.10% total GC content with identical GC content

across IR, LSC, and SSC regions . While minor size variations exist, the overall chloroplast genome structures are highly similar.

Annotation identified 133 genes in each species, including 88 protein-coding genes (CDS), 37 tRNA genes, and 8 rRNA genes . Seven tRNA genes (trnL-CAA, trnV-GAC, trnI-GAU, trnA-UGC, trnR-ACG, trnN-GUU, trnI-CAU) and four rRNA genes (rrn23S, rrn4.5S, rrn16S, rrn5S) plus eight CDS genes (rpl2, rpl23, rps12, rps7, ndhB, ycf1, ycf15, ycf2) are duplicated in the IR regions. Additionally, genes trnK-UUU, trnG-UCC, trnL-UAA, trnV-UAC, trnI-GAU, trnA-UGC, rpl16, rpl2, rps16, rpoC1, atpF, ndhA, ndhB, petB, and petD each contain one intron, while rps12, ycf3, and ClpP contain two introns.

2.2 Simple Sequence Repeat (SSR) Analysis The chloroplast genomes contain 40, 38, and 38 SSRs in *P. japonicus* var. *major*, *P. japonicus* var. *bipinnatifidus*, and *P. pseudoginseng* var. *elegantior*, respectively [FIGURE:2]. Mononucleotide repeats are most abundant (22, 17, and 17), followed by tetranucleotide repeats (8 each). Dinucleotide repeats number 4, 8, and 8, while tri- and pentanucleotide repeats occur 3, 2, and 2 times, respectively. No hexanucleotide repeats were detected in *P. japonicus* var. *major*, but one was found in each of the other two species. Most SSRs utilize A/T, AT/AT, AAT/ATT, AAAT/ATTT, AATT/AATT, and AAAAT/ATTTT motifs, accounting for 62.50%, 76.32%, and 76.32% of total SSRs, respectively, demonstrating a strong preference for A/T bases. Notably, only *P. japonicus* var. *major* contains C/G and AAAAT/ATTTT type SSRs, distinguishing it from the other two species.

2.3 Comparative Chloroplast Genome Analysis Using *P. japonicus* var. *major* as the reference, whole-genome alignment revealed high similarity and low divergence among six *Panax* species

. Non-coding regions exhibited greater sequence variation than coding regions, while IR regions were more conserved than single-copy regions, indicating high conservation and stability within the genus. Major sequence divergence regions include trnH-GUG~trnR-UCU, rpoC1~trnT-GGU, ndhF~trnL-UAG, rps12, rpl22, and ycf1. *P. pseudoginseng* var. *elegantior* and *P. japonicus* var. *bipinnatifidus* show higher sequence similarity to each other than to *P. japonicus* var. *major*.

2.4 Codon Usage Patterns GC content varies across codon positions, following the trend GC1 > GC2 > GC3 . Third-position GC content (GC3) ranges from 30.59% to 29.77%, indicating a preference for A/U-ending codons. Effective number of codons (ENC) values range from 20 (strong bias) to 61 (no bias). The three species show ENC values of 49.30, 47.89, and 49.23, all substantially >45, suggesting weak codon usage bias.

Among 59 synonymous codons, 29, 28, and 28 show RSCU>1 (high-frequency codons), all ending with A/U . High-expression codons (ΔRSCU$ $0.08) number

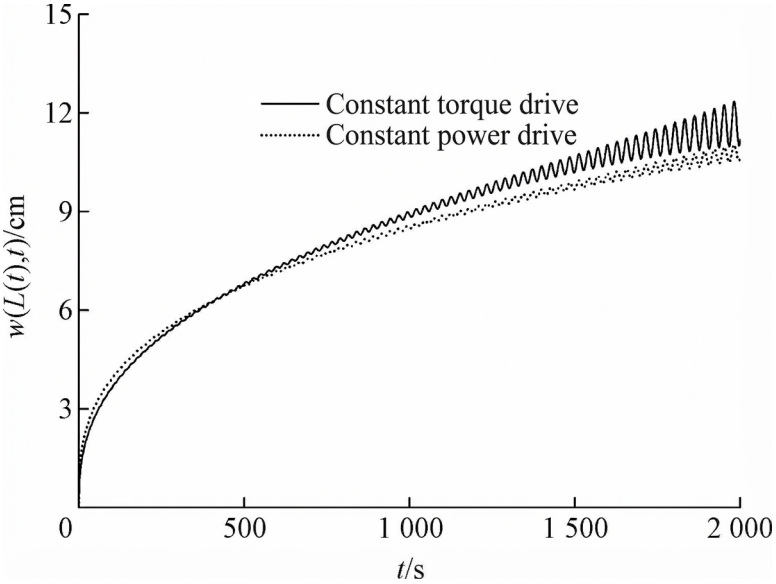

Figure 2: Figure 3

23, 23, and 24 . Intersection analysis identified 13 optimal codons common to all three species: GAA, GUA, UGU, AAA, AUU, CAA, CCU, UUA, GUU, UCU, ACU, GCU, and CGU. Five end with A and eight with U.

2.5.1 Neutral Plot Analysis Genes are predominantly distributed above the diagonal line [FIGURE:4]. Regression slopes between GC12 and GC3 range from 0.0111 to 0.0379 with $R^2>0$, indicating different mutation patterns between the first/second and third codon positions. Codon usage bias is primarily influenced by natural selection rather than mutational pressure.

2.5.2 ENC Plot Analysis Most genes cluster below the standard curve, with only a few on the curve [FIGURE:5], confirming that natural selection is the dominant force shaping codon preferences in these species.

2.5.3 PR2-Plot Analysis Gene distribution is non-uniform, concentrating in the lower-right quadrant [FIGURE:6]. Base usage frequencies show G>C and T>A, with pyrimidine usage exceeding purine usage at the third codon position. This pattern reflects combined effects of mutation and selective pressure on codon bias.

2.6 Phylogenetic Analysis To clarify relationships among the three Qinba *Panax* species, 19 Araliaceae chloroplast genomes were analyzed, including 16

downloaded from NCBI plus the three newly assembled genomes. The neighbor-joining phylogenetic tree

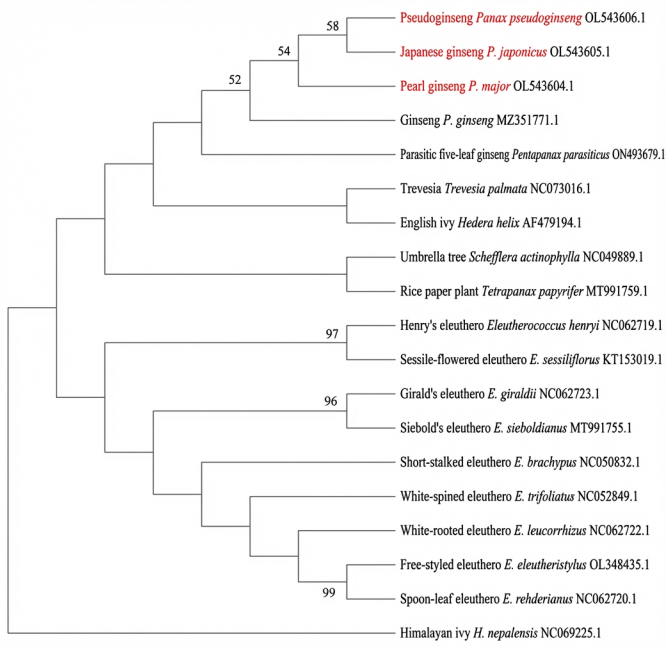

Figure 3: Figure 7

clusters *Panax*, *Trevesia*, *Hedera*, *Schefflera*, and *Tetrapanax* into one major clade, with *Acanthopanax* forming a separate clade. *Hedera nepalensis* is isolated as a distinct lineage. Within *Panax*, *P. pseudoginseng* var. *elegantior* and *P. japonicus* var. *bipinnatifidus* form a sister group to *P. japonicus* var. *major*, confirming the closer relationship between the former two species.

Discussion and Conclusion

This study presents comprehensive sequencing and comparative analysis of chloroplast genomes from three *Panax* medicinal species in the Qinba region. Genome lengths range from 156,071 to 156,104 bp, with size variations following the pattern SSC (26 bp) > LSC (17 bp) > IRs (5 bp), consistent with Yue (2021). The total GC content of 38.10% and higher, stable GC content in IR regions compared to SC regions reflect typical angiosperm chloroplast genome architecture (Zhang et al., 2012). Closely related species exhibit similar codon usage preferences (Sahoo et al., 2019). The A/U-ending codon bias primarily driven by natural selection aligns with patterns observed in other *Panax* species (Shi and Zhao, 2022), suggesting that congeneric species sharing similar environments may develop comparable codon preferences to adapt to common selective pressures.

Mutational "hotspot" regions in chloroplast genomes provide valuable genetic information for species identification (Li et al., 2018). The six variable gene fragments identified through mVISTA analysis show higher variability than universal plant DNA barcodes (rbcL, matK, and ITS) (Antil et al., 2023) and could serve as effective markers for *Panax* species identification and phylogenetic studies. SSRs are powerful tools for detecting genetic polymorphism at specific and whole-genome levels (Taheri et al., 2018; Ping et al., 2021). Identical SSR numbers and types in *P. japonicus* var. *bipinnatifidus* and *P. pseudoginseng* var. *elegantior* support their closer evolutionary relationship. The predominance of mononucleotide repeats and A/T base preference in SSRs and codons is consistent with other *Panax* species such as *P. notoginseng* and *P. quinquefolius* (Lu et al., 2022). Niu et al. (2017) demonstrated that A/T bases require less energy during mutation than G/C, explaining the A/T bias in SSRs and codons. The presence of two additional SSR types in *P. japonicus* var. *major* distinguishes it from the other species. Combined with consistent differences in genome size, gene composition, and GC3 content between *P. japonicus* var. *bipinnatifidus* and *P. pseudoginseng* var. *elegantior* versus *P. japonicus* var. *major*, these molecular features enable development of species-specific molecular markers for *P. japonicus* var. *major*.

Phylogenetic analysis provides crucial insights into evolutionary history. The current classification of these three species as varieties of *P. pseudoginseng* based on morphological characteristics of aerial parts (Editorial Board of Flora of China, 1978) has been maintained in *Flora of China*. According to this classification, *P. pseudoginseng* var. *elegantior* and *P. japonicus* var. *bipinnatifidus* are varieties of *Panax bipinnatifidus*, though supporting evidence remains unreported. Our NJ tree places *P. pseudoginseng* var. *elegantior* and *P. japonicus* var. *bipinnatifidus* as sister taxa to *P. japonicus* var. *major*, supporting their close relationship and providing scientific validation for the *Flora of China* classification. Chloroplast genomes thus prove highly effective for elucidating phylogenetic relationships within *Panax*. However, whether *P. pseudoginseng* var. *elegantior* can serve as an alternative or expanded source for pharmacopeial bead ginseng requires comprehensive evaluation of active compound content and clinical efficacy.

This study employs bioinformatic approaches to characterize chloroplast genomes of three *Panax* species from the Qinba region, establishing a close genetic relationship between *P. pseudoginseng* var. *elegantior* and the source plants of bead ginseng. This suggests potential for discovering similar or identical medicinal compounds, providing theoretical foundations for identifying new medicinal resources, developing novel drugs, and rationally utilizing traditional Chinese medicine resources such as bead ginseng.

References

Antil S, Abraham JS, Sripoorna S, et al., 2023. DNA barcoding, an effective tool for species identification: a review. *Mol Biol Rep*, 50(1):761-775.

Camiolo S, Melito S, Porceddu A, 2015. New insights into the interplay between codon bias determinants in plants. *DNA Res*, 22:461-470.

Editorial Board of Flora of China, Chinese Academy of Sciences, 2013. *Flora of China: Vol.13*. Beijing: Science Press:489-491.

Editorial Board of Flora of China, Chinese Academy of Sciences, 1978. *Flora of China Vol.54*. Beijing: Science Press: 179-181.

Li W, Liu Y, Yang Y, et al., 2018. Interspecific chloroplast genome sequence diversity and genomic resources in *Diospyros*. *BMC Plant Biol*, 18(1):210.

Liu C, Li M, Reng Y, et al., 2022. Characteristics and evolution of *Panax* chloroplast genomes. *Fujian J Agric Sci*, 37(7):886-896.

Lu Z, Tian W, Yang C, 2022. Chloroplast genome characteristics and codon usage bias analysis of *Panax* Linn. *Mol Plant Breed*:1-24.

Mao L, Huang Q, Long L, et al., 2022. Comparative analysis of codon usage bias in chloroplast genomes of seven *Nymphaea* species. *J NW For Univ*, 37(2):98-107.

Niu Z, Xue Q, Wang H, et al., 2017. Mutational biases and GC-biased gene conversion affect GC content in the plastomes of *Dendrobium* genus. *Int J Mol Sci*, 18(11):2307.

Ping J, Feng P, Li J, et al., 2021. Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*. *Ecol Evol*, 11(9):4786-4802.

Sahoo S, Dass S, Rakshit R, 2019. Codon usage pattern and predicted gene expression in *Arabidopsis thaliana*. *Gene*, 721:100012.

Shi Y, Zhao Y, 2022. Codon preference analysis of chloroplast genome of *Panax* L. *Mol Plant Breed*, 20(19):6350-6361.

Sun J, 2021. Comparative genomics study of *Panax japonicus* var. *major*, a medicinal plant of the genus Ginseng. Kunming: Kunming University of Science and Technology.

Taheri S, Lee Abdullah T, Yusop MR, et al., 2018. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. *Molecules*, 23(2):1-18.

Wang A, Wu H, Zhu X, et al., 2018. Species identification of *Conyza bonariensis* assisted by chloroplast genome sequencing. *Front Genet*, 11(9):374.

Wang F, Zhao W, Dong Z, et al., 2022. Analysis of chloroplast genome characteristics of *Prinsepia*. *Chin J Trop Crops*, 43(9):1759-1770.

Yu X, Song Y, Zhao Z, 2023. The complete chloroplast genome of *Elsholtzia fruticosa* (D. Don) Rehd. (Labiatae), an ornamental plant with high medicinal value. *Mitochondrial DNA Part B*, 8(3):336-341.

Yue J, 2021. Structural and phylogenetic analyses of the complete chloroplast genomes of four species within the genus *Panax* Linn. Hanzhong: Shaanxi University of Technology.

Zhang Y, Shen Z, Meng X, et al., 2022. Codon usage patterns across seven Rosales species. *BMC Plant Biol*, 22(1):65.

Zhang T, Fang Y, Wang X, et al., 2012. The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes. *PLoS ONE*, 7(1):e30531.