

Construction of a Molecular Traceability and Identification System for Edible Rose and Its Analogues Using Chloroplast Genomes and DNA Barcoding

Mengdi Zheng

hakuna_matata_mm@qq.com

Xi'an Medical University

Miaomiao Li

Xi'an Medical University

Yibo Ma

Shenyang Pharmaceutical University

Yangrui Li

Xi'an Medical University

Yulei Li

Xi'an Medical University

Research Article

Keywords: Rosa rugosa, Food authenticity, DNA barcoding, Chloroplast genomics, Chemical profiling

Posted Date: September 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10519297/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

1 **Title**
2 Construction of a Molecular Traceability and Identification System for Edible *Rose* and Its Analogues
3 Using Chloroplast Genomes and DNA Barcoding

4 **Authors**
5 Mengdi Zheng^{*1}, Miaomiao Li¹, Yibo Ma², Yangrui Li¹, Yulei Li¹

6
7 **Affiliations**
8 1. Xi'an Medical University, Xi'an, Shaanxi, 710021, China;
9 2. Shenyang Pharmaceutical University, Shenyang, Liaoning, 110016, China

10

11 **Corresponding Author**
12 Mengdi Zheng (Associate Professor)
13 Xi'an Medical University
14 Email: **hakuna_matata_mm@qq.com**
15 (This email address is identical to the one registered in the submission system.)

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

ABSTRACT

Ensuring the authenticity of edible *Rosa* raw materials is challenging due to morphological similarity and frequent adulteration of *Rosa rugosa* with related species. This study developed an analytical framework integrating chloroplast genomic signatures, DNA barcoding, and chemical profiling to authenticate *R. rugosa* and distinguish it from six congeners. Chloroplast genomes were compared for structural variations and codon usage bias. Three barcodes (ITS2, *matK*, *rbcL*) were evaluated via sequence divergence, phylogenetic reconstruction, secondary structure information, and machine learning classification. Six functional components were quantified in commercial samples. Results showed that chloroplast genomes are highly conserved but exhibit species-specific inverted repeat boundary variations and A/T compositional bias, with codon usage primarily shaped by natural selection (A/U-ending optimal codons). Among the three barcodes, ITS2 alone satisfied the barcode gap criterion and, when supplemented with its secondary structure, achieved complete species discrimination, whereas *matK* and *rbcL* showed limited resolution. Chemical profiling revealed cultivar-specific patterns: Kushui rose and *R. chinensis* were antioxidant-rich; Hotan and Pingyin roses displayed superior nutritional quality; and French rose exhibited uniformly lower values. This integrated method offers a reliable and practical tool for species-level authentication and quality screening of rose-derived food ingredients, effectively addressing the traceability demand in the rose food supply chain.

Keywords: *Rosa rugosa* · Food authenticity · DNA barcoding · Chloroplast genomics · Chemical profiling

INTRODUCTION

Rosa rugosa Thunb. (Rosaceae), a prominent species within the genus *Rosa*, is highly valued as a functional food ingredient and traditional edible resource (Wang 2023; Chinese Pharmacopoeia Commission 2025). Its flowers are widely consumed in various forms, including teas, beverages, jams, and confectionery products, and are also utilized in the production of essential oils and natural food flavorings. As a "medicinal and edible" plant, *R. rugosa* is rich in bioactive phytochemicals, notably flavonoids, polyphenols, and polysaccharides, which have been associated with significant antioxidant, anti-inflammatory, and cardioprotective activities, contributing to its growing popularity in the health food sector (Cui 2020; Jiang et al. 2003).

However, the commercial supply chain of rose-derived food products faces a considerable challenge of authenticity. The fresh or processed flowers of *R. rugosa* are frequently and inadvertently adulterated or substituted with those of closely related species and cultivars, including *Rosa chinensis*, *Rosa multiflora*, *Rosa gallica*, *Rosa davurica*, *Rosa sertata*, as well as cultivated varieties such as Kushui rose and Jinbian rose. These plant materials exhibit high morphological similarities in leaves, flower buds, and fruits, making visual identification error-prone (Jiang et al. 2003; Li et al. 2026; Xia et al. 2016). This issue is compounded during post-harvest processing, such as drying and grinding, where key diagnostic phenotypic features become obscured or destroyed. Traditional authentication methods, which rely on morphological markers like receptacle shape, style length, and leaf wrinkling, are therefore inadequate for ensuring the identity and purity of raw materials in the food industry. Consequently, the prevalence of adulteration with species such as *R. sertata*, *R. bella*, and *R. multiflora*

has been well-documented in market surveys, undermining both consumer trust and product quality (Luo et al. 2026; Cai et al. 2009; Shi et al. 2025; Guan et al. 2024).

DNA barcoding technology has been widely applied for authentication of plant resources, which could offer a more reliable solution to the adulteration issues in rose-derived food products (Zheng et al. 2026a; Zheng et al. 2026b). Previous studies have demonstrated that *matK* can distinguish *R. rugosa*, *R. chinensis*, and *R. multiflora* (Zhang et al. 2011), ITS2 effectively differentiates *R. chinensis* from *R. rugosa* (Zhang et al. 2021), and ISSR markers reveal rich genetic diversity within the genus (Xia et al. 2016). Nevertheless, the discriminatory power of single barcodes remains limited for closely related *Rosa* species.

Chloroplast genomes exhibit structural stability, maternal inheritance, and sequence conservation, offering unique advantages as "super-barcodes" for species identification (Jia et al. 2025a; Zhang et al. 2025). Codon usage bias, reflecting non-random synonymous codon usage influenced by mutation pressure, natural selection, and gene expression levels (Zhang et al. 2025; Dai et al. 2023), provides insights into evolutionary history and applications in genetic engineering (Yuan et al. 2025). In Rosaceae, codon preference in *R. chinensis* 'Old Blush' chloroplast genome was found to be primarily driven by natural selection (Hou et al. 2023), yet comparative studies among different *Rosa* species remain scarce.

Although numerous studies have reported on the chemical constituents of *Rosa* species and their close relatives, existing research has predominantly focused on single cultivars or isolated taxonomic groups (Jia et al. 2025b; Xie et al. 2022; Lu et al. 2026). Systematic studies integrating molecular identification, genomic comparison, codon usage bias analysis, and chemical composition determination remain scarce. In particular, comparative analyses of active constituent profiles and key indicator component contents among different cultivars and related species are still lacking, thereby constraining the precise evaluation and efficient utilization of *Rosa* germplasm resources.

To address these gaps, this study aims to: (i) evaluate the discriminatory capacity of ITS2, *matK*, and *rbcL* barcodes for *R. rugosa* and its relatives; (ii) compare chloroplast genome structural characteristics and identify hypervariable regions; (iii) systematically analyze codon usage bias patterns and their evolutionary mechanisms; (iv) determine contents of six functional and nutritional components in commercial samples; and (v) integrate these multi-level data to construct a comprehensive molecular–chemical authentication and quality evaluation system. This work will provide systematic molecular evidence for accurate identification, enrich *Rosa* chloroplast genome resources, and contribute to quality standard establishment and germplasm conservation.

3. MATERIALS AND METHODS

3.1 Plant Materials and Molecular Data Acquisition

Fresh leaves of *Rosa rugosa*, *R. chinensis*, *R. multiflora* (collected in Xi'an, April 2026), and *R. roxburghii* (purchased seedling) were prepared in triplicate, washed, surface-sterilized with 75% ethanol, frozen in liquid nitrogen, and ground for DNA extraction using the DP3-02 Kit (Tiangen Biotech). PCR amplification of ITS2, *matK*, and *rbcL* employed universal primers and reaction conditions as described (Zheng et al. 2026a; Zheng et al. 2026b). Complete chloroplast genomes of

seven *Rosa* species were retrieved from NCBI (MK641521.1, NC_065867.1, MG893867.1, OK509172.1, NC_079962.1, NC_032038.1, MT506347.1); CDS were extracted using CPGAVAS2 and screened (length ≥ 300 bp, intact codons, single-copy, no internal stops), yielding 54 genes (55 for *R. centifolia*). DNA barcode sequences were downloaded from NCBI, ITS2 ends trimmed using the ITS2 Database, and all sequences aligned with Clustal Omega. Laboratory-generated sequences were named using the format "species abbreviation + barcode type + number" (e.g., Rru-ITS2-1, Rru-matK-1, Rru-rbcL-1), with each barcode type for each species derived from three independent plant individuals (Table S1). Seven dried flower bud samples (French rose, Pingyin rose, Hotan rose, *R. chinensis*, Crimson Glory rose, Jinbian rose, Kushui rose) (Table S2 and Figure S1) were purchased in March 2026, authenticated by Dr. Mengdi Zheng, powdered (60-mesh), and stored at 4°C for chemical analysis.

3.2 Comparative Analysis of Chloroplast Genome Sequences

IR boundary dynamics were visualized using IRscope. SSRs were identified with MISA. Sequence variation was assessed by mVISTA alignment using *R. rugosa* (MK641521.1) as reference. Phylogenetic trees were reconstructed in MEGA12 using the Neighbor-Joining method with Kimura 2-parameter model and 1000 bootstrap replicates.

3.3 Chloroplast Codon Usage Bias Analysis

GC contents (GC1, GC2, GC3, overall) were calculated using EMBOSS CUSP; codon bias parameters (A3, T3, C3, G3, GC3s, CBI, Fop, CAI, RSCU, ENC) and correspondence analysis were computed using CodonW 1.4.2, with results analyzed within each software framework independently. Correlation analysis was performed using ChiPlot Heatmap module. Neutrality plots (GC3 vs. GC12), ENC-plots (ENC vs. GC3s with expected curve), and PR2-plots [$G3/(G3+C3)$ vs. $A3/(A3+T3)$] were constructed to assess evolutionary forces (Jia et al. 2025a). Optimal codons were determined by comparing RSCU values between top and bottom 10% ENC-ranked genes ($\Delta RSCU \geq 0.08$ and $RSCU > 1$). PCA was performed in GraphPad Prism 10.1.2 and Origin 2021 Pro using 59 synonymous codon frequencies (Zhang et al. 2025).

3.4 DNA Molecular Authentication

Sequence characteristics and divergence were analyzed in MEGA12. For ITS2, genetic distances were calculated using Kimura 2-parameter model, NJ trees were constructed, and secondary structures were predicted using the ITS2 Database; joint matrices were aligned in 4Sale and PNJ trees constructed in ProfDistS with TreeView visualization. For *matK* and *rbcL*, genetic distances and NJ trees were generated, and RNA secondary structures were predicted using RNAfold. Machine learning analysis employed BLOG algorithm for diagnostic site extraction and classification, and NaiveBayes, J48, SMO, and JRip classifiers on WEKA platform, evaluated by accuracy, error rate, and unclassified rate.

3.5 Chemical Content Determination

Six constituents (flavonoids, polyphenols, reducing substances as VCE, soluble sugars, polysaccharides, and amino acids) were quantified in dried flower buds. For the first four, 0.1 g powder was ultrasonicated in 10 mL 60% ethanol (40 °C, 150 W, 1 h) and centrifuged; supernatants were assayed by NaNO_2 – $\text{Al}(\text{NO}_3)_3$ – NaOH , Folin–Ciocalteu, phosphomolybdic acid, and ninhydrin methods,

respectively. Soluble sugars and polysaccharides were extracted separately and determined by anthrone–sulfuric acid (Xie et al. 2022; Lu et al. 2026). All tests were done in triplicate.

4. RESULTS

4.1 Comparative Analysis of Chloroplast Genomes

All seven *Rosa* chloroplast genomes exhibited the typical quadripartite structure, comprising LSC, SSC, and two IR regions (Fig. 1A). Total genome length varied minimally, from 156,591 bp in *R. chinensis* to 157,150 bp in *R. sertata*. The LSC region showed the greatest variation (85,707–86,233 bp), while SSC remained stable (18,753–18,820 bp). IR lengths were nearly identical within each species, ranging from 26,038 bp (*R. rugosa*) to 26,082 bp (*R. sertata*). At the LSC/IRb boundary (JLB), *rpl22* and *rpl2* flanked the junction, with *rps19* positioned 14–19 bp from the boundary; notably, *R. sertata* showed a distinct 64 bp distance for *rpl2*. The IRb/SSC boundary (JSB) occurred within *ndhF*, with 13–31 bp extending into IRb. At the SSC/IRa boundary (JSA), *ycf1* spanned the junction with 1,106–1,132 bp in IRa. The IRa/LSC boundary (JLA) featured *trnH* 3–12 bp from the junction.

SSR analysis detected 46–59 SSR loci across the seven species, with mononucleotide repeats predominating (82.61–97.93%), almost exclusively polyA/polyT, reflecting strong A/T bias (Fig. S2).. mVISTA alignment (Fig. 1B) using *R. rugosa* as reference revealed 67 conserved regions covering core photosynthetic genes (*psaA*, *psaB*, *rbcL*) and ATP synthase genes, with coding regions and conserved non-coding sequences showing high conservation; intergenic spacers, particularly near *trnM-CAU*, exhibited higher variation. Phylogenomic analysis based on complete chloroplast genomes resolved all seven species with 100% bootstrap support, dividing them into two clades: *R. chinensis*, *R. multiflora*, *R. gallica*, *R. centifolia*, and *R. roxburghii* formed a sister group, while *R. rugosa* and *R. sertata* clustered together (Fig. 2).

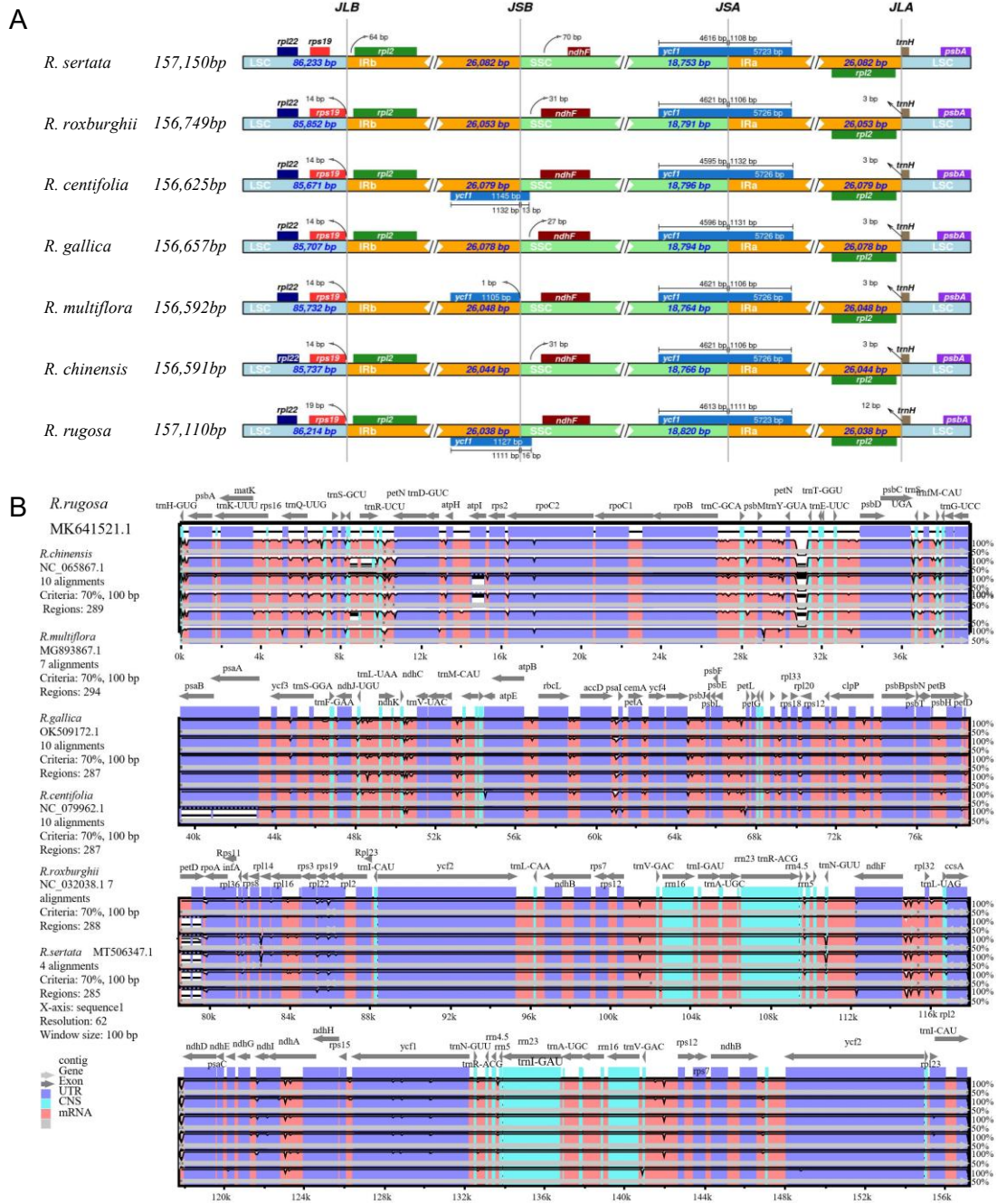


Fig. 1 Chloroplast genome features of seven *Rosa* species. **(A)** IR boundary dynamics across the LSC, SSC and IR regions (boxes and junctions not to scale); gene distances (bp) are marked at each junction. **(B)** mVISTA sequence identity plot with *R. rugosa* as reference; exons (dark blue), conserved non-coding sequences (red), and divergent regions (light blue) are shown. Coding regions are highly conserved across species, while intergenic spacers, especially *trnM*-CAU, display pronounced variation, serving as candidate markers for species discrimination.

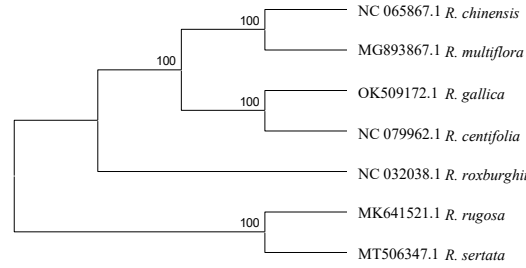


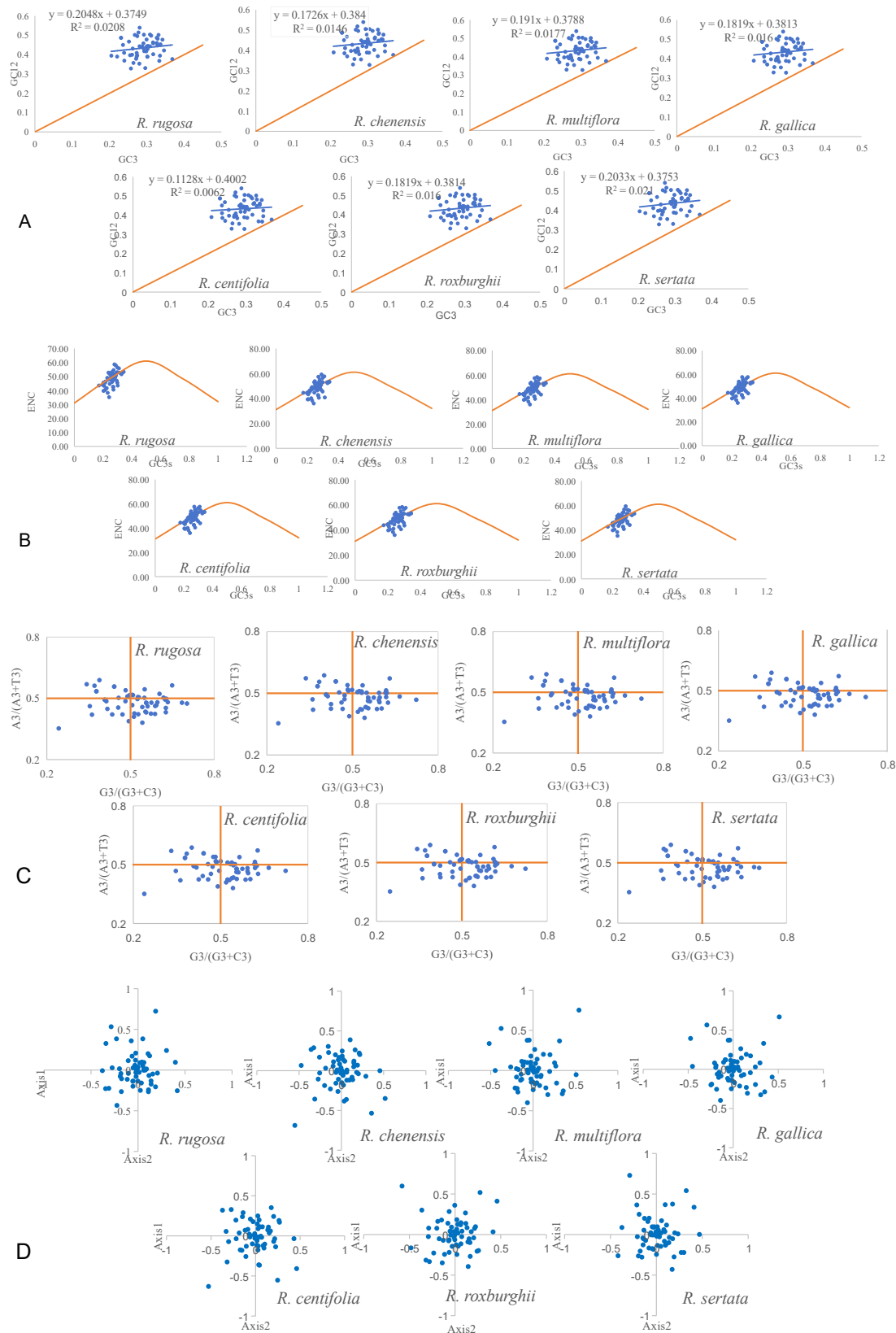
Fig. 2 NJ phylogenetic tree of seven *Rosa* species based on complete chloroplast genomes (Kimura 2-parameter, 1000 bootstrap replicates; support values at nodes). Two major clades are resolved, all with maximum bootstrap support, confirming robust intrageneric relationships.

4.2 Chloroplast Codon Usage Bias

GC3 content (29.15–30.53%) and GC3s (25.89–26.04%) were markedly lower than GC1 (45.77–47.15%) and GC2 (37.81–39.81%), indicating strong A/U preference at the third codon position. Overall GC (38.02–38.71%) and GC12 (41.73–41.81%) were highly conserved across species. CAI values were low (0.16994–0.17031), CBI was negative (–0.093 to –0.0915), Fop approximated 35%, and ENC ranged 48.27–48.42, collectively indicating weak codon usage bias in *Rosa* chloroplast genomes. Correlation analysis revealed strong positive correlations among CAI, CBI, and Fop; GC_{all} correlated significantly with GC1, GC2, and GC12; GC3 showed significant correlation with ENC ($p > 0.48$) in Table S3 and Figure S3.

Neutrality plots (Fig. 3A) showed all genes distributed above the diagonal with regression slopes far below 1 and R^2 values ranging 0.0062–0.021, indicating natural selection as the dominant evolutionary force. ENC-plot analysis (Fig. 3B) revealed most genes clustered near or below the expected curve, with *R. rugosa* showing greater dispersion and higher proportions in the [0.10, 0.30) ENC ratio interval (25.93%) compared to other species (~16.7%), suggesting stronger selective constraints. PR2-plot (Fig. 3C) analysis demonstrated consistent T > A and G > C preference at the third position across all species, with points predominantly in quadrant IV [$G3/(G3+C3) > 0.5$, $A3/(A3+T3) < 0.5$]. Correspondence analysis (Fig. 3D) revealed species-specific distribution patterns, with *R. multiflora*, *R. gallica*, *R. sertata*, and *R. rugosa* genes mainly in quadrant I, *R. chinensis* and *R. centifolia* in quadrant III, and *R. roxburghii* showing more dispersed distribution. Optimal codon analysis identified highly conserved optimal codon sets among the seven species, with UUU, UCA, UAA, ACA, AAU, CAU, and GGA shared universally, all ending in A/U (Fig. 4). PCA showed extensive overlap among species across PC1–PC4, with no species-specific clustering, confirming

216 overall conservation of codon usage patterns; *R. rugosa* exhibited typical and conserved patterns with
217 only minor deviations (Fig. 5).



219

220

Fig. 3 Codon usage bias analysis in seven *Rosa* chloroplast genomes. (A) Neutrality plots. All genes distribute above the diagonal with slopes $\ll 1$ ($R^2 = 0.0062\text{--}0.021$), indicating natural selection dominates. (B) ENC-plots. Points cluster at GC3s 0.2–0.5 (ENC 30–60), mostly near or below the expected curve; *R. rugosa* shows greater dispersion. (C) PR2-plots. Points predominantly fall in quadrant IV ($G > C$, $T > A$), with asymmetric distribution confirming selection-driven codon usage. (D) Correspondence analysis of RSCU. Species show distinct distribution patterns: *R. multiflora*, *R. gallica*, *R. sertata*, and *R. rugosa* in quadrant I; *R. chinensis* and *R. centifolia* in quadrant III; *R. roxburghii* with unique balanced distribution. Point 41 (codon GGG) occupies the most extreme position, notably in *R. sertata*, suggesting co-dominance of mutation pressure and natural selection.

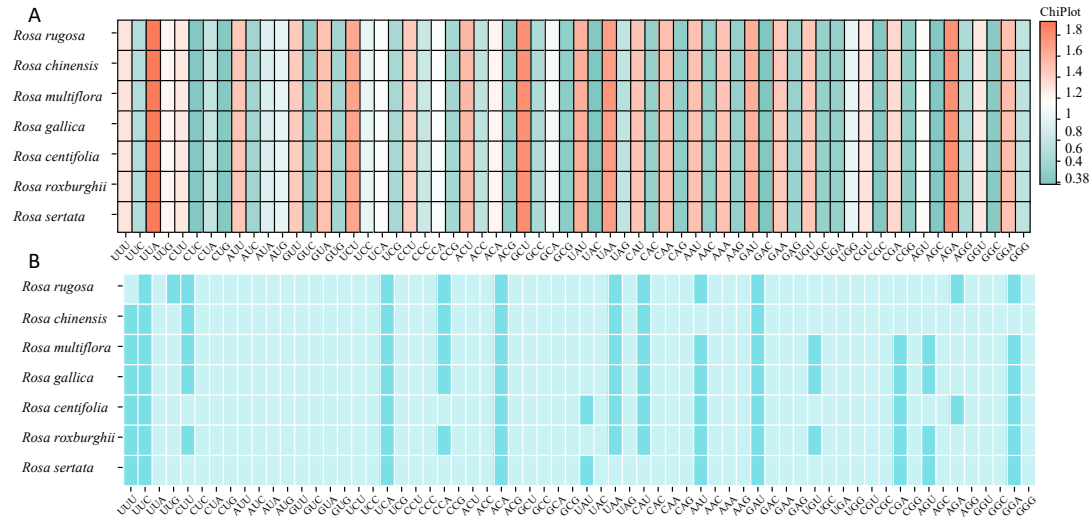


Fig. 4 Analysis of codon usage patterns in seven *Rosa* species chloroplast genomes. A: Heatmap of RSCU values for 59 sense codons. Note: The color scale indicates RSCU values, with red (RSCU > 1) representing frequently used codons and green (RSCU < 1) representing rarely used codons. B: Identification of optimal codons based on RSCU and Δ RSCU criteria. Note: Dark blue squares: optimal codons with RSCU > 1 and Δ RSCU ≥ 0.08 ; light blue squares: non-optimal codons with RSCU > 1 but Δ RSCU < 0.08.

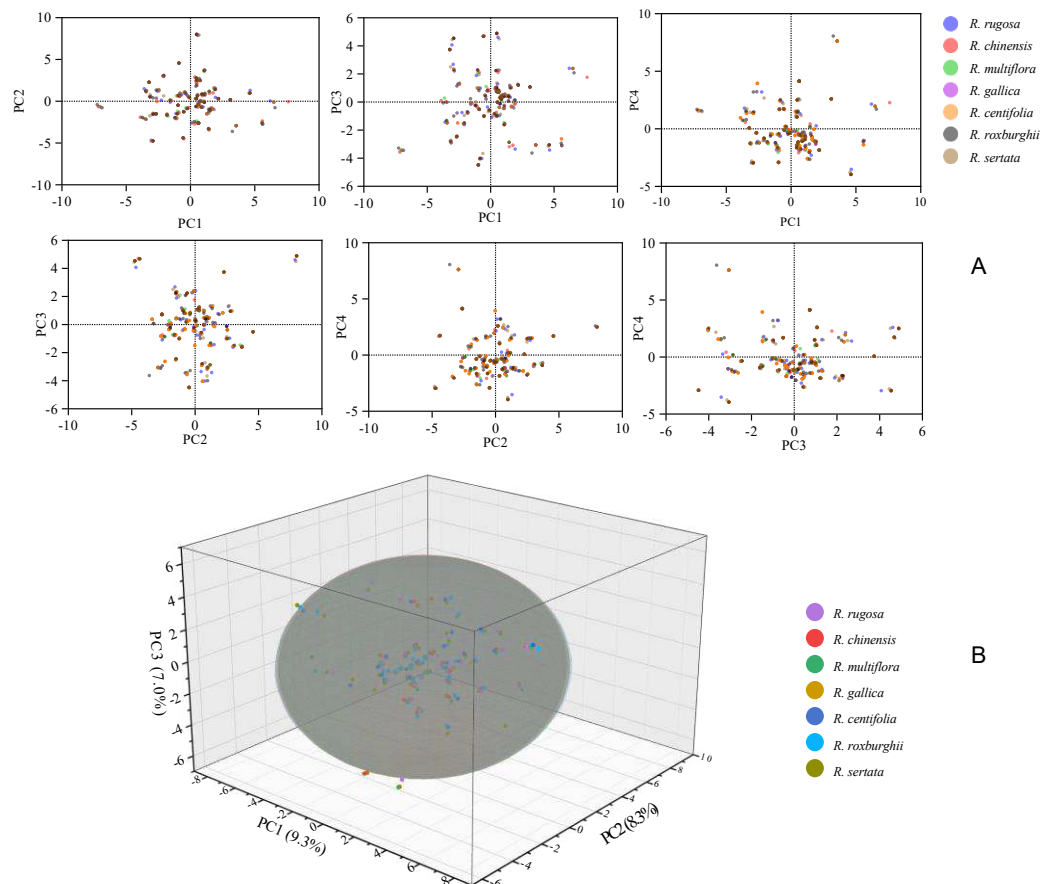


Fig. 5 Principal component analysis of codon usage patterns in seven *Rosa* species. (A) Two-dimensional PCA plots (PC1–PC2, PC1–PC3, PC1–PC4, PC2–PC3, PC2–PC4, PC3–PC4) based on RSCU values of 59 synonymous codons from chloroplast protein-coding genes. (B) Three-dimensional PCA plot (PC1–PC2–PC3). All seven species show extensive overlap across dimensions with no species-specific clustering, indicating highly conserved codon usage patterns among *Rosa* chloroplast genomes.

4.3 DNA Barcoding-Based Molecular Authentication

ITS2, *matK*, and *rbcL* sequences from 37, 38, and 38 samples respectively were analyzed. *matK* was longest (1369 bp) with the most variable and parsimony-informative sites (41.42%), followed by *rbcL* (709 bp, 47.11% variable sites), while ITS2 was shortest (344 bp) with the lowest variation (38.08%). ITS2 satisfied the barcode gap criterion with minimum interspecific distance (0.00798) exceeding maximum intraspecific distance (0.00533), enabling clear species discrimination. *matK* failed this criterion (minimum interspecific distance=0), and *rbcL* showed only a marginal gap (0.002052 vs. 0.001173) with limited discriminatory power (Table S4 and S5).

Phylogenetic analysis of ITS2 using NJ showed incomplete resolution with *R. rugosa* and *R. centifolia* samples intermixed (Fig. 6A). PNJ trees incorporating secondary structure information

markedly improved resolution, with all seven species forming distinct monophyletic clades: six *R. rugosa* samples clustered as a compact monophyletic group; *R. multiflora* and *R. chinensis* each formed independent branches; *R. gallica* formed a separate clade (86% support); and *R. sertata*, *R. roxburghii*, and *R. centifolia* each formed exclusive lineages (Fig. 6B). ITS2 secondary structures consistently comprised one central loop and four helical domains (I–IV), with *R. rugosa* exhibiting a characteristic "four-arm loop" architecture with elongated, symmetrically distributed side loops, providing structural basis for its distinct phylogenetic position (Fig. 7A). For *matK*, *R. rugosa* did not form an independent clade but clustered with some *R. centifolia* samples; *R. gallica* was dispersed; and *R. roxburghii* and *R. sertata* were intermixed (Fig. 6C); Fast-evolving *matK* shows variable stem-loop complexity among seven *Rosa* species. *R. rugosa*, *R. gallica*, and *R. roxburghii* are most complex (dense branches); *R. multiflora* and *R. sericea* are simplest (sparse); *R. chinensis* and *R. centifolia* are intermediate (Y- and T-shaped). Core loops are conserved; peripheral variations reflect molecular divergence (Fig. 7B). For *rbcL*, *R. rugosa*, *R. chinensis*, and *R. sertata* each formed independent branches, while *R. gallica* and *R. centifolia* clustered together, as did *R. multiflora* and *R. roxburghii* (Fig. 6D); Highly conserved *rbcL* has simpler, more elongated structures than *matK*, with the same complexity ranking but fewer stem-loops and more linear profiles. Core loops persist; overall simplification likely results from stronger selection pressure on this photosynthetic enzyme gene (Fig. 7C).

Machine learning analysis using BLOG, NaiveBayes, J48, SMO, and JRip classifiers demonstrated that BLOG achieved 100% correct classification rates for all three barcodes. Among other models, *matK* maintained stable 100% accuracy, ITS2 ranged 85.29–91.18%, and *rbcL* performed poorly at 60–65.71%; JRip showed the worst overall performance. Diagnostic nucleotide positions identified by BLOG revealed *matK* with the strongest species-specific variation, ITS2 relying on position 194 for discrimination, and *rbcL* with limited informative sites (Table S6 and Fig. 8).

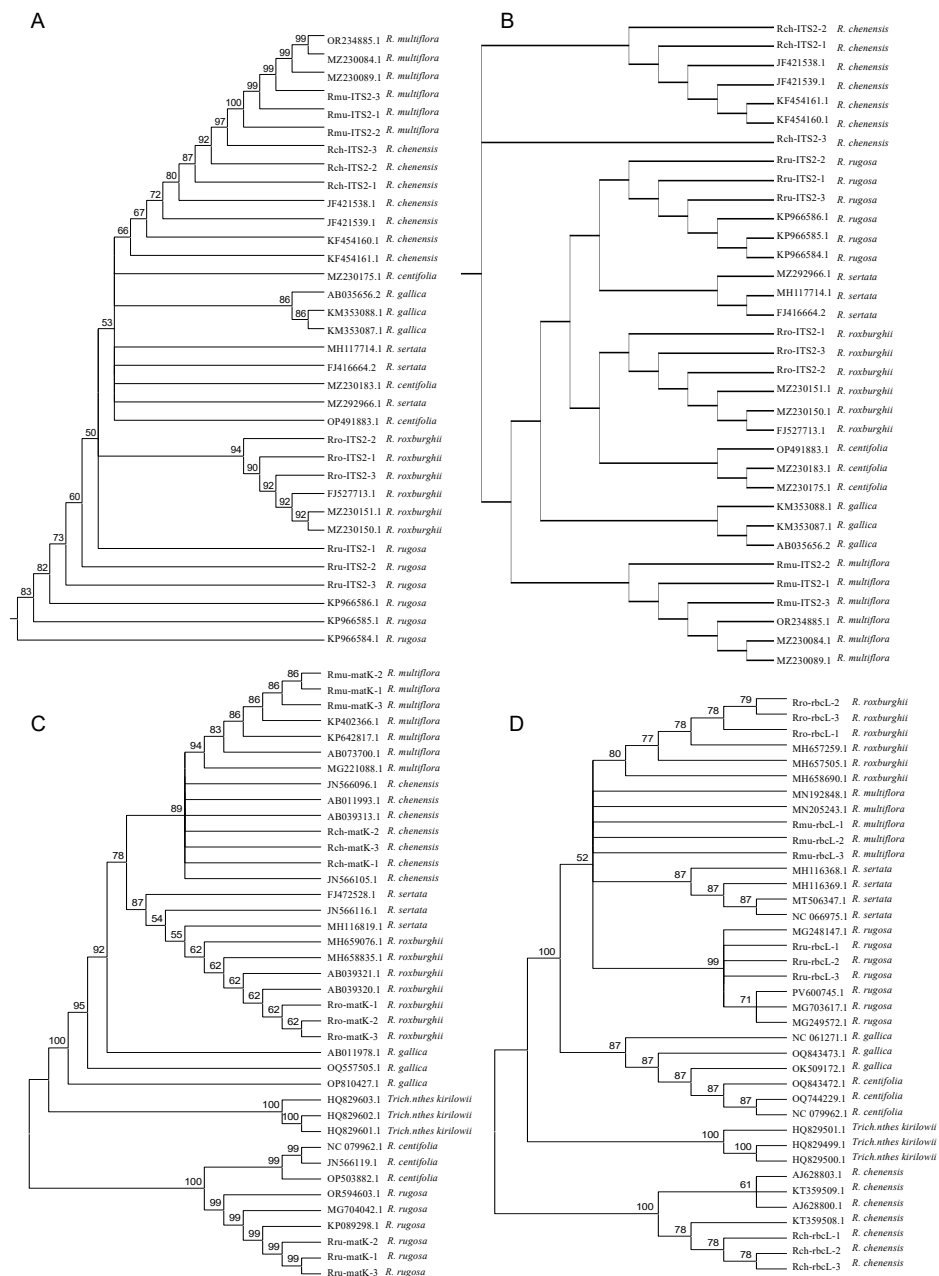


Fig. 6 Phylogenetic trees of seven *Rosa* species based on DNA barcodes. (A) NJ tree based on ITS2 sequences; (B) PNJ tree based on ITS2 sequence-structure matrices, showing improved resolution with secondary structure information. (C) NJ tree based on *matK* sequences; (D) NJ tree based on *rbcL* sequences. Numbers at nodes indicate bootstrap support values from 1000 replicates.

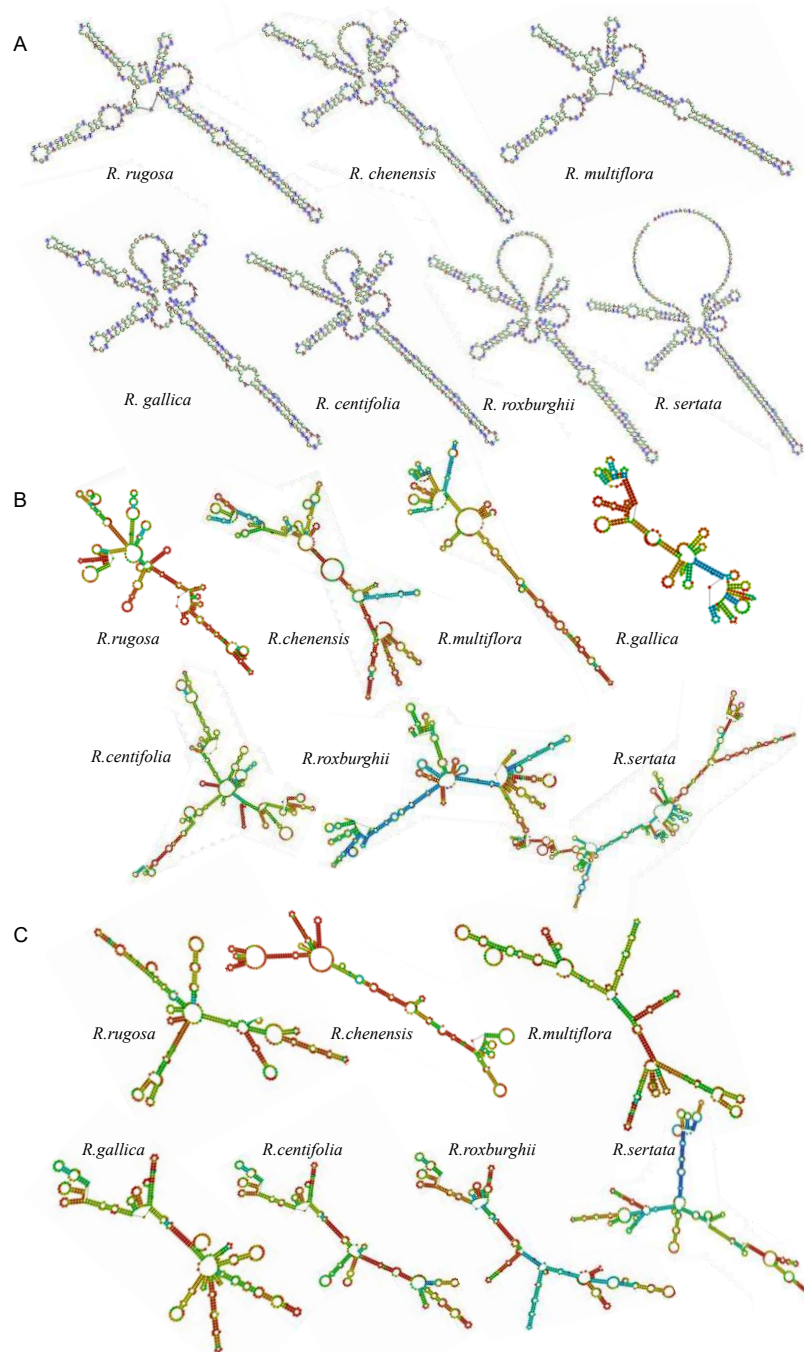


Fig. 7 Secondary structure predictions of DNA barcodes in seven *Rosa* species. (A) ITS2: four-arm loop with central loop and four helical domains. (B) *matK*: variable stem-loops and multi-branched structures. (C) *rbcL*: conserved backbone with divergent side branches. All seven species exhibit distinct species-specific secondary structures for each barcode.

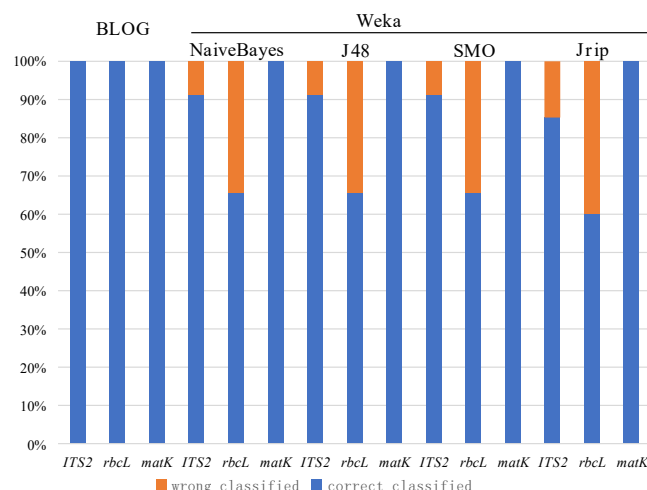


Fig. 8 Bar chart of machine learning recognition success rates based on three DNA barcodes using BLOG and WEKA classifiers. Blue and orange indicate correctly classified, incorrectly classified.

4.4 Chemical Composition Profiling

Six functional and nutritional components were determined in seven commercial *Rose* samples (Table 4). Total flavonoids ranged 26.84–53.25 mg/g, with Kushui rose highest and French rose lowest. Total polyphenols varied 16.7–80.0 mg GAE/g, with *R. chinensis* highest and French rose lowest. Total reducing substances ranged 1.59–3.50 mg/g. Soluble sugars were 0.49–0.86%, total polysaccharides 1.15–2.41%, and total amino acids 2.16–6.84%. Hotan rose and Pingyin rose showed superior comprehensive nutritional quality with balanced high levels of active and nutritional components. Kushui rose and *R. chinensis* were enriched in antioxidant-active components. French rose exhibited consistently low values across all indices. Jinbian rose and Crimson Glory rose showed distinct differential component profiles.

Table 1 Contents of functional and nutritional components in seven *Rosa* samples

Sample	Species / cultivar	TF (mg/g)	TP (mg GAE/g)	TRS (mg/g)	SS (%)	TPS (%)	AA (%)
French rose	<i>R. gallica</i>	26.84	16.7	1.59	0.76	1.42	4.68
Pingyin rose	<i>R. Rugosa</i> ‘Pingyin’	35.51	41.9	3.06	0.82	2.03	6.28
Hotan rose	<i>R. Rugosa</i> ‘Hotan’	39.49	40.4	2.32	0.86	2.41	6.84
<i>Rosa chinensis</i>	<i>R. chinensis</i>	47.75	80	3.5	0.69	1.15	5.6
Crimson Glory rose	<i>R. chinensis</i> ‘Crimson Glory’	40.16	46.8	2.17	0.58	1.39	2.16
Jinbian rose	<i>R. Chinensis</i> ‘JinBian’	30.75	44.7	2.47	0.49	1.69	6.25
Kushui rose	<i>R. rugosa</i> cv. Kushui	53.25	74.9	2.11	0.72	1.28	5.1
Mean ± SD		39.11±9.21	49.34±21.69	2.75±0.66	0.70±0.13	1.62±0.45	5.42±1.24

Note: total flavonoids (TF), total polyphenols (TP, as GAE), total reducing substances (TRS), soluble sugars (SS), total polysaccharides (TPS), and amino acids (AA). Mean ± SD is shown in the last row. Units: mg/g for TF, TP, TRS, AA; % for SS and TPS.

5. DISCUSSION

5.1 Chloroplast Genome Features and Codon Usage Bias

The chloroplast genomes of seven *Rosa* species exhibited highly conserved overall architecture, with species-specific variations at IR boundaries providing diagnostic markers for closely related species. The distinct spacing between *trnH* and the JLA boundary in *R. rugosa*, and the unique *rpl2-psbA* arrangement in *R. sertata*, represent neutral mutations accumulated in non-functional terminal regions under weak purifying selection, suitable for developing novel molecular markers (Zheng et al. 2026b; Sun et al. 2025). The overwhelming predominance of polyA/polyT mononucleotide SSRs reflects universal A/T bias in *Rosa* chloroplast genomes, while core photosynthetic genes and conserved non-coding sequences remained highly conserved; conversely, intergenic spacers such as *trnM-CAU* showed abundant variation, consistent with Rosaceae chloroplast genome evolution and providing sequence foundations for marker development and population genetics.

Codon analysis revealed GC3 contents of merely 29.15–30.53% and optimal codons universally ending in A/U, with neutrality plots confirming natural selection as the dominant force shaping codon usage patterns, mutation pressure playing only an auxiliary role. This conclusion aligns closely with previous codon bias studies in *R. chinensis* chloroplast genomes, corroborating shared evolutionary features within the genus (Jia et al. 2025a; Zhang et al. 2025). Species differed in selection intensity: *R. centifolia* experienced the strongest purifying selection, while *R. sertata* was relatively more influenced by mutation pressure; *R. rugosa* occupied intermediate positions for all parameters. Combined ENC-plot and PCA analyses indicated that *R. rugosa* codon usage was more prominently affected by mutation pressure, yet simultaneously exhibited the most typical and conserved patterns among all species, representing characteristic homogenizing evolution driven by maternal inheritance and long-term purifying selection in chloroplast genomes (Dai et al. 2023), with only *R. roxburghii* showing notable deviation. The A/T bias in *Rosa* chloroplasts directly impacts evolutionary rates and discriminatory capacity of non-coding sequences; future research should explore correlations between codon bias intensity and DNA barcode substitution rates. Furthermore, applying the elucidated codon usage patterns to optimize codons for key genes in flavonoid and polyphenol biosynthesis, combined with chloroplast expression systems featuring high expression efficiency and maternal inheritance, offers novel technical avenues for improving medicinal quality and enhancing active component production (Yuan et al. 2025; Cui et al. 2022; Gao et al. 2023).

5.2 Phylogenetic Relationships and DNA Barcoding Efficacy

The NJ phylogenetic tree based on complete chloroplast genomes divided seven *Rosa* species into two major clades, with *R. rugosa* and *R. sertata* as closest relatives, corroborating previous ISSR-based phylogenetic conclusions (Xia et al. 2016a). However, *matK*-based trees showed *R. rugosa* intermixed with *R. centifolia*, suggesting potential gene flow or shared maternal ancestry (Zheng et al. 2026a; Zheng et al. 2026b). This study systematically evaluated ITS2, *matK*, and *rbcL* using sequence characteristics, genetic distances, phylogenetic trees, and machine learning algorithms. Only ITS2 strictly satisfied the DNA barcode gap criterion, with interspecific distances significantly exceeding intraspecific distances and excellent resolution of closely related species. *matK* evolved too rapidly,

with intraspecific variation masking interspecific differences, while *rbcL* as a core photosynthetic gene was excessively conserved with insufficient informative sites; neither alone achieved accurate identification of *Rosa* relatives, consistent with previous barcode studies (Zhang et al. 2021). Integrating ITS2 secondary structure information into PNJ trees substantially improved resolution, demonstrating that secondary structures provide complementary discriminatory information beyond primary sequences. Machine learning results showed *matK* achieving 100% classification accuracy, followed by ITS2, with *rbcL* performing poorest. Collectively, ITS2 was identified as the core DNA barcode for *R. rugosa* and relatives, with *matK* and *rbcL* as auxiliary markers; combined usage constructs a more stable and accurate molecular identification system.

5.3 Multi-component Nutritional and Bioactive Composition Differences

All six determined components fell within previously reported normal ranges, with inter-cultivar variation patterns consistent with existing literature. Total flavonoids (26.84–53.25 mg/g) aligned with the 20–80 mg/g range for edible-medicinal roses; Kushui rose and *R. chinensis* showed elevated levels, while French rose was lowest, matching reports that high altitude and strong light promote flavonoid synthesis while oil-type roses accumulate less (Jia et al. 2025b). Total polyphenols (16.7–80.0 mg GAE/g) fit the 15–85 mg GAE/g general range for *Rosa* flowers; *R. chinensis* and Kushui rose were highest, French rose lowest, consistent with stronger phenolic metabolic activity in *R. chinensis*, stress-induced polyphenol synthesis, and metabolic antagonism between aromatic and phenolic compounds (Xie et al. 2022; Lu et al. 2026). Total reducing substances (1.59–3.50 mg/g) correlated significantly with flavonoid and polyphenol contents, consistent with antioxidant system studies. Soluble sugars (0.49–0.86%) matched the 0.4–1.0% reference range; Hotan and Pingyin roses accumulated more sugars due to ample light and large diurnal temperature variation in production areas, while Jinbian rose as a Yunnan characteristic cultivar showed weaker sugar synthesis. Total polysaccharides (1.15–2.41%) fell within the 1.0–2.6% range; Hotan and Pingyin roses as traditional edible types showed prominent polysaccharide accumulation, whereas Kushui rose prioritized antioxidant component synthesis with suppressed polysaccharide production, and *R. chinensis* inherently accumulated less polysaccharide than *R. rugosa*. Total amino acids (2.16–6.84%) agreed with the 3.0–7.0% normal range for dried roses; superior production environments conferred higher contents in Hotan and Pingyin roses, while Crimson Glory and French roses showed deficient accumulation, consistent with findings that native edible roses outperform introduced oil-type roses in nutrient accumulation. Collectively, inter-cultivar differences result from combined effects of genotype, cultivation environment, and breeding direction: Pingyin and Hotan roses balance nutritional and active components with optimal comprehensive quality; Kushui rose and *R. chinensis* are enriched in antioxidants; French rose shows uniformly inferior quality; Jinbian rose and Crimson Glory rose each possess distinctive germplasm characteristics. These compositional features serve not only for cultivar quality evaluation but also as effective auxiliary evidence for distinguishing *R. rugosa* and relatives.

5.4 Comprehensive Evaluation of the Multidimensional Authentication System

Traditional morphological identification relies on external characteristics that are easily lost after drying and processing, suffering from subjectivity and substantial error. Chloroplast genome analysis and DNA barcoding alone accomplish species discrimination at the genetic level without evaluating

practical application quality. Chemical composition testing intuitively reflects phenotypic and utilization value but cannot distinguish morphologically and genetically similar relatives. This study integrated chloroplast genome analysis, codon usage bias analysis, DNA barcoding molecular identification, and multi-component content determination to construct a molecular-chemical multidimensional authentication and quality evaluation system for *R. rugosa* and relatives. This system achieves precise identification of closely related species through molecular markers and quantitative quality assessment through multi-component contents, compensating for limitations of individual techniques while fulfilling dual functions of species authentication and quality evaluation with enhanced stability and practicality. It provides comprehensive scientific foundations for quality control of rose flower medicinal materials, conservation of *Rosa* germplasm resources, elite cultivar breeding, and deep-processing industry development.

6. CONCLUSIONS

The chloroplast genomes of seven *Rosa* species are highly conserved overall, with stable species-specific variations at IR boundaries and intergenic regions developable as novel molecular diagnostic markers. Chloroplast SSRs are predominantly polyA/polyT mononucleotide repeats, with whole genomes showing significant A/T bias. Codon usage patterns are primarily regulated by natural selection; optimal codons universally end in A/U, and *R. rugosa* exhibits conservative and typical codon usage patterns. The elucidated codon evolution principles provide theoretical support for *Rosa* chloroplast genetic engineering, heterologous gene codon optimization, and efficient expression.

Chloroplast genome-based phylogenetic trees indicate *R. rugosa* and *R. sertata* as closest relatives. Integrating sequence characteristics, genetic distances, phylogenetic trees, and machine learning results, ITS2 is identified as the core DNA barcode for *R. rugosa* and relatives, with its secondary structure further enhancing discriminatory resolution; *matK* and *rbcL* are limited by sequence variation or excessive conservation and serve only as auxiliary molecular markers.

Six classes of nutritional and active components in seven tested samples showed obvious cultivar specificity: Kushui rose and *R. chinensis* are prominent in antioxidant components; Hotan rose and Pingyin rose exhibit optimal comprehensive nutritional quality; French rose shows uniformly low component indices. These results corroborate multiple previous studies. Compositional characteristics provide effective auxiliary evidence for *Rosa* cultivar discrimination and quality evaluation.

This study successfully constructed a multidimensional molecular-chemical authentication and quality evaluation system for *R. rugosa* and relatives, overcoming limitations of single-technique approaches. It provides solid theoretical and technical support for authenticity identification of rose flowers, quality standard establishment, germplasm resource development, and industrialization.

Author contributions Mengdi Zheng: Conceptualization, Methodology, Supervision, Writing- original draft, Writing-review & editing, Funding acquisition, Project administration. Miaomiao Li: Investigation, Data curation, Visualization. Yibo Ma: Resources, Investigation, Formal analysis. Yangrui Li: Formal analysis, Investigation. Yulei Li: Data curation.

Funding Xi'an Medical University 2024 Annual Special Program for Scientific and Technological Capability Enhancement (2024NLTS122).

Data availability The authors confirm that the data supporting the findings of this study are available within the article and as its supplementary materials.

References

- Cai F, Sun S Q, Yan WR et al (2009) Analysis and identification of medicinal rose authenticity by infrared spectroscopy. *Spectrosc Spectr Anal* 29(9): 2429-2433.
- Chinese Pharmacopoeia Commission (2025) Pharmacopoeia of the People's Republic of China: 2025 Edition, Volume I. China Medical Science Press, Beijing.
- Cui WH, Du X, Zhong MC et al (2022) Complex and reticulate origin of edible roses (*Rosa*, Rosaceae) in China. *Hortic Res* 9:uhab051.
- Cui WQ (2020) Preliminary study on genetic differentiation of rose populations from different provenances based on SSR analysis. Master's Thesis, Shandong University, Jinan.
- Dai J, Liu Q, Xu X et al (2023) Comparative and phylogenetic analysis of the complete chloroplast genomes of *Uncaria* (Rubiaceae) species. *Front Plant Sci* 14:1271689.
- Gao CW, Li T, Zhao X et al (2023) Comparative analysis of the chloroplast genomes of *Rosa* species and RNA editing analysis. *BMC Plant Biol* 23:318.
- Guan HL, Huang BY, Yan XY et al (2024) Identification of distinct roses suitable for future breeding by phenotypic and genotypic evaluations of 192 rose germplasms. *Hortic Adv* 2(1):5.
- Hou Z, Li A, Huang CB (2023) Codon usage bias analysis of chloroplast genome of Chinese ancient rose 'Old Blush'. *Seed* 42(1):58-65.
- Jia HM, Li YW, Shi BS et al (2025) Study on antioxidant active components and stability of cold brew extract of rose. *Storage Process* 25(12):57-64,74.
- Jia XB, Wei JQ, Chen YW et al (2025) Codon usage patterns and genomic variation analysis of chloroplast genomes provides new insights into the evolution of Aroideae. *Sci Rep* 15:4333.
- Jiang JX, Li SF (2003) Identification of authenticity and quality of *Rosa rugosa* and *Rosa chinensis*. *Chin Tradit Herb Drugs* 34(7):667-668.
- Li KJ, Yuan M, Yang Q et al (2026) Herbal textual research on *Rosa rugosa*. *Chin Tradit Pat Med* 48(6):1989-1995.
- Lu J, Wang J, Wu QK, et al (2026) Evaluation of functional quality of petals from different edible and medicinal rose cultivars. *J Nanjing For Univ (Nat Sci Ed)* DOI: 10.12302/j.issn.1000-2006.202508030.
- Luo L, Lu X Y, Zhang C J et al (2026) Investigation, classification and breeding of *Rosa* germplasm resources in China. *Acta Horti Sin*, 1-28 [online]. DOI: 10.16420/j.issn.0513-353x.2025-1027.
- Shi QQ, Fan LP, Shen XJ et al (2025) Effects of different fermentation methods on quality and flavor of rose fermented beverage. *Food Ferment Ind* 51(22): 304-314. DOI: 10.13995/j.cnki.11-1802/ts.042182.
- Wang T (2023) DNA molecular identification and storage preservation of edible rose cultivars. Master's Thesis, Xinjiang University, Urumqi. DOI: 10.27429/d.cnki.gxjdu.2023.002528.

- 467 Xia Z, Zhou Y, Wang L J et al (2016) ISSR analysis and identification of *Rosa chinensis*, *Rosa*
468 *rugosa* and their related species. *Chin Tradit Herb Drugs* 47(24): 4397-4402. DOI:
469 10.7501/j.issn.0253-2670.2016.24.023.
- 470 Xie LM, Yan HJ, Tang KX et al (2022) Yield evaluation and nutritional composition analysis of four
471 main edible rose cultivars in Yunnan. *Southwest China J Agric Sci* 35(11):2627-2632. DOI:
472 10.16213/j.cnki.scjas.2022.11.021.
- 473 Yuan Y, He LX, Yuan BX et al (2025) Comparative Genomics and Phylogenetic Analysis of
474 Chloroplast Genomes in the *Hevea* Genus. *Genes* 16:201.
- 475 Zhang JS, Feng M (2025) Analysis of the Codon Usage Bias Pattern in the Chloroplast Genomes
476 of *Chloranthus* Species (Chloranthaceae). *Genes* 16:186.
- 477 Zhang P, Huang C L, Wang X L et al (2021) Analysis and identification of ITS and ITS2 sequences of
478 Kushui rose. *Chin Med Clin* 12(7):1190-1194. DOI: 10.13313/j.issn.1673-4890.20200423001.
- 479 Zhang W, Li Y (2011) DNA barcoding identification of *Rosa rugosa*, *Rosa multiflora* and *Rosa*
480 *chinensis*. *Jiangsu Agric Sci* 39(3):59,62. DOI: 10.3969/j.issn.1002-1302.2011.03.020.
- 481 Zheng MD, Gao MC, Zhang ZR et al (2026a) Integrating DNA barcoding and machine learning for
482 species identification: Comparative genomics and codon usage bias of chloroplasts
483 in *Gentiana* sect. *Cruciata*. *J Plant Physiol* 319:154730. DOI: 10.1016/j.jplph.2026.154730.
- 484 Zheng MD, Zhao Y, Gao MC et al (2026b) Chloroplast structure, codon usage bias, and machine
485 learning-based molecular identification using DNA barcoding of *Sophorae Tonkinensis* Radix et
486 Rhizoma (Shan Dou Gen) and its analogues. *Fitoterapia* 88:107005. DOI:
487 10.1016/j.fitote.2025.107005.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FoodAnalyticalMethodssupplementary0728.docx](#)