BMC Plant Biology

RESEARCH Open Access

Plastid genomic features and phylogenetic placement in *Rosa* (Rosaceae) through comparative analysis

Hui Jiang[1†], Shuilian He[1†], Jun He[2], Yunjuan Zuo[3], Wenling Guan[1], Yan Zhao[1], Xuejiao Li[1] and Jing Meng[1*]

Abstract

Background Species of the genus *Rosa* are among the commercially important exploited groups of ornamental plants in the world. Despite its wide application, the phylogenetic placement of many subgenera and sections of the genus is still unresolved due to hybridization, polyploidization, incomplete lineage sorting, low differentiation among the genus, and even their complex history of cultivation and breeding. Through more comprehensive taxon sampling, this study analyzed 18 representative *Rosa* plastid genomes, including 13 new sequences, to elucidate their phylogeny within the genus as well as the variation patterns in the plastid genomes.

Results The results revealed that the length of 106 complete *Rosa* plastomes varied between 156,333 bp and 157,396 bp, with closed circular tetrad structures of the SSC and LSC regions separated by two IR regions. Comparative analysis subsequently revealed high similarity in the total GC content, gene order and PCGs (79) of *Rosa* plastomes. No significant contraction or expansion of the IR boundary was noted in most *Rosa* species, except for the *trnH-GUG* gene, which is found mainly in the LSC region but crosses the IRa/LSC boundary in basal taxa of the *Rosa* phylogenetic tree. Abundant SSRs (73–87) and long repeat sequences (36–52) were detected in *Rosa* plastomes, and most of these repeats could be found within the IGS region. Eight IGS regions were identified as highly variable regions, which provides potential information for developing molecular markers. Nineteen genes were discovered to have undergone significant positive selection. Phylogenetic analyses based on PCGs and complete plastome sequences indicated that the genus *Rosa* was monophyletic well grouped into seven major clades with high bootstrap support. Most previously-defined subgenera and sections were paraphyletic.

Conclusions By assembling the largest known dataset of *Rosa* plastomes, the plastid genomic features across the genus were comprehensively studied before reconstructing a phylogenetic tree with a well-resolved backbone. However, the current study also shows the limitations of using plastomes to infer the phylogeny of some difficult taxa, and combining plastome, morphological and nuclear data together is recommended. This work offers valuable and basic sequence information for phylogenetic studies, species identification, *Rosa* species breeding and molecular genetics studies.

†Hui Jiang and Shuilian He contributed equally to this work.

*Correspondence:
Jing Meng
mengjing@ynau.edu.cn

Full list of author information is available at the end of the article

Keywords *Rosa*, Phylogeny, Plastid genome, Comparative analysis, Rosaceae

Background

The genus *Rosa* L. (roses; Rosoideae, Rosaceae) comprises approximately 150–200 species [1, 2], with ploidy levels ranging from diploid (2n = 2x = 14) to decaploid (2n = 10x = 70). These species are spread across the subtropical and temperate areas of the northern hemisphere [3–5]. *Rosa* is not only an economically important genus in ornamental horticulture but also widely known for its essential oil as well as its applications in perfumes, cosmetics, food and pharmaceuticals [6–9]. Fossil evidence indicates that rose species have existed for at least 30 million years [8, 10], with cultivation likely originated in China around 5000 years ago [6, 11, 12]. The taxonomic classification of *Rosa* species faces persistent ambiguities. The most widely adopted framework is the morphological system proposed by Rehder in 1940 [3], which was later revised by Wissemann (2003) [4] to integrate new phylogenetic data. This system recognizes four subgenera, namely *R.* subg. *Rosa* (formerly *Eurosa*), *R.* subg. *Hesperhodos* Cockerell, *R.* subg. *Hulthemia* (Dumort.) Focke, and *R.* subg. *Platyrhodon* (Hurst) Rehder. The latter three subgenera containing only one section and only one or two species. In contrast, the subgenus *Rosa* encompasses approximately 95% of the species. Within subgenus *Rosa*, species are further grouped into ten sections: *R.* sect. *Caninae* (DC.) Ser. (~ 50 species), *R.* sect. *Cinnamomeae* (DC.) Ser. (~ 80 species), *R.* sect. *Synstylae* DC. (~ 36 species), *R.* sect. *Pimpinellifoliae* (DC.) Ser. (~ 15 species), and four sections with fewer than five species, *R.* sect. *Gallicanae* (DC.) Ser. *R.* sect. *Carolinae* Crép., *R.* sect. *Chinenses* (DC.) Ser. = *R.* sect. *Indicae* Thory, *R.* sect. *Bracteatae* Thory, *R.* sect. *Laevigatae* Thory, *R.* sect. *Banksianae* Lindl. The most recent type of species in the genus is *R. cinnamomea* L. (syn. *R. majalis* Herrm.) in *R.* sect. *Cinnamomeae* [13, 14]; thus, *R.* sect. *Rosa* (DC.) Ser. is used rather than *R.* sect. *Cinnamomeae*. Some studies tend to use *R.* sect. *Microphyllae* Crep. to refer to *R.* subg. *Platyrhodon* within *R.* subg. *Rosa* [5, 10], but this paper retains *Platyrhodon* as a subgeneric status.

Previous studies have employed various molecular markers in attempts to resolve the genetic relationships among *Rosa* species. These include restriction fragment length polymorphism (RFLP) [15], random-amplified polymorphic DNA (RAPD) [16–23], amplified fragment length polymorphism (AFLP) [24–29], simple sequence repeat or microsatellite (SSR) [30–43], inter simple sequence repeat (ISSR) [44–47], and single-nucleotide polymorphism (SNP) [48–50] as well as DNA sequences such as those of chloroplast DNA (cpDNA; e.g., *matK*, *rbcL*, *psbA-trnH*), nuclear ribosomal DNA (nrDNA; e.g., ITS, *GAPDH*) or a combination of both [10, 51–74]. Furthermore, phylogenetic studies often integrate DNA sequences with molecular markers to increase resolution [75–80]. Despite this extensive research, molecular data have offered limited support for the current sections recognized by Rehder and later refined by Wissemann. In some cases, the results even contradict their frameworks. Furthermore, in some subgenera, sections or species, low bootstrap support (based on extremely low sequence divergence levels), poor internal resolution [57, 60] as well as incongruent phylogenies based on nuclear and plastid genes [10, 60, 63, 65, 72] still result in taxonomical and phylogenetic uncertainties. It is generally believed that small phenotypic variation, incomplete lineage sorting, extensive hybridization, recent radiation, low sequence divergence levels, ambiguous species definitions, chloroplast capture, and polyploidization complicate phylogenetic reconstruction of the genus *Rosa* [7–9, 16, 27, 42, 44, 46, 57–61]. Additional complications also arise from the lack of standardized species names as well as inadequate or biased sampling, especially when cultivars are involved or when sampling is limited to specific geographical areas or groups [10].

With the advent of next-generation sequencing technologies, increasing attention has turned toward the whole plastome, an important semiautonomous and uniparentally (usually maternally)-inherited organelle in most angiosperm species [81–85]. In angiosperms, typical cp genomes exhibit highly conserved quadripartite circular structures that consist of two single-copy regions (the small single-copy (SSC) and large single-copy (LSC) regions of sizes 16–27 kb and 80–90 kb, respectively) as well as a pair of inverted repeat (IR) regions (approximately 20–28 kb) [86]. These genomes generally ranges from 120 to 130 genes and 120–160 kb in length and have an overall GC content of 30–40% [87–89]. Compared to mitochondrial and nuclear genomes, plastomes are smaller, have reduced nucleotide substitution rates, exhibit lower recombination rates, and offer distinct genetic information for taxonomic and phylogenetic purposes at multiple taxonomic levels, especially for complex evolutionary relationships [90, 91]. Comparative and phylogenetic analyses of complete *Rosa* plastomes, have yielded improvement to a certain extent [92–98]. However, the phylogenetic relationships of *Rosa* remain unclear, including the delineation of its three subgenera except for *R.* subg. *Rosa*, the identity of the earliest diverging clade, and conflicts between nuclear and plastid data [128].

This study analyzed 18 representative *Rosa* plastid genomes, including 13 newly sequenced. The main objectives were: (1) to compare plastome structures within

the *Rosa* genus and provide genetic resources for future research; (2) to use larger datasets for reconstructing a more comprehensive phylogeny of the genus *Rosa*, with better-resolved phylogenetic relationships; and (3) to identify potential DNA markers within plastid genomes for phylogenetic analyses and classifying *Rosa* species. It is expected that the findings will provide valuable and basic sequence information for phylogenetic studies, species identification, molecular genetics studies and *Rosa* species breeding.

Results

General characteristics of *Rosa* plastomes

All 106 complete *Rosa* plastid genomes, for which sizes varied between 156,333 bp (*R. laevigata*) and 157,396 bp (*R. minutifolia*), presented the typical quadripartite structures (including the pair of IR regions, SSC, and LSC) of angiosperms (Fig. 1; Table S1), with the lengths of these four regions being as follows: LSC 85,452–86,539 bp; SSC 18,657–18,879 bp; IRA and IRB 26,008–26,082 bp (Table 1; Table S2). The total GC contents of these plastomes were from 37.15 to 37.30%. Moreover, the GC content was higher for the IR region (42.69–42.76%) in comparison with the SSC (30.94–31.37%) and LSC (35.09–35.31%) ones (Table 1; Table S2).

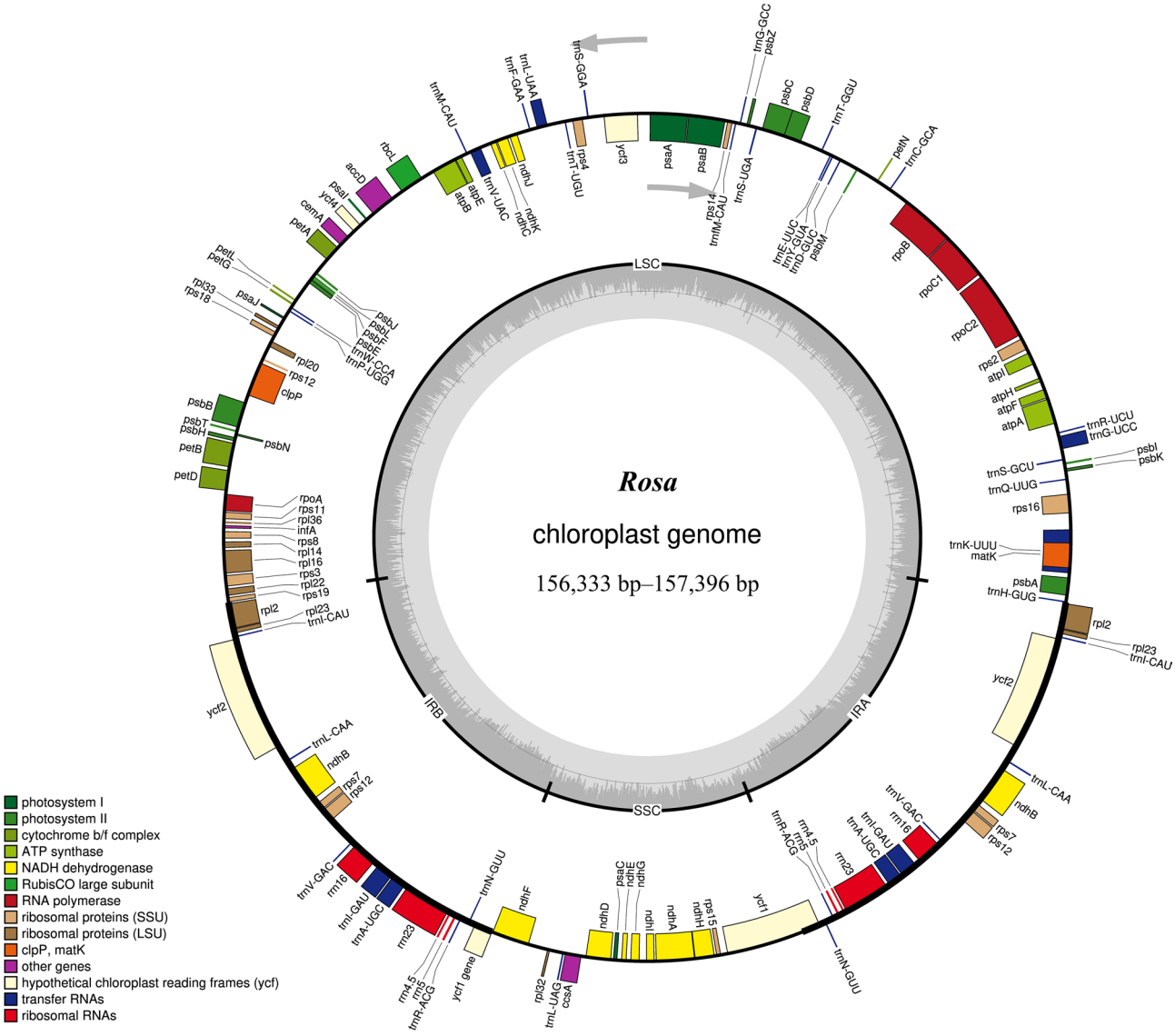

Fig. 1 Gene map of *Rosa* plastomes. Genes for which transcription occurs in a clockwise direction are shown on the inside of the circle, while those on the outside are transcribed in a counterclockwise direction. Genes assigned to different functional groups are represented by colored bars. For the inner circle, the lighter gray indicates the AT content of the genome, with the darker gray indicating GC content

Table 1 Characteristics of the 18 representative *Rosa* plastomes from different subgenera and sections

Species	GenBank accession	Length (bp)				Number of genes					GC content (%)			
		Total	LSC	SSC	IR	Total	PCG	tRNA	rRNA	ψycf1	Total	LSC	SSC	IR
R. moschata	PP976092	156,575	85,724	18,767	26,042	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.24	35.20	31.29	42.73
R. taiwanensis	PP976065	156,600	85,715	18,769	26,058	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.24	35.19	31.32	42.73
R. lucidissima	PP976062	156,583	85,735	18,760	26,044	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.24	35.21	31.32	42.71
R. odorata var. *pseudindica**	MK116518	156,652	85,785	18,761	26,053	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.22	35.17	31.28	42.72
R. × damascena	PP976090	156,575	85,724	18,767	26,042	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.24	35.20	31.29	42.73
R. canina	PP976078	156,589	85,590	18,861	26,069	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.24	35.23	31.23	42.71
R. laevigata	PP976063	156,377	85,495	18,784	26,049	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.29	35.30	31.25	42.74
R. cymosa	PP976091	156,567	85,719	18,782	26,033	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.21	35.19	31.18	42.73
R. bracteata	PP976079	156,783	85,893	18,792	26,049	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.17	35.11	31.28	42.70
*R. roxburghii***	KX768420	156,749	85,852	18,791	26,053	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.23	35.21	31.31	42.70
*R. rugosa**	MK641521	157,110	86,214	18,820	26,038	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.21	35.18	31.10	42.76
R. nitida	PP976095	156,998	86,095	18,799	26,052	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.23	35.23	31.10	42.75
R. palustris	PP976094	157,069	86,161	18,806	26,051	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.22	35.22	31.10	42.75
R. virginiana	PP976093	157,087	86,177	18,808	26,051	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.23	35.22	31.10	42.75
*R. minutifolia***	MT755634	157,396	86,539	18,755	26,041	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.21	35.17	31.19	42.76
*R. persica***	MZ261891	156,807	85,973	18,742	26,046	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.21	35.18	31.18	42.72
R. omeiensis	PP976098	157,151	86,293	18,694	26,082	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.22	35.20	31.14	42.73
R. xanthina f. hugonis	PP976080	157,214	86,270	18,810	26,067	131 (114)	85 (79)	37 (30)	8 (4)	1 (1)	37.18	35.18	31.03	42.72

IR: inverted repeat; LSC: large single-copy region; SSC: small single-copy region; GC: guanine-cytosine; rRNA: ribosomal RNA. tRNA: transfer RNA. * Sequences published by the authors of this article; ** Sequences downloaded from NCBI. * and ** in the following tables and figures indicate the same meanings as above. The other sequences with bold font indicated newly sequenced in this study. The unique genes in the *Rosa* plastomes are shown as numbers in parentheses

Overall, 113–114 unique genes were annotated in the 106 *Rosa* plastomes, including 29–30 transfer RNA (tRNA) genes, 4 ribosomal RNA (rRNA) genes, 1 pseudogene (ψ*ycf1*), and 79 protein-coding genes (PCGs) (Table 1; Table S2). Loss of the *trnD-GUC* gene (74 bp) resulting in its unique 29 tRNAs of *R. roxburghii* f. *normalis* KY419960. However, this gene was normal in *R. roxburghii* f. *normalis* MZ261869, *R. roxburghii* KX768420 and other *Rosa* plastomes. After careful alignment, the loss of *trnD-GUC* in *R. roxburghii* f. *normalis* KY419960 was due to a sequencing error. Furthermore, the distribution and order of genes in all the *Rosa* plastomes were also similar. These genes could be classified into four types, namely those for genetic systems, photosynthesis systems, biosynthesis, and those with unknown functions (Table S3). Although most genes occurred as single copies within the SSC or LSC, eighteen duplicated ones were identified in the IRs, and these included all four rRNA genes, seven tRNA genes and seven PCGs (Table S3). Seventeen genes presented introns: six tRNA genes and eight PCGs presented one intron, with two introns in the case of three PCGs (Table S4). Additionally, the exons of these intron-containing genes were the same length, but the intron lengths were mostly different. The *ycf1* gene at the IRb/SSC boundary was annotated as a pseudogene because incomplete duplication resulted in a premature stop codon.

Relative synonymous codon usage (RSCU)

Evaluating the codon usage pattern of 85 PCGs in the 18 *Rosa* germplasms revealed that the PCG sequences, with a total length of 78,750–78,822 bp, consisted of 26,250–26,274 codons. Furthermore, 64 types of codons (excluding stop codons) encoded 20 amino acids (Fig. 2; Table S5). The average RSCU value of each codon is shown in Fig. 2 because of the similar codon usage in the 18 *Rosa* plastomes. Among the 64 codons, cysteine (Cys, 1.16%) was the least encoded amino acid, while leucine (Leu, 10.43–10.54%) was the most commonly encoded one (Table S5). Moreover, except for methionine (Met) and tryptophan (Trp) (RSCU = 1), almost all amino acids had at least two synonymous codons. Leaving aside the three

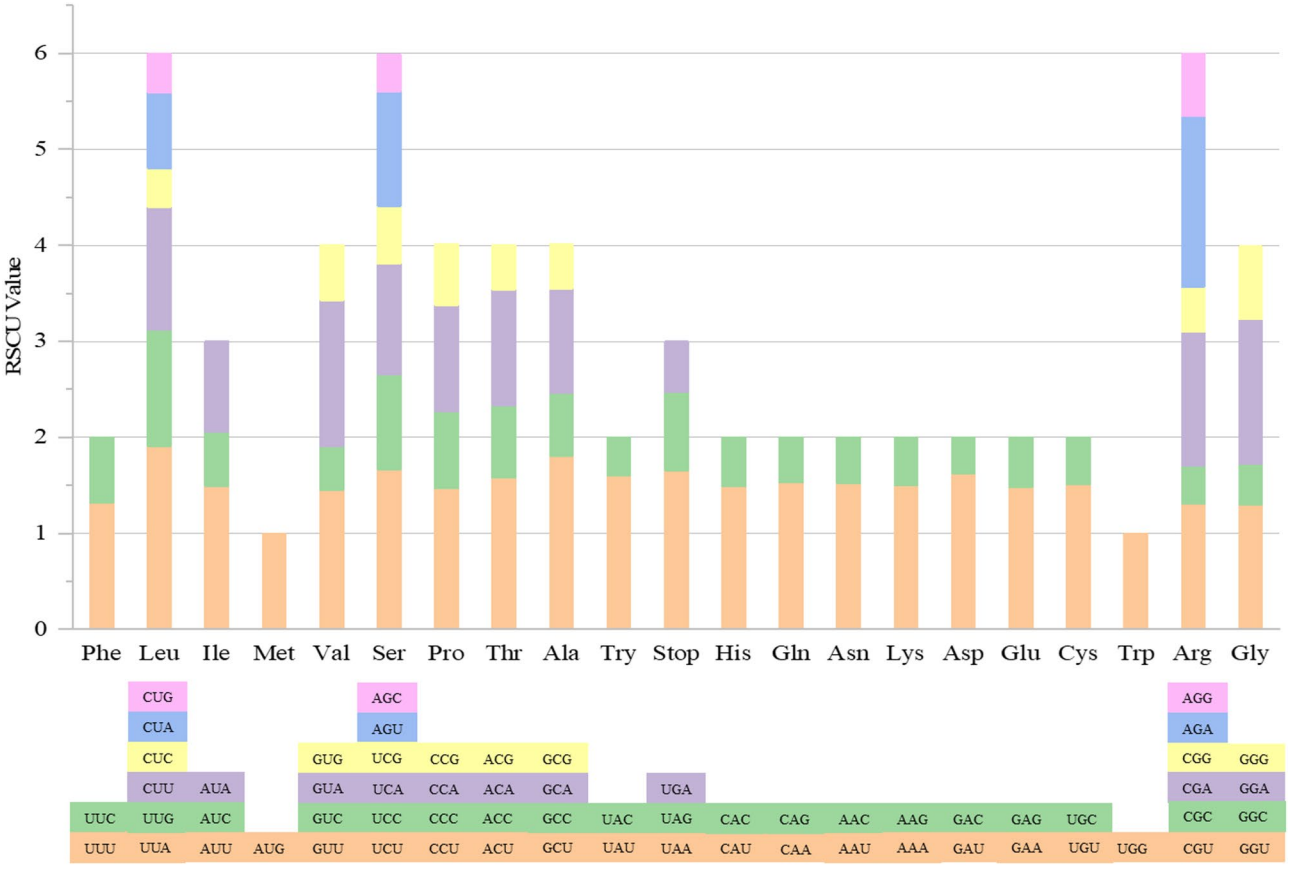

Fig. 2 Relative synonymous codon usage (RSCU) values of stop codons and 20 amino acids based on all coding sequences of the 18 representative *Rosa* plastomes. The codon usage of *Rosa* plastome sequences is indicated by a histogram above each amino acid, with the bars' colors corresponding to those of the codons. Ala (A): alanine; Tyr (Y): tyrosine; Pro (P): proline; His (H): histidine; Trp (A): tryptophan; Gly (G): glycine; Thr (T): threonine; Glu (E): glutamic; Ser (S): serine; Gln (Q): glutamine; Met (M): methionine; Cys (C): cysteine; Phe (F): phenylalanine; Asp (D): asparagine; Val (V): valine; Asn (N): asparagine; Lys (K): lysine; Leu (L): leucine; Ile (I): isoleucine; Arg (R): arginine

stop codons, all *Rosa* plastid genomic sequences had 30 codons RSCU > 1 except for the UCC codon RSCU > 1 of *R. lucidissima*, and most of the codons having RSCU of > 1 ended with A/U(T) (except UUG). In addition, the start codons of almost all PCGs of these *Rosa* plastomes were the standard ATG/CAT (RSCU = 1), except for *rps19* which started with CAC/GTG.

Repeat sequences

In the 18 *Rosa* plastomes, a range of 73 (*R. omeiensis*) to 87 (*R. lucidissima*) SSRs were detected, which consisted of 41–56 mononucleotide repeats, 10–15 dinucleotide repeats, 4–7 trinucleotide repeats, 9–13 tetranucleotide repeats, 0–2 pentanucleotide repeats as well as 0–3 hexanucleotide repeats (Fig. 3A; Table S6). Most SSRs were identified within noncoding regions (IGS and introns) (Fig. 3B). Moreover, most mononucleotides were A and T repeats, with A and T combinations also detected in the repeat units of other SSRs (Fig. 3C; Table S6). Additionally, we observed 36 (*R. nitida*)–52 (*R. bracteata*) long repeat elements, including 25–40 tandem (T), 4–7 forward (F) repeats, 3–6 palindromic (P), and 0–1 reverse (R) repeats (Fig. 3D; Table S7). Only one reverse repeat sequence was found in *R. laevigata*, *R. bracteata*, *R. minutifolia* and *R. persica*. The length of these long repeats was variable, ranging from 10 to 138 bp (*R. minutifolia*) (Fig. 3E; Table S7). Most long repeats were found within the IGS region; moreover, long repeated sequences contained only four coding regions (*rpoC2*, *infA*, *ndhF*, *ycf1* and *ycf2*).

Structural variation

Comparative analysis of the 18 representative *Rosa* plastomes revealed that, despite their highly conserved nature, the LSC/IR/SSC boundaries still exhibited some structural variations (Fig. 4). Specifically, the IR/SC boundaries were of two types: Type I, found in 12 plastomes, was the most common one, and it had JLB (junctions of LSC/IRb) within *rps19-rpl2*, JSA (junctions of SSC/IRa) within *ycf1*, JSB (junctions of IRb/SSC) within ψ*ycf1* as well as JLA (junctions of IRa/LSC) within *rpl2-trnH-GUG* (Fig. 4). The type II JLAs within *trnH-GUG* and the other three boundaries were similar; this type was present mainly in *R.* sect. *Pimpinellifoliae*, *R.* sect. *Carolinae* and *R.* subg. *Hulthemia*. Additionally, collinearity analyses based on 18 *Ros*a plastome alignments showed the absence of large-scale structural variations, such as rearrangements or inversions (Fig. S1).

Divergence of plastome sequences

An analysis of sequence divergence revealed very similar sequence identity plots (Fig. 5) and nucleotide polymorphism (Pi) values (Fig. 6). In particular, the protein-coding and IR regions were more conserved compared with the noncoding and SC regions, respectively. All protein-coding regions also had Pi values in the range of 0 to 0.00946 (*rps19*), with a mean of 0.00265, and thus were lower than 0.02. However, in the case of the noncoding regions, the Pi values were in the range of 0 to 0.02970, with a mean of 0.00764. Additionally, high Pi values (≥ 0.02) were noted for eight variable regions, six of which were within the IGS region (*trnG-GCC-trnfM-CAU*, *psbT-psbN*, *petD-rpoA*, *rps3-rpl22*, *ndhI-ndhA*, and *rpl2-trnH-GUG*) of the SC region, along with one in the IR region (*rps19-rpl2*) and another at the LSC/IRa boundary (*trnH-GUG-psbA*). The IGS region *trnG-GCC-trnfM-CAU* showed the highest Pi value (0.02970), indicating it was the most divergent. Comparing analysis with the complete 106-plastome dataset (Fig. S2) revealed minor differences in nucleotide diversity (pi values) across the eight variable regions.

Positive selection analysis

Likelihood ratio tests (LRTs) supported the presence of positively selected sites in the M8 (beta & ω > 1) model (Tables S8 and S9). Using the Bayes empirical Bayes (BEB) method, 56 significant positive selection sites, corresponding to 19 genes, were identified. These included four genes for the genetic system (*rps19*, *rpoA*, *rpoC1*, *rpoC2*), 11 genes for the photosynthetic system (*atpB*, *atpF*, *psbA*, *psbB*, *psbC*, and *rbcL*, *psaA*, *psaB*, *ndhF*, *ndhH*, *ndhI*), two genes for biosynthesis (*matK* and *accD*), and two unknown functional genes (*ycf1* and *ycf2*) (Table 2 and S8). Interestingly, the *ycf1* gene, found at the SSC/IRa junction, presented the greatest number (21) of positive selection sites, followed by *rbcL* (8), *ndhF* (5), and *rpoA* (4). Beside *matK*, *psaA* and *psbC*, which contained two significant positive selection sites, each of the other genes harboured a single significant positively selected site.

Phylogenetic analyses

The two datasets (79 plastid protein-coding sequences and complete plastome sequences) produced phylogenetic trees of similar topologies. In fact, analyzing the same dataset via the Bayesian inference (BI) and maximum likelihood (ML) methods yielded similar results (Figs. 7 and S3), that is a well-resolved phylogenetic tree of the *Rosa* genus where most nodes had bootstrap support of over 90%. All the results strongly supported the monophyly of the genus *Rosa* and that the subgenus *Rosa* was polyphyletic (PP = 1.00, BS = 100%). The 106 *Rosa* samples were grouped into seven major clades (A–G), and the positions of the other six clades were relatively stable, except for the D clade, which varied slightly in the phylogeny obtained from the different datasets. The conflicting nodes presented low node support and short branch lengths. The D clade based on the 79-CDS matrix

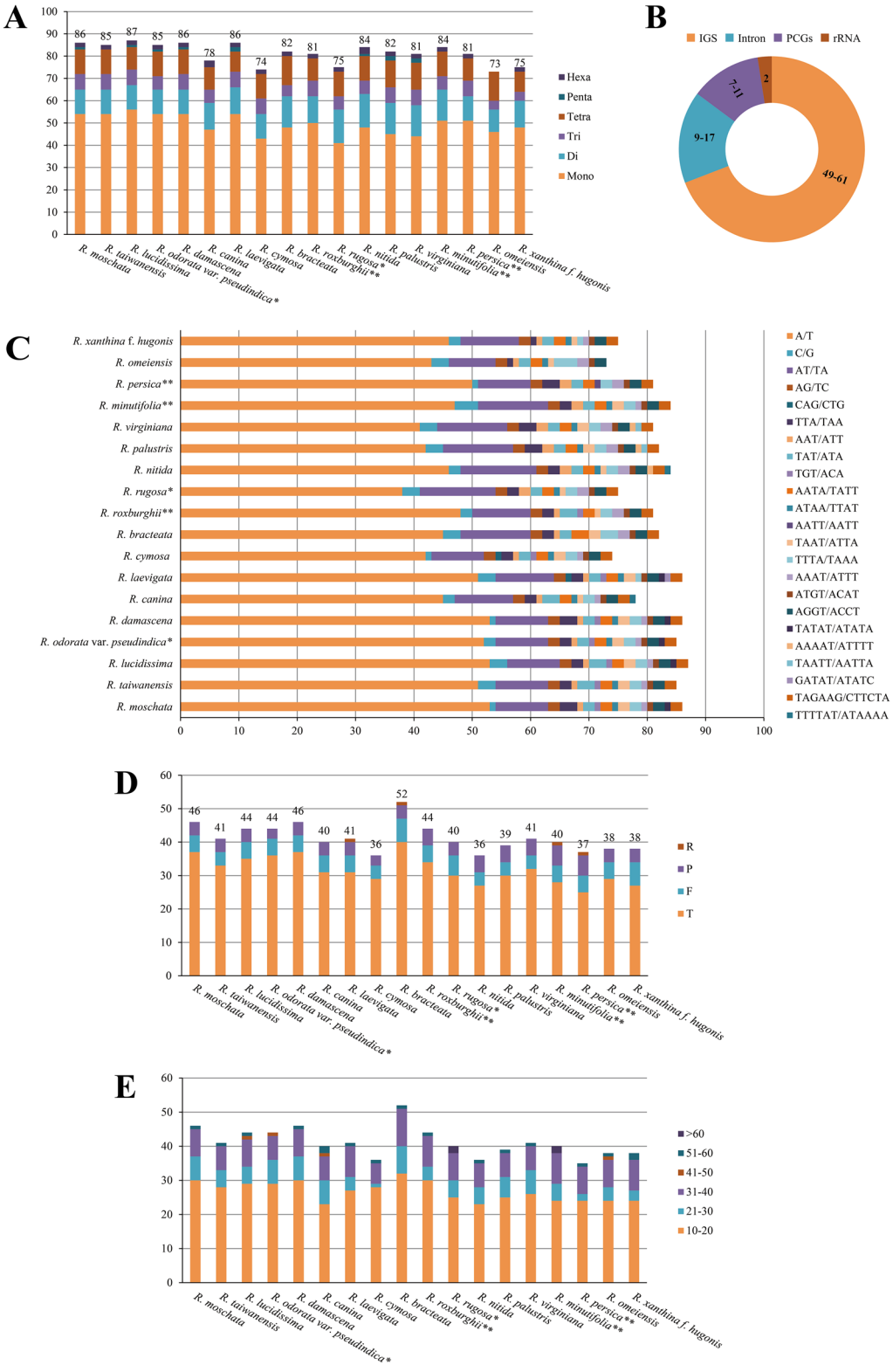

Fig. 3 Analysis of the repeat sequences of 18 representative *Rosa* plastomes. (**A**: Six types of SSRs and their frequency; **B**: number of SSRs distributed across different regions; **C**: Types and frequency of different SSR repeat units; **D**: Four types of long repeat sequences and their frequency; **E**: length of four types of long repeat sequences). * Sequences published by the authors of this article, ** sequences downloaded from NCBI

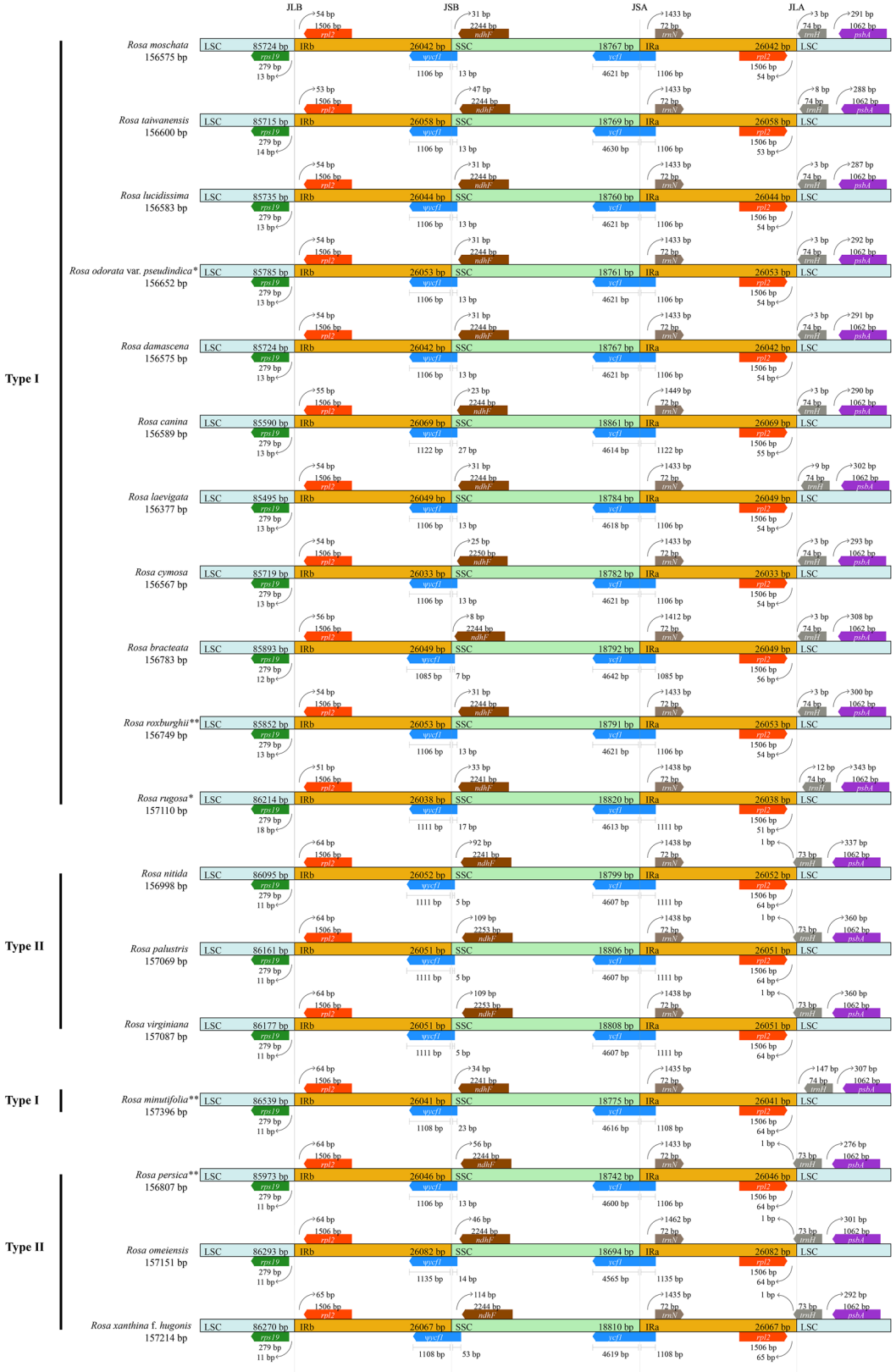

Fig. 4 Comparisons of the boundaries at the IR, SSC and LSC regions for the 18 representative *Rosa* plastomes. Colored boxes indicate genes, with numbers above the gene features representing the length between the gene's end and the borders. The location of the distance is marked by arrows. * Sequences published by the authors of this article, ** sequences downloaded from NCBI. This figure is not to scale

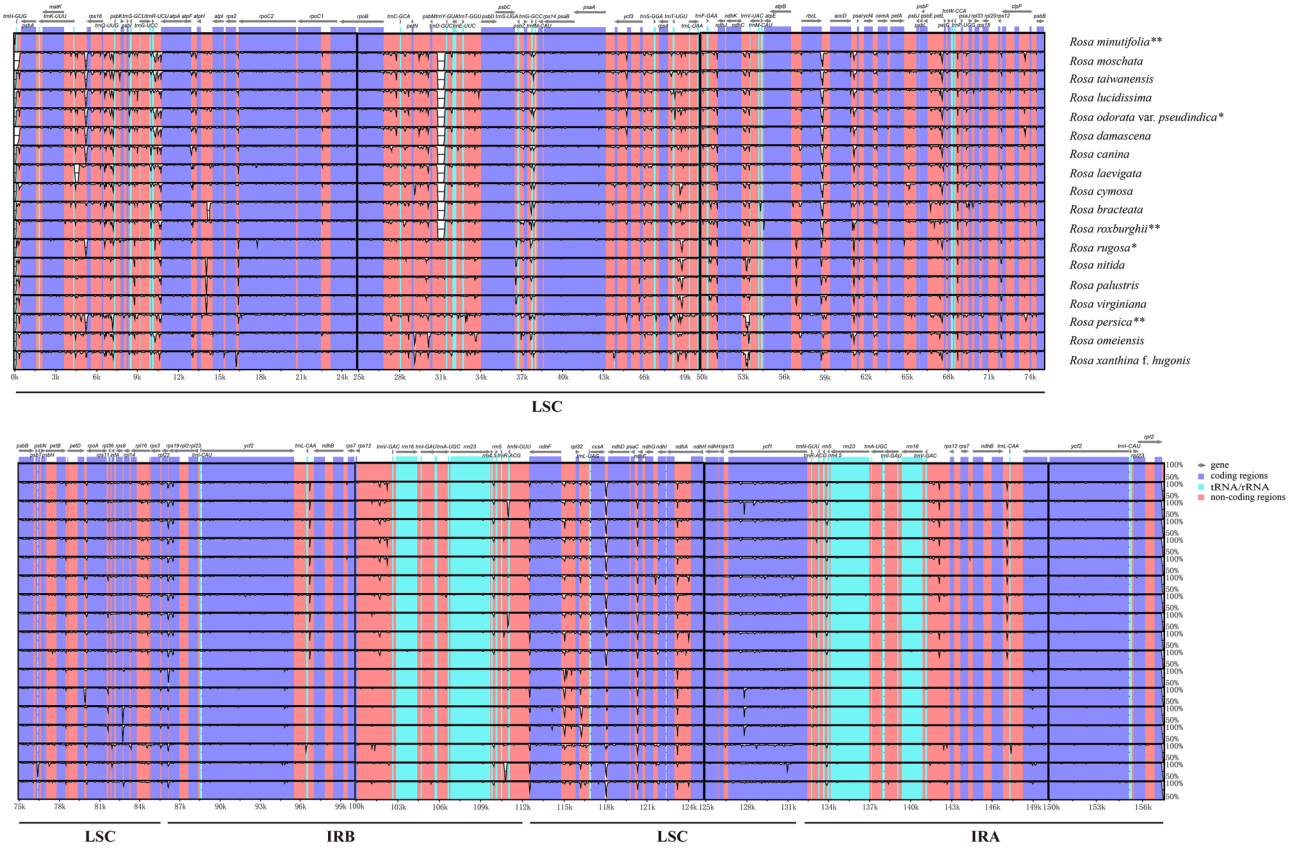

Fig. 5 Sequence identity plot of 18 representative *Rosa* plastomes. The X-axis shows the sequence length, while the Y-axis indicates percentage identity to the reference. The direction in which transcription occurs is shown by arrows below the genes. The different colored bars represent gene positions in the plastome (with *R. minutifolia* as a reference). * Sequences published by the authors of this article, ** sequences downloaded from NCBI

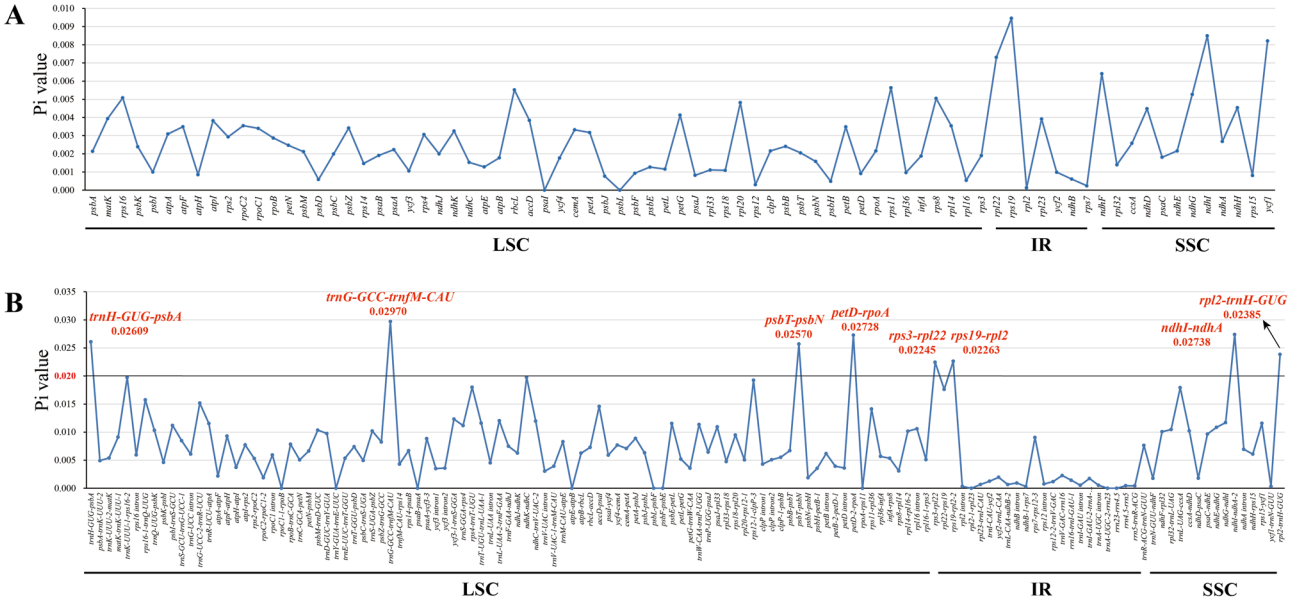

Fig. 6 Comparison of nucleotide polymorphisms (Pi) across 18 representative *Rosa* plastomes. (**A**) Protein-coding regions; (**B**) Noncoding regions. The X-axis indicates the position of genes in the *Rosa* plastomes, while the Y-axis depicts the Pi values. Gene names highlighted in red indicate those for which the Pi values were above 0.02

Table 2 Positively selected sites identified for the 18 representative *Rosa* plastomes

M8	Region	Gene Name	Selected Sites	Pr (ω > 1)	Number of Selected Sites
BEB	LSC	*accD*	98 K	0.973*	1
	LSC	*atpB*	1077 D	0.970*	1
	LSC	*atpF*	1727 W	0.977*	1
	LSC	*matK*	3006 I, 3018 L	0.971*, 0.997**	2
	SSC	*ndhF*	5230 S, 5500 M, 5607 N, 5609 T, 5666 W	0.976*, 0.973*, 0.976*, 0.977*, 1.000**	5
	SSC	*ndhH*	6204 M	0.972*	1
	SSC	*ndhI*	6492 L	0.991**	1
	LSC	*psaA*	7817 S, 7878 G	0.974*, 0.974*	2
	LSC	*psaB*	8571 G	0.978*	1
	LSC	*psbA*	9470 T	0.972*	1
	LSC	*psbB*	10,162 T	0.998**	1
	LSC	*psbC*	10,453 A, 10,603 S	0.998**, 0.974*	2
	LSC	*rbcL*	11,580 E, 11,777 I, 11,780 A, 11,807 V, 11,831 T, 11,880 A, 11,892 D, 12,027 L	0.975*, 1.000**, 0.978*, 0.973*, 0.978*, 0.973*, 0.974*, 1.000**	8
	LSC	*rpoA*	13,333 S, 13,388 I, 13,390 K, 13,391 H	0.973*, 1.000**, 1.000**, 1.000**	4
	LSC	*rpoC1*	15,027 Y	0.974*	1
	LSC	*rpoC2*	16,147 H	0.973*	1
	LSC	*rps19*	17,260 G	0.972*	1
	SSC	*ycf1*	18,565 F – 20,040 S (19656 I)	0.970* – 0.998**	21
	IR	*ycf2*	21,033 L	0.998**	1

*: $p < 0.05$; **: $p < 0.01$

was separated into three strongly-supported independent clades (D1–D3) (PP = 1.000, BS = 100%), which represented species of *R.* subg. *Platyrhodon* (main parts), *R.* sect. *Banksianae*, and *R.* sect. *Bracteatae*, respectively (Fig. 7). They were successive sisters to clades E + F + G, and all had weak nodal support in addition to the D3 clade. However, the D clade based on the whole plastome matrix, yielded a polytomy clade with poor node support (PP = 0.668, BS < 50%) and was strongly supported as a sister of clades E + F + G (PP = 1.000, BS = 100%) (Fig. S3). In addition, *R.* sect. *Laevigatae*, constituting the E clade, was a sister clade to F and G with weak node support (PP = 0.993 or 0.929, BS = 70% or < 50%). Along the phylogenetic backbone of the genus *Rosa*, *R.* sect. *Pimpinellifoliae* (main parts; A clade) was identified as the most basal clade. This was then followed by *R.* subg. *Hulthemia*, which composed the B clade. The C clade included all species of *R.* subg. *Hesperhodos*, *R.* sect. *Carolinae* and *R.* sect. *Rosa*, as well as a few species of *R.* subg. *Platyrhodon* (*R. praelucens*), *R.* sect. *Pimpinellifoliae* (*R. spinosissima* and *R. kokanica*) and *R.* sect. *Synstylae* (*R. glomerata*). *R.* subg. *Platyrhodon* and *R.* sect. *Rosa* were both nonmonophyletic. The F and G clades were the most recently differentiated, where the F clade consisted of species from the *Caninae* and *Gallicanae* sections (main parts) as well as *R. arvensis* from *R.* sect. *Synstylae*. *R.* sect. *Caninae* was also nonmonophyletic. Furthermore, *R.* sect. *Chinenses* and *R.* sect. *Synstylae* were likewise not monophyletic but polyphyletic and formed the G clade with *R. sterilis* and *R. kweichowensis* from *R.* subg. *Platyrhodon* and *R.* × *damascena* from *R.* sect. *Gallicanae*. Interestingly, different individuals or variants of a species did not cluster together in the phylogenetic tree but with other species. For example, *R. persica*, *R. canina*, *R. lucidissima*, and *R. chinensis*.

Discussion

Plastome structure comparisons and sequence divergence hotspots

Rosa species exhibit quite conserved plastomes, with similar overall genome structures, gene numbers, gene components, gene orders and even total length (ranging from 156 kb to 157 kb) (Table 2 and S2). Consistent with previous analyses on *Rosa* plastomes [92–98], the GC content (37.15–37.30%) and codon usage of the *Rosa* plastomes remained highly conserved (Tables S2 and S5). However, the GC content was higher for the IR region compared to the SC regions, and the majority of codons (> 85%) having RSCU > 1 ended in A/U(T). These results were favourable for the stability of the plastid genome [99]. Plastomes at the long length are mainly concentrated in the basal taxa of the *Rosa* genus: the *Rosa*, *Carolinae* and *Pimpinellifoliae* sections of the subgenus *Rosa* as well as the subgenera of *Hulthemia* and *Hesperhodos* (Fig. 7; Table S2). Previous studies have reported that differences in plastome length are caused mainly by variations in the

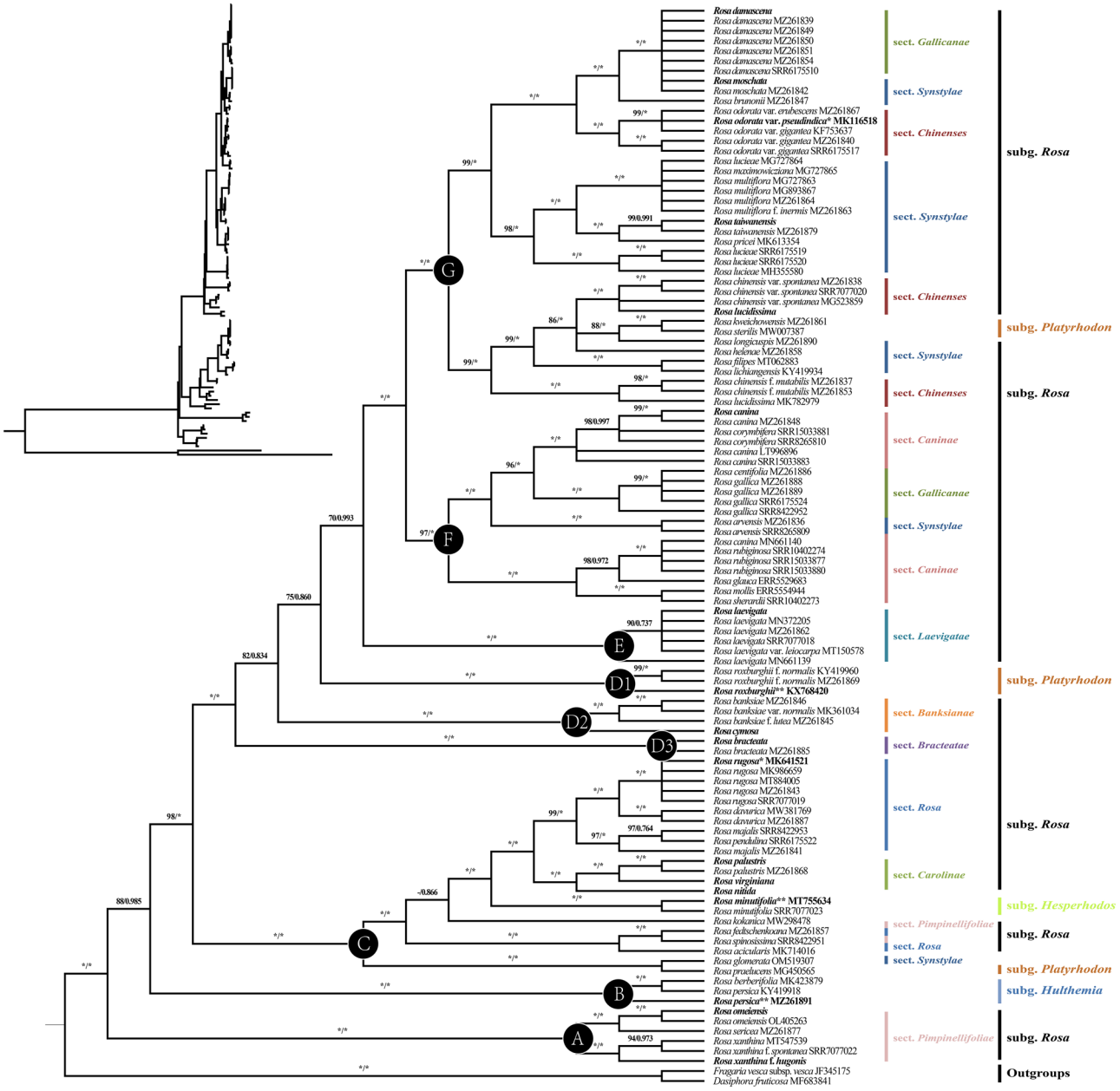

Fig. 7 Consensus phylogenetic tree obtained from 79 protein-coding sequences (CDSs) of 106 *Rosa* samples and two outgroups using the Bayesian inference (BI) and maximum likelihood (ML) methods. ML bootstrap support (BS)/BI posterior probabilities (PP) are indicated by numbers above the branches. "*" indicates PP = 1.00 or BS = 100%. "-" indicates PP < 0.50 or BS < 50%. Distance between species are indicated by branch lengths in the upper left figure

IR regions, variations in noncoding regions (introns and intergenic regions), and gene loss [90, 100]. In the genus *Rosa*, variation in the noncoding regions as well as slight expansions of the IR regions could largely explain the variations in plastome length.

Repeated sequences are crucial for genome rearrangements and variations [101, 102], and they are mostly found in IGSs rather than coding sequences [103]. In this study, abundant SSRs (73–87) and long repeats (36–52; including tandem repeats and interspersed repeats) were detected across 18 representative *Rosa* species (Fig. 3C; Table S6). Most SSRs, tandem repeats (T) and interspersed repeats (FPRs), are distributed in noncoding regions (IGS and intron regions), with most SSRs being mononucleotide repeats (A or T) and most long repeats being tandem and forward repeats, respectively. These findings align with the results of other plastome studies, although the number of repetitive sequences has varied across studies [94, 97]. Such repeat sequences can provide valuable information for developing genetic markers that can be applied in phylogenetic and population studies [104]. In *Rosa* plastomes, we also detected several

Table 3 Divergent hotspot statistics of *Rosa*

Sampling	Pi threshold	Divergent regions	References
Five *Rosa* samples	> 0.006	*psbI-trnS-GCU-trnG-UCC*, 5′*matK-trnK-UUU*, *rps16-trnQ-UUG*, *rpoB-trnC-GCA*, *rps4-trnT-UGU*, *ycf1*	Jeon and Kim, 2019 [92]
Five *Rosa* samples	Uncertain	*psbL-trnS-GCU*, *psbM-trnD-GUC*, *trnM-CAU-aptE*, *trnG-UCC-trnfM-CAU*, *psbZ-trnG-UCC*, *trnR-UCU-atpA*, *trnH-GUG-psbA*, *trnT-UGU-trnL-UAA*, *psbE-petL*, *trnP-UGG-psaJ*, *trnK-UUU-rps16*, *rps16-trnQ-UUG*, *psaJ-rpl33*, *trnS-GGA-rps4*, *rps2-rpoC2*, *aptF-aptH*, *rpl12-clpP*, *rpoB-trnC-GCA*, *trnD-GUC-trnY-GUA*, *ndhA*	Li et al., 2020 [94]
13 *Rosa* samples	> 0.007	*trnK-UUU*, *rps16-trnQ-UUG*, *ycf1*, *trnT-UGU-trnL-UAA*	Shen et al., 2022 [97]
28 *Rosa* and 2 *Geum* samples	> 0.010	*ycf1*, *rps16-trnQ-UUG*, *psbE-petL*, *trnT-UGU-trnL-UAA*	
12 *Rosa* samples (covered all subgenera and sections)	> 0.021	*trnL-UAA-trnF-GAA*, *trnT-UGU-trnL-UAA*, *ndhC-trnV-UAC*, *psbZ-trnG-GCC*, *psbE-petL*, *ycf1*	Zhang et al., 2022 [96]
24 *Rosa* samples	> 0.013	*ycf3-trnS*, *trnT-trnL*, *psbE-petL*, *ycf1*	Gao et al., 2023 [98]
18 *Rosa* samples (covered all subgenera and sections)	≥ 0.020	*trnH-GUG-psbA*, *trnG-GCC-trnfM-CAU*, *psbT-psbN*, *petD-rpoA*, *rps3-rpl22*, *ndhI-ndhA*, *rpl2-trnH-GUG*, *rps19-rpl2*	The current study

specific SSRs. For instanceSSR (TTTTAT), which was exclusively found in the *R. canina* plastome; SSR (ATA-AAA), which was detected only in the *R. nitida* plastome; and SSR (TAGAAG/CTTCTA), which was only absent in the *R. omeiensis* plastome. The SSR (AAAAT) was present only in *R.* sect. *Carolinae*. These special SSRs can potentially act as molecular markers that help in identifying *Rosa* species.

IR regions are crucial for maintaining the plastid genome's structural stability, and their length variations are generally recognized as the most significant factor contributing to variations in plastome size [87, 105]. However, no significant variation in the IR region was detected in *Rosa* plastomes- an observation consistent with several previous studies [97, 98]. As in most dicots, the *rps19*, *rpl2*, *ndhF*, *ycf1*, *trnN-GUU* and *trnH-GUG* genes in *Rosa* plastomes are consistently located at the boundaries of the SC and IR [91]. We classified the boundaries of *Rosa* plastomes into two types according to whether the *trnH-GUG* gene spanned the IRa/LSC junction, with type I (the *trnH-GUG* gene was 3–147 bp away from JLA, mostly 3 bp) being the most prevalent, which further revealed that the boundary structures of *Rosa* plastomes are relatively stable.

In this study, mVISTA analysis revealed that greater variations in the SSC and LSC regions compared with the IR ones. Furthermore, the protein-coding regions had fewer variable sites in comparison with the noncoding ones (Fig. 6). This was consistent with the prevalent diversity patterns observed in many angiosperms [106, 107]. Eight highly variable regions (Pi ≥ 0.020) were initially identified as divergence hotspots based on 18 representative *Rosa* plastomes, including six in the IGS regions of the SC regions (*trnG-GCC-trnfM-CAU*, *psbT-psbN*, *petD-rpoA*, *rps3-rpl22*, *ndhI-ndhA*, and *rpl2-trnH-GUG*), one in the IGS region of the IR region (*rps19-rpl2*), and one at the LSC/IRa junction (*trnH-GUG-psbA*). Notably, *trnH-GUG-psbA* has frequently been used as a molecular marker for studying the phylogeny of the genus *Rosa* in previous studies [2, 10, 71, 74]. To evaluate their robustness across broader taxonomic sampling, we expanded the analysis to 106 *Rosa* plastomes. This validation revealed three consistently variable regions: *trnG-GCC-trnfM-CAU*, *psbT-psbN*, and *rpl22-rps19*. Among them, two (*trnG-GCC-trnfM-CAU* and *psbT-psbN*) overlapped with the initially identified hotspots, whereas the remaining highly variable regions from the 18-species subset showed slightly lower pi values and were not among the top-ranked in the full dataset (Fig. S2). Among these, two highly variable regions (*trnG-GCC-trnfM-CAU* and *psbT-psbN*) overlapped with the initially identified hotspots, while the remaining six highly variable regions from the 18-species subset showed slightly lower pi values in the full 106-sample dataset, although the differences were minor. This discrepancy is likely attributed to differences in sampling strategy. The 18 representative *Rosa* plastomes were selected to maximize phylogenetic diversity, thereby capturing lineage-specific polymorphisms that may not persist in broader sampling. In contrast, the full dataset of 106 *Rosa* plastomes reflects a more inclusive species spectrum, which can dilute rare variants and emphasize only the most conserved hotspots across the genus. These observations underscore that hotspot identification is sensitive to dataset composition, alignment strategy, and threshold selection. Previous *Rosa* plastome studies have similarly reported different sets of variable regions depending on sampling schemes, different threshold selection and sequence alignment approaches (Table 3), highlighting the influence of study-specific parameters. In addition, variation in sequencing strategies - such as the use of whole genome shotgun sequencing- may also contribute to differences in polymorphism detection across studies [92, 94, 96–98]. Although the seven regions identified from the 18 representative *Rosa* plastomes offer promising candidates for DNA barcoding, only two remained

consistently variable across the full dataset of 106 *Rosa* plastomes. This discrepancy emphasizes the need for further phylogenetic validation before adopting these markers universally. A comprehensive approach integrating large-scale sampling and multi-locus comparisons will be essential for resolving relationships among closely related *Rosa* species and informing breeding and conservation strategies.

Adaptive evolution

Positive selection plays a crucial role in driving molecular adaptation and functional divergence of plastid genes under environmental pressures [108]. In this study, 19 chloroplast genes exhibiting signals of positive selection were identified. These genes span genetic system (*rps19*, *rpoA*, *rpoC1*, *rpoC2*), photosynthesis (*atpB*, *atpF*, *psbA*, *psbB*, *psbC*, *rbcL*, *psaA*, *psaB*, *ndhF*, *ndhH*, *ndhI*), biosynthesis (*matK*, *accD*), and open reading frames of unknown function (*ycf1*, *ycf2*) (Table 2 and S8). Although the precise functions of *ycf1* and *ycf2* remain unclear, their high variability made them used in phylogenetic studies [109]. Notably, *ycf1* harbors the highest number of positively selected sites identified in this study, and it also exhibits significant signatures of adaptive evolution across various plant lineages [81, 110]. The *matK* and *accD* genes are associated with protein biosynthesis. *matK* encodes a maturase essential for the splicing of multiple chloroplast introns, playing a pivotal role in maintaining chloroplast function, and exhibits adaptive evolution signals in several lineages [111, 112]. Additionally, *matK* is frequently employed as an effective molecular marker in phylogenetic studies of the genus *Rosa* [57, 65, 77]. The *accD* gene, which is involved in fatty acid synthesis, contributes to the stability of the chloroplast membrane and may enhance plant responses to environmental stresses such as temperature and drought [113]. Photosynthesis-related genes such as *atpB/F*, *psaA/B*, *psbA/B/C*, *rbcL*, and *ndhF/H/I* are under positive selection pressure in this study. The *atpB* and *atpF* encode subunits of the ATP synthase complex, integral to energy conversion [114]. The *psaA/B* and *psbA/B/C* are core components of Photosystem I and II, respectively and arecrucial for maintaining photosynthetic efficiency [81, 87, 115]. Their adaptive variations may be linked to ecological adaptations of *Rosa* species under varying light intensities and altitudinal gradients. The *ndh* gene cluster encodes subunits of the NADH dehydrogenase complex, participating in photosynthetic electron transport and energy regulation under photo-oxidative stress [116, 117]. Specifically, *ndhF* has been shown to be under positive selection in multiple studies on plant adaptive evolution [118, 119]. The *rbcL* gene encodes the large subunit of Rubisco, central to carbon assimilation, and is widely reported as a target of adaptive evolution in photosynthetic systems, particularly under drought or low-temperature conditions [81, 120, 121]. Furthermore, the *rps* gene family plays significant roles in cell biology, including participation in protein synthesis, maintenance of cell growth, regulation of the cell cycle, and involvement in cell signal transduction [122]. The *rpoA*, *rpoC1*, and *rpoC2* genes encode chloroplast RNA polymerase subunits, forming core components in the regulation of chloroplast gene expression. These genes frequently appear in lists of positively selected genes in angiosperms, while their variations likely supporting dynamic regulation of gene expression in response to environmental changes [123, 124]. *Rosa* species are widely distributed across temperate to subtropical regions of the Northern Hemisphere, with some extending into the frigid zone, inhabiting diverse ecological types including low-altitude plains, alpine meadows, and forest edges [3–5]. Facing variable ecological factors such as light, moisture, and temperature, the positive selection signals observed in the aforementioned genes may represent molecular manifestations of their adaptive evolution to complex environments. Particularly, genes related to photosynthesis and transcription/translation may play key adaptive roles under different ecological conditions (e.g., drought, low temperature, high radiation). Moreover, several of the positively selected genes identified in our analysis—such as *accD*, *matK*, *ndhF*, *rbcL*, and *ycf1*—have also been frequently reported to be under positive selection in other angiosperms (including *Rosa*) [96, 119, 125–126], indicating that they may represent conserved targets of adaptive evolution across flowering plants. These results underscore the importance of integrating these key molecular markers in studies of phylogeny, germplasm conservation, and adaptive evolution.

Phylogenetic relationships

Despite its long-standing ornamental and economic significance, the genus *Rosa* remains particularly challenging for taxonomic classification and phylogenetic reconstruction [4, 10, 15, 25, 27, 57–80, 92–98]. Traditional classification systems, proposed by Rehder (1940) and Wissemann (2017) [3, 127], recognize four *Rosa* subgenera, with *R.* subg. *Rosa* further classified into 10 sections. Although support for deeper phylogenetic nodes remains limited, our study reconstructs a robust phylogeny of *Rosa* based on complete chloroplast genomes. The plastome phylogenetic framework revealed that most previously-defined subgenera and sections were paraphyletic in origin. These aligned with the results reported previously [95, 96, 128] and could be attributed to chloroplast capture, introgressive hybridization, differences between the evolutionary rate of nuclear and plastid genes, or incomplete lineage sorting. In our plastid sequence analyses, the three previous subgenera

(*Platyrhodon*, *Hulthemia* and *Hesperhodos*) seemed to best sink into *R.* subg. *Rosa* as sections, which did not occur as a sister to *R.* subg. *Rosa* [4, 95, 128]. Based on its morphological similarity and molecular evidence, *R.* sect. *Carolinae* has been resolved as paraphyletic with the other species in *R.* sect. *Rosa* [15, 25, 55, 57, 59, 60, 62, 68, 128, 129] and is now part of *R.* sect. *Rosa*, which was confirmed by our analysis. The species of *R.* sect. *Chinenses* nested within *R.* sect. *Synstylae* as previously described [54, 60, 62, 72, 75, 96, 128, 130, 131]. Interestingly, *R. glomerata* from *R.* sect. *Synstylae* in our study is the only member emerging in the C Clade, in agreement with early analyses of single-copy nuclear genes [98, 128], revealing its putative hybrid origin. Hence, the merging of *R.* sect. *Chinenses* with *R.* sect. *Synstylae* is being proposed. The *Laevigatae*, *Banksianae* and *Bracteatae* sections are all monophyletic in our study [10, 96], located at the base of the *Chinenses-Synstylae* Clade with *R.* subg. *Platyrhodon* with moderate support except for the polyploid sections (*Gallicanae* and *Caninae*). Furthermore, *R.* sect. *Pimpinellifoliae* from *R.* subg. *Rosa* is polyphyletic, in line with previous observations through chloroplast or nuclear ITS sequences [54, 55, 60, 62].

However, certain inconsistencies were observed between our findings and those inferred from nuclear gene data [128, 131]. Our core *Pimpinellifoliae* species (A Clade in Fig. 7) diverged the earliest in the genus *Rosa*, with *R.* subg. *Hulthemia* as sister to the remaining *Rosa* species (excluding *R.* subg. *Hesperhodos*). This aligns with previous plastome-based phylogenies in roses [95], including the same sequence of *R. minutifolia* (*R.* subg. *Hesperhodos*) from Zhao et al. [132]. Through analysis of nuclear allele SCO_{Tag} sequences using amplicon sequencing, Debray et al. [128] proposed that *R.* subg. *Hesperhodos* represents the earliest-diverging lineage in *Rosa*, with *R.* subg. *Hulthemia* and core sect. *Pimpinellifoliae* forming a monophyletic clade - a finding consistent with earlier plastome-based phylogenetic analyses [96]. In this study, both sequences of *R. minutifolia* obtained from the NCBI GenBank were reportedly collected from France, suggesting they may represent cultivated rather than native Californian populations [62]. The placement of *R.* subg. *Hesperhodos* outside the most basal group may reflect taxonomic misclassification rather than its true evolutionary position. Furthermore, phylogenetic analysis based on SNPs from haplotype-resolved genome assemblies and resequencing data [131] revealed strong support for a clade comprising sections *Laevigatae*, *Banksianae* and *Bracteatae* positioned basally to the other *Rosa* accessions, without representatives of *R.* subg. *Hulthemia* and *R.* subg. *Hesperhodos* in the sampling. All three sampled *Pimpinellifoliae* species formed a well-supported clade, which was the sisters to sections *Rosa-Carolinae* and the other polyploid *Pimpinellifoliae* species [131]. *R.* sect. *Banksianae* in our study is monophyletic [10, 96, 131], unlike the results reported previously [128], regarding *R. cymosa* as a hybrid from other sections. The observed phylogenetic discrepancies may stem from our reliance solely on chloroplast genome data, as their uniparental inheritance limits detection of reticulate evolutionary processes involving polyploidization or hybridization [10, 58, 62, 96, 128].

Polyploidy is well-documented as a significant phenomenon in evolution and a crucial cytogenetic mechanism in the process of speciation [133], contributing to the complexity and intrigue of classifying the genus *Rosa*. Our analysis of the cp genome revealed that the polyphyletic nature of several sections, such as *R.* sect. *Gallicanae* and *R.* sect. *Caninae*, can be attributed to certain polyploid accessions, as these sections consist solely of polyploid species. All of the accessions of *R.* × *damascena* in *R.* sect. *Gallicanae* in our plastid phylogenetic tree, clustered with *R. moschata* and *R. brunonii* (*R.* sect. *Synstylae*), which has been proven to constitute the maternal lineage of *R.* × *damascena* [128]. Furthermore, *R.* sect. *Caninae*, consisting of only allopolyploid species of hybrid origin, clustered with some members of the *Synstylae* and *Gallicanae* sections, in accordance with previous results [128]. Two members of *R.* sect. *Gallicanae*, *R. gallica* and *R. centifolia*, are sisters to parts of *R.* sect. *Caninae*, which might indicate their origination by hybridization with *R.* sect. *Caninae* as their maternal progenitor [96, 128, 131]. *Rosa kokanica* and *R. spinosissima* from *R.* sect. *Pimpinellifoliae* are also allopolyploid, emerging in C clade, including members of sections *Rosa*, *Carolinae* and *Synstylae*, which indicates a close relationship among them. Debray et al. [128] confirmed the hybrid origin of the *R. kokanica* and *R. spinosissima* likely derived from crosses with species between *R.* sect. *Rosa* (as seed parents) and core *Pimpinellifoliae* species (as pollen parents). In addition, *R. praelucens* (*R.* subg. *Platyrhodon*, 10x), the highest naturally occurring ploidy in the genus *Rosa*, is a sister to *R. glomerata* (*R.* sect. *Synstylae*), clustering with other species from several sections (*Rosa*, *Pimpinellifoliae* and *Carolinae*) far from *R. roxburghii* in the same subgenera in our study. Prior cytological study has identified *R. praelucens* as an allopolyploid [2], closely related to *R.* sect. *Rosa* according to combined cpDNA and nrDNA data [10, 65] and possibly derived from *R. sweginzowii* in Asia [128]. Therefore, we propose that *R. roxburghii* should be retained as the sole representative of the newly established section *Platyrhodon* [10, 128].

Conclusions

This study assembled the most comprehensive plastome dataset to study plastid genomic features across the *Rosa* genus before reconstructing a plastome phylogeny with a

well-resolved backbone. Comparative analysis of plastid genomic analyses revealed that gene content, gene order, collinear structure and codon usage were conserved within the *Rosa* plastomes. Additionally, eight highly diverse hotspots - *trnH-GUG-psbA*, *trnG-GCC-trnfM-CAU*, *psbT-psbN*, *petD-rpoA*, *rps3-rpl22*, *ndhI-ndhA*, *rps19-rpl2*, and *rpl2-trnH-GUG* - were initially identified from 18 representative *Rosa* plastomes as candidate regions for shallower DNA barcoding and phylogenetic analyses in *Rosa*. However, broader sampling is required to validate their reliability and phylogenetic utility. Selection pressure analyses revealed 19 positively selected genes related to genetic, photosynthetic, and biosynthetic functions, indicating adaptive evolution in *Rosa* plastomes. Moreover, a plastome phylogeny that resolved most of the intersection relationships was successfully established. The phylogenetic backbone is stable except for that of *R.* subg. *Platyrhodon*, sections *Banksianae* and *Bracteatae*, which are not well resolved. The status and position of *R.* sect. *Pimpinellifoliae* and the nonmonophyletic *R.* subg. *Rosa*, *Platyrhodon*, *R.* sect. *Synstylae*, and *Chinenses*, as well as the conflicting relationships of those with the nuclear phylogeny, suggest that this study about the phylogeny of *Rosa* can be further improved in terms of intrageneric relationships. In addition, the current study also shows the limitations of using plastomes to infer the phylogeny of some difficult taxa, such as *R. praelucens*, *R. glomerata*, *R. kokanica*, *R. acicularis*, *R. spinosissima*, *R. fedtschenkoana*, *R. arvensis*, *R. sterilis* and *R. kweichowensis*, which may be involved in hybridization and polyploidization. Future *Rosa* studies can be based on a combination of plastome, morphological and nuclear data.

Materials and methods

Taxon sampling and DNA extraction

Overall, 18 *Rosa* species were selected for comparative genomic analysis in this study, including 13 sequences newly sequenced, and the other five retrieved from the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/nuccore/) [134, 135]. The 18 materials sampled here rep resented all four subgenera (*Rosa*, *Hulthemia*, *Hesperhodos* and *Platyrhodon*) of the *Rosa* genus as well as all ten sections in the *Rosa* subgenus according to Reder's system (1940) [3]. This sampling strategy prioritized newly sequenced plastomes for data consistency, while ensuring representation of all major taxonomic lineages in *Rosa*. Additionally, 64 *Rosa* plastomes (excluding cultivars) were obtained from the NCBI GenBank database, while raw sequences from the NCBI-SRA database (https://www.ncbi.nlm.nih.gov/sra) were also reassembled into 24 *Rosa* plastomes for phylogenetic analysis. Two plastomes of Tribe Potentilleae (Rosaceae), a close relative to *Rosa* [92, 136–138], were chosen as outgroups, i.e., *Fragaria vesca* subsp. *vesca* (JF345175) and *Potentilla micrantha* (HG931056). Consequently, 108 complete plastomes (106 *Rosa* and two outgroups) were obtained for phylogenetic analysis, ensuring that each subgenus or section contained at least two samples. The GenBank accessions, including detailed sample information, are presented in Table S1. In this study, the 13 newly sequenced species were those of fresh or silica-dried leaves which were obtained following DNA extraction using a modified cetyltrimethylammonium bromide (CTAB) method [139]. The DNA's integrity was then checked by electrophoresis on a 1% (w/v) agarose gel, and after assessing its purity with a NanoDrop spectrophotometer 2000, the concentration was determined via a Qubit 2.0 instrument for precise quantification.

Plastome sequencing, assembly and annotation

Purified genomic DNA of high-quality was sheared into short 350-bp fragments to construct paired-end (PE) libraries. This was followed by sequencing on an Illumina HiSeq 2500 platform at Novogene Company (Tianjin, China) to yield 150-bp paired-end reads. Each sample generated around 4 Gb of clean data which were imported into GetOrganelle v1.7.5.3 with the parameters suggested by the developers of the software (https://github.com/Kinggerm/GetOrganelle) for assembly into the plastome sequences. This was followed by automatic annotation of the finished plastomes using the Plastid Genome Annotator (PGA) [140] before manually adjusting the intron/exon boundaries as well as start/stop codons in Geneious v8.0.2 based on multiple *Rosa* complete plastome sequences. All SRA data and plastome sequences obtained from the NCBI databases were also subjected to reassembly and reannotation as it was the case for the newly generated sequences. After drawing the structural features of the *Rosa* plastome map online via Organelle Genome DRAW (OGDRAW) (http://ogdraw.mpimp-golm.mpg.de/) [141], the newly sequenced plastid genomes and the plastomes extracted from the NCBI-SRA data were deposited into NCBI GenBank under the accession numbers provided in Table 1 and S2.

Genome features and comparative plastid genomic analysis

The general characteristics (including number of genes, GC content, length and gene categories) of the complete *Rosa* plastomes were analyzed in Geneious v8.0.2 (Table 1 and S2). Table 1 shows the basic characteristics of the 18 complete *Rosa* plastomes only, and Table S2 presents the basic characteristics of the whole plastomes of the 108 samples. For an in-depth analysis of plastome features and divergence, 18 whole plastome sequences, representing all *Rosa* subgenera and sections, were selected for comparative analysis (Table 1 and S1). Firstly, Geneious

v8.0.2 was used for extracting all protein-coding genes of each plastome sequence, with their RSCU subsequently examined with MEGA v7.0 [142]. Additionally, these plastome sequences were analyzed in the online software MISA to identify SSRs, especially mono-, di-, tri-, tetra-, penta-, and hexanucleotides using parameters set to 10, 5, 4, 3, 3, and 3, respectively. Additionally, the online program REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) [143] was used, with the parameters set to a mini mum repeat size of 30 bp as well as a Hamming distance of 3, to identify long repeat sequences, including reverse (R), forward (F), and palindromic (P) repeats. Finally, tandem repeats (T) were discovered using Tandem Repeats Finder v4.07 (https://tandem.bu.edu/trf/trf.html) [144] with the default settings.

To determine whether there are structural differences among these *Rosa* plastomes, the SC/IR boundaries were mapped with IRscope [145] to assess IR expansion/contraction, with shifts in the IR/SC boundary being attributed to different causes, i.e., IR expansion/contraction or gene loss. Genome rearrangement and inversions were detected via the Mauve [146] plugin in Geneious v8.0.2, while divergence in the *Rosa* plastome was plotted in Shuffle-LAGAN mode using the mVISTA online program (http://genome.lbl.gov/vista/mvista/submit.shtml) [147], with the *R. minutifolia* plastome (MT755634) as the reference.

Potential hotspots of nucleotide diversity were also identified in *Rosa* plastomes using the Perl scripts "2_extract_bed_CDS_RNA_and_intergenic.pl" and "generate_gene_matrix_from_one_ fasta_file.pl" (https://github.com/quxiaojian/Bioinformatic_Scripts) to automatica lly extract noncoding and coding parts from plastomes. Overall, 79 PCGs and 132 noncoding regions (including introns and intergenic regions) were extracted and separately aligned via the MAFFT v7.471 plugin integrated into PhyloSuite v1.2.2, after which the nucleotide polymorphism (Pi) values of each region were evaluated via DnaSP v6.12.03 [148].

To detect the positively selected sites of protein-coding sequences (CDSs) in *Rosa* plastomes, the ratio of synonymous (dS) and nonsynonymous (dN) substitutions was determined according to the formula $\omega = dN/dS$ using the CodeML algorithm in EasyCodeML v1.4 [149], with ratios of $\omega < 1$, $\omega = 1$, and $\omega > 1$ indicating negative, neutral and positive selections, respectively. After aligning each single-copy CDS in codon mode, it was concatenated via PhyloSuite v1.2.2 into one matrix in PAML format as the input sequence file [150]. Similarly, the ML tree, generated with IQ-TREE v1.6.8 [151], was used as an input tree. In preset mode, potential positive selection was tested with a site model, while the analyses were performed on the basis of four site models (M1a vs. M2a, M0 vs. M3, M8a vs. M8, and M7 vs. M8). In this case, an LRT threshold of $p < 0.05$ signalled adaptation within the genome. By comparing four pairs of site-specific models, M7 vs. M8 was used for estimating positive selection sites based on ω and LRT values. BEB [152] analysis was then implemented in the M8 model for detecting positively selected sites of specific genes (Table S9).

Phylogenetic analysis

To investigate how members within the genus *Rosa* were related, a total of 106 representative *Rosa* plastid genomes (13 were newly sequenced) were used to construct phylogenetic trees, with *Fragaria vesca* subsp. *vesca* (JF345175) and *Potentilla micrantha* (HG931056) in Potentilleae (Rosaceae) selected as outgroups (Table S1). The trees, based on 79 CDSs and complete plastid genome sequences, were generated using BI and ML methods. PhyloSuite v1.2.2 was then used to perform the subsequent series of analyses. The 79 unique CDSs and whole plastome sequences were aligned separately via MAFFT v7.471 [153], and this was followed by an incongruence length difference test in PAUP v4.0a168 to assess the possibility of combining data from different genes ($P < 0.01$). The aligned CDS matrices were individually checked and manually adjusted for small errors in AliView [154] prior to concatenation in PhyloSuite v1.2.2 to yield a single 68,703 bp supermatrix. DAMBE v 7.0.68 [155] was then used to assess substitutional saturation of the concatenated matrix. Regarding the phylogenies, IQ-TREE v1.6.8 was used with the 'Auto' option to automatically select the model for 5000 ultrafast bootstraps (the ultrafast bootstrap option using 5000 replicates) before constructing the ML tree, with a Shimodaira–Hasegawa-like approximate likelihood-ratio test (SH-aLRT) also run for branches. Similarly, BI phylogenies were inferred via MrBayes v3.2 [156] with ModelFinder used to identify the best-fitting substitution model [157]. The Markov chain Monte Carlo algorithm was executed for 2,000,000 generations, with sampling performed every 1000 generations. Convergence and stationarity were confirmed when the average standard deviations of split frequencies fell below 0.01. The first 25% of the sampled trees were discarded as burn-in, and the remaining ones were used to construct a majority-rule consensus tree. FigTree v1.4.4 (https://tree.bio.ed.ac.uk/software/figtree/) and TreeGraph v2.15.0–887 beta [158] were used to visualize and annotate the final trees.

Abbreviations

BI	Bayesian inference
ML	Maximum likelihood
SSR	Simple sequence repeats
Pi	DNA polymorphism
RSCU	Relative synonymous codon usage
CDS	Coding sequence
IR	Inverted repeat
SC	Single copy

SSC Small single copy
LSC Large single copy
cp Chloroplast

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06734-0.

Supplementary Material 1: Table S1: Taxonomic information and GenBank accession numbers of 108 samples used for comparative genomics and phylogenetic analyses. **Table S2**: Basic characteristics of all sampled plastomes (106 *Rosa* and two outgroups). **Table S5**: Codon usage in the plastid genomes of 18 *Rosa* species. **Table S6**: Distribution of simple sequence repeats (SSR) in the plastid genomes of 18 *Rosa* species. **Table S7**: Distribution of long repeat sequences (TFPR) in the plastid genomes of 18 *Rosa* species. **Table S8**: Bayesian empirical Bayesian (BEB) analysis based on the M8 model detected positive selection sites in 79 shared CDSs of 18 *Rosa* plastomes. **Table S9**: Comparison of site models and LRT results for 79 shared CDSs in the 18 *Rosa* plastomes.

Supplementary Material 2: Table S3: Gene compositions of the *Rosa* plastomes. **Table S4**: Genes in the *Rosa* plastomes containing introns, along with the intron and exon lengths.

Supplementary Material 3: Figure S1: Collinearity analysis of 18 representative *Rosa* plastomes from different subgenera and sections. **Figure S2**: Comparison of nucleotide polymorphisms (Pi) across all 106 *Rosa* plastomes. **Figure S3**: Consensus phylogenetic tree of complete plastome sequences of 106 *Rosa* samples and two outgroups using Bayesian inference (BI) and maximum likelihood (ML) methods.

Acknowledgements

We are grateful to Molecular Biology Experiment Center, Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences for providing the experimental platform and data. We sincerely thank Soulaiman Sakr for his valuable comments and English language editing of the manuscript. We finally thank the editor and the constructive comments from the anonymous reviewers who helped in improving the manuscript.

Author contributions

JM and SLH conceived designed the research and acquired these fundings. HJ performed the experiment, analyzed the data and drafted the manuscript. JH and YJZ analyzed the data. WLG, YZ and XJL provided suggestions on structuring the article and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Yunnan Fundamental Research Projects (grant NO. 202201AT070256 and 202301BD070001-032).

Data availability

All sequences in this study are available in the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/nuccore/), with GenBank accession numbers shown in Table 1 and Table S1.

Declarations

Ethics approval and consent to participate

The authors confirm that all methods comply with local and national regulations. No materials from animal or human were used in this research.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]College of Landscape and Horticulture, Yunnan Agricultural University, Kunming 650201, China
[2]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
[3]Southeast Asia Biodiversity Research Institute, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences & Center for Integrative Conservation, Chinese Academy of Sciences, Menglun 666303, China

Received: 27 November 2024 / Accepted: 16 May 2025
Published online: 04 June 2025

References

1. Lim KY, Werlemark G, Matyasek R, Bringloe JB, Sieber V, El Mokadem H, Meynet J, Hemming J, Leitch AR, Roberts AV. Evolutionary implications of permanent odd polyploidy in the stable sexual, pentaploid of *Rosa canina* L. Heredity. 2005;94(5):501–6.
2. Jian HY, Zhang H, Tang KX, Li SF, Wang QG, Zhang T, Qiu XQ, Yan HJ. Decaploidy in *Rosa praelucens* Byhouwer (Rosaceae) endemic to Zhongdian plateau, Yunnan, China. Caryologia. 2010;63(2):162–7.
3. Rehder A. Manual of cultivated trees and shrubs hardy in North America. Collier Macmillan Ltd: New York 1940;426–51.
4. Wissemann V. Conventional taxonomy (wild roses). London: Academic 2003;111–7.
5. Ku TC, Robertson KR. *Rosa* Linnaeus. In: Wu CY, Raven PH, editors. Flora of China. Volume 9. Beijing/St, Louis, Science Press/Missouri Botanical Garden Press; 2003. p. 339–81.
6. Gudin S, Rose. Genet Breed. 2000;17:160–89.
7. Debener T, Linde M. Exploring complex ornamental genomes: the rose as a model plant. Crit Rev Plant Sci. 2009;28(4):267–80.
8. Tomljenovic N, Pejic I. Taxonomic review of the genus *Rosa*. Agriculturae Conspectus Scientificus. 2018;83(3):139–47.
9. Leus L, Van Laere K, De Riek J, Van Huylenbroeck J. Rose. In: Van Huylenbroeck J, editor. Ornamental crops. Springer International Publishing: Cham; 2018;719–67.
10. Fougère-Danezan M, Joly S, Bruneau A, Gao XF, Zhang LB. Phylogeny and biogeography of wild roses with specific attention to polyploids. Ann Bot. 2015;115(2):275–91.
11. Shepherd RE. History of the rose. Soil science. Volume 77. Macmillan: New York 1954;416.
12. De Vries DP, Dubois LAM. Rose breeding past, present, prospects. 1996;242–8.
13. Jarvis CE. Seventy-two proposals for the conservation of types of selected linnaean generic names, the report of subcommittee 3 C on the lectotypification of linnaean generic names. Taxon. 1992;41(3):552–83.
14. Barrie FR. Report of the general committee. Taxa. 2006;55:795–800.
15. Matsumoto S, Wakita H, Fukui H. Molecular classification of wild roses using organelle DNA probes. Sci Hortic. 1997;68(1):191–6.
16. Olsson Å, Nybom H, Prentice HC. Relationships between nordic Dogroses (*Rosa* L. sect. *Caninae*, Rosaceae) assessed by RAPDs and elliptic fourier analysis of leaflet shape. Syst Bot. 2000;25(3):511–21.
17. Jan CH, Byrne DH, Manhart J, Wilson H. Rose germplasm analysis with RAPD markers. HortScience. 1999;34(2):341–5.
18. Millan T, Osuna F, Cobos S, Torres AM, Cubero JI. Using RAPDs to study phylogenetic relationships in *Rosa*. Theor Appl Genet. 1996;92(2):273–7.
19. Debener T, Bartels C, Mattiesch L. RAPD analysis of genetic variation between a group of rose cultivars and selected wild rose species. Mol Breed. 1996;2(4):321–7.
20. Atienza SG, Torres AM, Millan T, Cubero JI. Genetic diversity in *Rosa* as revealed by RAPDs. Agriculturae Conspectus Scientificus. 2005;70(3):75–85.
21. Martin M, Piola F, Chessel D, Jay M, Heizmann P. The domestication process of the Modern Rose: genetic structure and allelic composition of the rose complex. Theor Appl Genet. 2001;102(2):398–404.
22. Jürgens AH, Seitz B, Kowarik I. Genetic differentiation of *Rosa canina* (L.) at regional and continental scales. Plant Syst Evol. 2007;269(1–2):39–53.
23. Mirzaei L, Rahmani F. Genetic relationships among *Rosa* species based on random amplified polymorphic DNA (RAPD). Afr J Biotechnol. 2011;10(55):11373–7.
24. De Cock K, Vander Mijnsbrugge K, Breyne P, Van Bockstaele E, Van Slycken J. Morphological and AFLP–based differentiation within the taxonomical complex section *Caninae* (subgenus *Rosa*). Ann Bot. 2008;102(5):685–97.

25. Joly S, Bruneau A. Delimiting species boundaries in *Rosa* sect. *Cinnamomeae* (Rosaceae) in Eastern North America. Syst Bot. 2007;32(4):819–36.
26. Basaki T, Mardi M, Kermani MJ, Pirseyedi SM, Ghaffari MR, Haghnazari A, Shanjani PS, Koobaz P. Assessing *Rosa persica* genetic diversity using amplified fragment length polymorphisms analysis. Sci Hortic. 2009;120(0):538–43.
27. Koopman WJM, Wissemann V, De Cock K, Van Huylenbroeck J, De Riek J, Sabatino GJH, Visser D, Vosman B, Ritz CM, Maes B, Werlemark G, Nybom H, De Keyser E, Van Laere K, Smulders MJM. AFLP markers as a tool to reconstruct complex relationships: A case study in *Rosa* (Rosaceae). Am J Bot. 2008;95:353–66.
28. Riek JD, Cock KD, Smulders MJM, Nybom H. AFLP-based population structure analysis as a means to validate the complex taxonomy of Dogroses (*Rosa* section *Caninae*). Mol Phylogenet Evol. 2013;67(3):547–59.
29. Vogt JC, Herklotz V, Ritz CM. Epigenetic patterns in genetically imbalanced polyploid dog rose hybrids (*Rosa* L. Sect. *Caninae* (DC.) Ser.) revealed by cDNA–amplified fragment length polymorphisms and methylation-sensitive amplified polymorphisms. Int J Plant Sci. 2015;176(5):433–45.
30. Rusanov K, Kovacheva N, Vosman B, Zhang L, Rajapakse S, Atanassov A, Atanassov I. Microsatellite analysis of *Rosa damascena* Mill. accessions reveals genetic similarity between genotypes used for rose oil production and old Damask rose varieties. Theor Appl Genet. 2005;111(4):804–9.
31. Nybom H, Esselink GD, Werlemark G, Vosman B. Microsatellite DNA marker inheritance indicates preferential pairing between two highly homologous genomes in polyploid and hemisexual dog-roses, *Rosa* L. Sect. *Caninae* DC. Heredity. 2004;92(3):139–50.
32. Babaei A, Tabaei-Aghdaei SR, Khosh-Khui M, Omidbaigi R, Naghavi MR, Esselink GD, Smulders MJ. Microsatellite analysis of Damask rose (*Rosa damascena* Mill.) accessions from various regions in Iran reveals multiple genotypes. BMC Plant Biol. 2007;7:12.
33. Scariot V, Akkak A, Botta R. Characterization and genetic relationships of wild species and old garden roses based on microsatellite analysis. J Am Soc Hortic Sci. 2006;131(1):66–73.
34. Meng J, Li DZ, Yi TS, Yang JB, Zhao XF. Development and characterization of microsatellite loci for *Rosa odorata* var. *gigantea* Rehder & E.H.Wilson (Rosaceae). Conserv Genet. 2009;10(6):1973.
35. Tang KX, Qiu XQ, Zhang H, Li SF, Wang QG, Jian HY, Yan B, Huang XQ. Study on genetic diversity of some *Rosa* germplasm in Yunnan based on SSR markers. Acta Horticulturae Sinica. 2008;35(08):1227–32.
36. Alsemaan T, Albatal N, Baydar H, Almaarri K. Genetic diversity and qulaitative variation of *Rosa damascena* in Syria. Int J Agric Res. 2011;6(5):429–36.
37. Samiei L, Naderi R, Khalighi A, Shahnejat Bushehriet AA, Mozaffarian V, Esselink GD, Osaloo SK, Smulders MJM. Genetic diversity and genetic similarities between Irianian rose species. J Hortic Sci Biotechnol. 2010;85(3):231–7.
38. Kiani M, Zamani Z, Khalighi A, Fatahi R, Byrne DH. Microsatellite analysis of Iranian Damask rose (*Rosa damascena* Mill.) germplasm. Plant Breeding. 2010;129:551–7.
39. Xu F. Study on genetic diversity of 90 *Rosa* germplasm based on SSR markers. Master. Southwest University, Chongqing 2009.
40. Qiu XQ, Zhang H, Li SF, Jian HY, Tang KX. Relative relationships analysis of rose germplasm in Yunnan based on SSR markers. Acta Bot Boreali-Occidentalia Sinica. 2009;29(9):1764–71.
41. Akond M, Jin S, Wang X. Molecular characterization of selected wild species and miniature roses based on SSR markers. Sci Hortic. 2012;147:89–97.
42. Ritz CM, Wissemann V. Microsatellite analyses of artificial and spontaneous dogrose hybrids reveal the hybridogenic origin of *Rosa micrantha* by the contribution of unreduced gametes. J Hered. 2011;102(2):217–27.
43. Herklotz V, Ritz CM. Multiple and asymmetrical origin of polyploid dog rose hybrids (*Rosa* L. sect. *Caninae*(DC.) Ser.) involving unreduced gametes. Ann Bot. 2017; 120(2):209–20.
44. Schanzer I, Vagina A. ISSR (Inter Simple Sequence Repeat) markers reveal natural intersectional hybridization in wild roses [*Rosa* L., sect. *Caninae* (DC.) Ser. and sect. *Cinnamomeae* (DC.) Ser.]. Kärntner Botanikzentrum. 2007;14.
45. Yang F, Zeng I, Ye K, Zhao ZG, Zhang P, Gong XW, Yin Q, Sun Q. Relationship of 17 *Rosa* plants detected by morphology and ISSR analysis. Bull Bot Res. 2011;31(02):193–8.
46. Schanzer IA, Kutlunina NA. Interspecific hybridization in wild roses (*Rosa* L. sect. *Caninae* DC). Biol Bull. 2010;37(5):480–8.
47. Jabbarzadeh Z, Khosh-Khui M, Salehi H, Shahsavar A, Saberivand A. Assessment of genetic relatedness in roses by ISSR markers. World Appl Sci J. 2013;28:2085–90.
48. Vukosavljev M, Arens P, Voorrips RE, van 't Westende WP, Esselink GD, Bourke PM, Cox P, van de Weg WE, Visser RG, Maliepaard C, Smulders MJ. High-density SNP-based genetic maps for the parents of an outcrossed and a selfed tetraploid garden rose cross, inferred from admixed progeny using the 68k Rose SNP array. Hortic Res. 2016;3:16052.
49. Zhang J, Esselink GD, Che D, Fougère-Danezan M, Arens P, Smulders MJM. The diploid origins of allopolyploid rose species studied using single nucleotide polymorphism haplotypes flanking a microsatellite repeat. J Hortic Sci Biotechnol. 2013;88(1):85–92.
50. Yan M, Byrne DH, Klein PE, Yang J, Dong Q, Anderson N. Genotyping-by-sequencing application on diploid rose and a resulting high-density SNP-based consensus map. Hortic Res. 2018;5:17.
51. Wissemann V. Molecular evidence for allopolyploid origin of the *Rosa canina*-complex (Rosaceae, Rosoideae). J Appl Bot. 2002;76(5–6):176–8.
52. Wu S, Ueda Y, Nishihara S, Matsumoto S. Phylogenetic analysis of Japanese *Rosa* species using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS). J Hortic Sci Biotechnol. 2001;76(2):127–32.
53. Matsumoto S, Nishio H, Ueda Y, Fukui H. Phylogenetic analyses of genus *Rosa*: Polyphyly of section *Pimpinellifoliae* and origin of *Rosa* x *fortuniana* Lindl. Proceedings of the Third International Symposium on Rose Research and Cultivation. 2001;(547):357–63.
54. Wu SQ, Ueda Y, He HY, Nishihara S, Matsumoto S. Phylogenetic analysis of Japanese *Rosa* species using *matK* sequences. Breed Sci. 2000;50(4):275–81.
55. Matsumoto S, Kouchi M, Fukui H, Ueda Y. Phylogenetic analyses of the subgenus *Eurosa* using the ITS nrDNA sequence. Acta Hortic. 2000;521:193–202.
56. Wissemann V. Genetic constitution of *Rosa* sect. *Caninae* (*R. canina*, *R. jundzillii*) and Sect. *Gallicanae* (*R. gallica*). J Appl Bot. 1999;73(5–6):191–6.
57. Matsumoto S, Kouchi M, Yabuki J, Kusunoki M, Ueda Y, Fukui H. Phylogenetic analyses of the genus *Rosa* using the *matK* sequence: molecular evidence for the narrow genetic background of modern roses. Sci Hortic. 1998;77(1–2):73–82.
58. Joly S, Starr JR, Lewis WH, Bruneau A. Polyploid and hybrid evolution in roses east of the Rocky Mountains. Am J Bot. 2006;93(3):412–25.
59. Joly S, Bruneau A. Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. Syst Biol. 2006;55(4):623–36.
60. Wissemann V, Ritz CM. The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS-1 and *atpB-rbcL* intergenic spacer (IGS) versus conventional taxonomy. Bot J Linn Soc. 2005;147(3):275–90.
61. Ritz CM, Schmuths H, Wissemann V. Evolution by reticulation: European dogroses originated by multiple hybridization across the genus *Rosa*. J Hered. 2005;96(1):4–14.
62. Bruneau A, Starr JR, Joly S. Phylogenetic relationships in the genus *Rosa*: new evidence from chloroplast DNA sequences and an appraisal of current knowledge. Syst Bot. 2007;32(2):366–78.
63. Qiu XQ, Zhang H, Jian HY, Wang QG, Zhou NN, Yan HJ, Zhang T, Tang KX. Genetic relationships of wild roses, old garden roses, and modern roses based on internal transcribed spacers and *mat*K sequences. HortScience. 2013;48(12):1445–51.
64. Che DD, Zhang JZ, Wang JG, Fan JP, Gong SF. In analysis of *ITS1* sequences of nrDNA in *Rosa* species from Northeast China. Leuven, Belgium: International Society for Horticultural Science (ISHS); 2013. pp. 341–4.
65. Qiu XQ, Zhang H, Wang QG, Jian HY, Yan HJ, Zhang T, Wang JH, Tang KX. Phylogenetic relationships of wild roses in China based on nrDNA and *mat*K data. Sci Hortic. 2012;140(6):45–51.
66. Meng J, Fougère-Danezan M, Zhang LB, Li DZ, Yi TS. Untangling the hybrid origin of the Chinese tea roses: evidence from DNA sequences of single-copy nuclear and chloroplast genes. Plant Syst Evol. 2011;297(3):157–70.
67. Zhu ZM, Gao XF. Molecular evidence for the hybrid origin of *Rosa lichiangensis* (Rosaceae). Phytotaxa. 2015;222(3):221–8.
68. Liu CY, Wang GL, Wang H, Xia T, Zhang SZ, Wang QG, et al. Phylogenetic relationships in the genus *Rosa* revisited based on *rpl16*, *trnL-F*, and *atpB-rbcL* sequences. HortScience. 2015;50(11):1618–24.
69. Jian HY, Tang KX, Sun H. Phylogeography of *Rosa soulieana* (Rosaceae) in the Hengduan Mountains: refugia and 'melting' pots in the Quaternary climate oscillations. Plant Syst Evol. 2015;301(7):1819–30.
70. Zhu ZM. Molecular phylogenetic studies on the sections *Chinenses* and *Synstylae* of the genus *Rosa*. PhD, University of Chinese Academy of Sciences, China, 2015.
71. Deng HN. Molecular phylogeny and speciation of *Rosa* section *Microphyllae*. Master, Southwest University, Chongqing 2015.
72. Zhu ZM, Gao XF, Fougère-Danezan M. Phylogeny of *Rosa* sections *Chinenses* and *Synstylae* (Rosaceae) based on chloroplast and nuclear markers. Mol Phylogenet Evol. 2015;87(0):50–64.

73. Ahmed SM. Phylogenetic analysis of *Rosa damascena* L. from Taif using DNA barcoding approach. Pak J Bot. 2019;51(1):157–64.
74. Zhou YQ. Molecular phylogeny of genus *Rosa* L. and the possible origin of several cultivars. Master, Yunnan Normal University, Kunming 2016.
75. Wu S, Nishihara S, Ueda Y. Phylogenetic analysis of section *Synstylae* in the genus *Rosa* based on RAPD markers. Acta Hortic. 2001;547:391–402.
76. Iwata H, Kato T, Ohno S. Triparental origin of Damask roses. Gene. 2000;259:53–9.
77. Leus L, Jeanneteau F, Van Huylenbroeck J, Van Bockstaele E, De Rick J. Molecular evaluation of a collection of rose species and cultivars by AFLP, ITS, *rbc*L and *mat*K. Acta Hortic. 2004;651:141–7.
78. Gao YD, Zhang Y, Gao XF, Zhu ZM. Pleistocene glaciations, demographic expansion and subsequent isolation promoted morphological heterogeneity: A phylogeographic study of the alpine *Rosa sericea* complex (Rosaceae). Sci Rep. 2015;5:11698.
79. Meng J, He SL, Li DZ, Yi TS. Nuclear genetic varation of *Rosa odorata* var. *gigantea* (Rosaceae): population structure and conservation implications. Tree Genet Genomes. 2016;0(1):65–79.
80. Yang CY, Yu C, Ma YJ, Luo L, Pan HT, Zhang QX. Phylogenetic relationships of the genus *Rosa* based on SSR markers and single copy nuclear gene. J Beijing Forestry Univ. 2018;40(12):85–96.
81. Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76:273–97.
82. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci USA. 2007;104(49):19363–8.
83. Jansen RK, Cai ZQ, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK. el at. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA. 2007;104:19369–74.
84. Barrett CF, Davis JI, Leebens-Mack J, Conran JG, Stevenson DW. Plastid genomes and deep relationships among the commelinid monocot angiosperms. Cladistics. 2013;29(1):65–87.
85. Birky CW. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc Natl Acad Sci USA. 1995;92(25):11331.
86. Jansen RK, Raubeson LA, Boore JL, Depamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ. el at. Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol. 2005;395:348–84.
87. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. In Chloroplast biotechnology: methods and protocols. 2014;3–38.
88. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134.
89. Palmer JD. Comparative organization of chloroplast genomes. Annu Rev Genet. 1985;19(1):325–54.
90. Zheng XM, Wang JR, Feng L, Liu S, Pang HB, Qi L, Li J, Sun Y, Qiao WH, Zhang LF, Cheng YL, Yang QW. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. Sci Rep. 2017;7(1):1555.
91. Ravi V, Khurana JP, Tyagi AK, Khurana P. An update on chloroplast genomes. Plant Syst Evol. 2008;271(1–2):101–22.
92. Jeon JH, Kim SC. Comparative analysis of the complete chloroplast genome sequences of three closely related East-Asian wild roses (*Rosa* sect. *Synstylae*; Rosaceae). Genes. 2019;10(1):23.
93. Yin XM, Liao BS, Guo S, Liang CL, Pei J, Xu J, Chen SL. The chloroplasts genomic analyses of *Rosa laevigata*, *R. rugosa* and *R. canina*. Chin Med. 2020;15:18.
94. Li CH, Zheng YQ, Huang P. Molecular markers from the chloroplast genome of rose provide a complementary tool for variety discrimination and profiling. Sci Rep. 2020;10(1):12188.
95. Cui WH, Du XY, Zhong MC, Fang W, Suo ZQ, Wang D, Dong X, Jiang XD, Hu JY. Complex and reticulate origin of edible roses (*Rosa*, Rosaceae) in China. Hortic Res. 2022;9:uhab051.
96. Zhang C, Li SQ, Xie HH, Liu JQ, Gao XF. Comparative plastid genome analyses of *Rosa*: insights into the phylogeny and gene divergence. Tree Genet Genomes. 2022;18(3):20.
97. Shen WX, Dong ZH, Zhao WZ, Ma LY, Wang F, Li WY, Xin PY. Complete chloroplast genome sequence of *Rosa lucieae* and its characteristics. Horticulturae. 2022;8(9):788.
98. Gao CW, Li T, Zhao X, Wu CH, Zhang Q, Zhao XZ, Wu MX, Lian YH, Li ZQ. Comparative analysis of the chloroplast genomes of *Rosa* species and RNA editing analysis. BMC Plant Biol. 2023;23(1):318.
99. Ren QM, Wang YC, Lin YN, Zhen ZH, Cui YL, Qin S. The extremely large chloroplast genome of the green alga *Haematococcus pluvialis*: genome structure, and comparative analysis. Algal Res. 2021;56.
100. Turmel M, Otis C, Lemieux C. Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. Sci Rep. 2017;7(1):994.
101. Weng ML, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol. 2013;31(3):645–59.
102. Wu S, Chen JY, Li Y, Liu A, Li A, Yin M, Shrestha N, Liu JQ, Ren GP. Extensive genomic rearrangements mediated by repetitive sequences in plastomes of *Medicago* and its relatives. BMC Plant Biol. 2021;21(1):421.
103. Jiang H, Tian J, Yang JX, Dong X, Zhong ZX, Mwachala G, Zhang CF, Hu GW, Wang QF. Comparative and phylogenetic analyses of six Kenya *Polystachya* (Orchidaceae) species based on the complete Chloroplast genome sequences. BMC Plant Biol. 2022;22(1):177.
104. Nie XJ, Lv SZ, Zhang YX, Du XH, Wang L, Biradar SS, Tan XF, Wan FH, Song WN. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). PLoS ONE. 2012;7(5):e36869.
105. Zhu AD, Guo WH, Gupta S, Fan WS, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016;209(4):1747–56.
106. Yu JJ, Fu J, Fang YP, Xiang J, Dong HJ. Complete chloroplast genomes of *Rubus* species (Rosaceae) and comparative analysis within the genus. BMC Genomics. 2022;23(1):32.
107. Sebastin R, Kim J, Jo IH, Yu JK, Jang W, Han S, Park HS, AlGarawi AM, Hatamleh AA, So YS, Shim D, Chung JW. Comparative Chloroplast genome analyses of cultivated and wild *Capsicum* species shed light on evolution and phylogeny. BMC Plant Biol. 2024;24(1):797.
108. Gu XL, Li LL, Li SC, Shi WX, Zhong XN, Su YJ, Wang T. Adaptive evolution and co-evolution of chloroplast genomes in Pteridaceae species occupying different habitats: overlapping residues are always highly mutated. BMC Plant Biol. 2023;23:511.
109. Dong WP, Xu C, Li CH, Sun JH, Zuo YJ, Shi S, Cheng T, Guo JJ, Zhou SL. *ycf1*, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5:8348.
110. Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M. Uncovering the protein translocon at the chloroplast inner envelope membrane. Science. 2013;339:571–4.
111. Hao DC, Chen SL, Xiao PG. Molecular evolution and positive darwinian selection of the chloroplast maturase *matK*. J Plant Res. 2010;123:241–7.
112. Stern DB, Goldschmidt-Clermont M, Hanson MR. Chloroplast RNA metabolism. Annu Rev Plant Biol. 2010;61:125–55.
113. Kode V, Mudd EA, Iamtham S, Day A. The tobacco plastid *accd* gene is essential and is required for leaf development. Plant J. 2005;44:237–44.
114. Mahendrarajah TA, Moody ERR, Schrempf D, Szánthó LL, Dombrowski N, Davín AA, Pisani D, Donoghue PCJ, Szöllősi GJ, Williams TA, Spang A. ATP synthase evolution on a cross-braced dated tree of life. Nat Commun. 2023;14:7456.
115. Ferreira KN, Iverson TM, Maghlaoui K, Barber J, Iwata S. Architecture of the photosynthetic oxygen-evolving center. Science. 2004;303:1831–8.
116. Martín M, Sabater B. Plastid *ndh* genes in plant evolution. Plant Physiol Biochem. 2010;48:636–45.
117. Yamori W, Shikanai T. Physiological functions of cyclic electron transport around photosystem I in sustaining photosynthesis and plant growth. Annu Rev Plant Biol. 2016;67:81–106.
118. Wen F, Wu XZ, Li TJ, Jia ML, Liu XS, Liao L. The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China. BMC Genomics. 2021;22:161.
119. Wang Y, Wen F, Hong X, Li ZL, Mi YL, Zhao B. Comparative chloroplast genome analyses of *Paraboea* (Gesneriaceae): insights into adaptive evolution and phylogenetic analysis. Front Plant Sci. 2024;13:1019831.
120. Piot A, Hackel J, Christin PA, Besnard G. One-third of the chloroplast genes evolved under positive selection in PACMAD grasses. Planta. 2018;247:255–66.
121. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. In: Maliga P, editor. Chloroplast Biotechnology. Methods Mol Biol. 2014;1132:3–38.
122. Saha A, Das S, Moin M, Dutta M, Bakshi A, Madhav MS, Kirti PB. Genome-wide identification and comprehensive expression profiling of ribosomal protein

small subunit (RPS) genes and their comparative analysis with the large subunit (RPL) genes in rice. Front Plant Sci. 2017;8:1553.
123. Serino G, Maliga P. RNA polymerase subunits encoded by the plastid *rpo* genes are not shared with the nucleus-encoded plastid enzyme. Plant Physiol. 1998;117:1165–70.
124. Zhang Y, Tian L, Lu CM. Chloroplast gene expression: recent advances and perspectives. Plant Commun. 2023;4:100611.
125. Liu HZ, Ye H, Zhang NY, Ma JY, Wang JT, Hu GJ, et al. Comparative analyses of chloroplast genomes provide comprehensive insights into the adaptive evolution of *Paphiopedilum* (Orchidaceae). Horticulturae. 2022;8(5):391.
126. Geng LY, Jiang TY, Chen X, Li Q, Ma JH, Hou WX, Tang CQ, Wang Q, Deng YF. Plastome structure, phylogeny and evolution of plastid genes in *Reevesia* (Helicteroideae, Malvaceae). J Plant Res. 2024;137:589–604.
127. Wissemann V. Conventional taxonomy (Wild Roses). Reference module in life sciences. 2017;1–6.
128. Debray K, Paslier MCL, Bérard A, Thouroude T, Michel G, Marie-Magdelaine J, Bruneau A, Foucher F, Malécot V. Unveiling the patterns of reticulated evolutionary processes with phylogenomics: hybridization and polyploidy in the genus *Rosa*. Syst Biol. 2022;71(3):547–69.
129. Debray K, Marie-Magdelaine J, Ruttink T, Clotault J, Foucher F, Malécot V. Identification and assessment of variable single-copy orthologous (SCO) nuclear loci for low-level phylogenomics: a case study in the genus *Rosa* (Rosaceae). BMC Evol Biol. 2019;19:1–19.
130. Yang C, Ma YJ, Cheng BX, Zhou LJ, Yu C, Luo L, Pan HT, Zhang QX. Molecular markers indicate that the narrow Québec endemics *Rosa rousseauiorum* and *Rosa williamsii* are synonymous with the widespread *Rosa blanda*. Can J Bot. 2005;83(4):386–98.
131. Zhang Z, Yang T, Liu Y, Wu S, Sun HH, Wu J, Li YH, Zheng Y, Ren HR, Yang YY, Shi SC, Wang WY, Pan Q, Lian LJ, Duan SW, Zhu YX, Cai YM, Zhou HG, Zhang H, Tang KX, Cui JP, Gao D, Chen LY, Jiang YH, Sun XM, Zhou XF, Fei ZJ, Ma N, Gao JP. Haplotype-resolved genome assembly and resequencing provide insights into the origin and breeding of modern rose. Nat Plants. 2024;10:1659–71.
132. Zhao X, Gao CW. The complete chloroplast genome sequence of *Rosa minutifolia*. Mitochondrial DNA Part B. 2020;5:3320–1.
133. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci USA. 2009;106:13875–9.
134. Meng J, Jiang H, Zhang LN, He J. Characterization of the complete plastid genome of an important Chinese Old Rose *Rosa odorata* var. *pseudindica*. Mitochondrial DNA Part B. 2019;4:679–80.
135. Jiang H, He J, Meng J. Characterization of the complete plastid genome of a Chinese endangered species *Rosa rugosa* Thunb. Mitochondrial DNA Part B. 2019;4:1679–80.
136. Jian HY, Zhang SD, Zhang T, Qiu XQ, Yan HJ, Li SB, Wang QG, Tang KX. Characterization of the complete chloroplast genome of a critically endangered decaploid rose species, *Rosa praelucens* (Rosaceae). Conserv Genet Resour. 2018;10:851–4.
137. Group TAP, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS. el at. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181(1):1–20.
138. Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, Yang JB, Li DZ, Yi TS. Diversification of Rosaceae since the late cretaceous based on plastid phylogenomics. New Phytol. 2017;0(214):1355–67.
139. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19(1):11–5.
140. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15(1):50.
141. Lohse M, Drechsel O, Kahlau S, Bock R. Organellar genome DRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 2013;41(W1):W575–81.
142. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
143. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
144. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
145. Amiryousefi A, Hyvonen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1.
146. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394–403.
147. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32(S2):W273–9.
148. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.
149. Gao FL, Chen CJ, Arab DA, Du ZG, He YH, Ho SYW. EasyCodeML: a visual tool for analysis of selection using CodeML. Ecol Evol. 2019;9(7):3891–8.
150. Zhang D, Gao FL, Jakovlic I, Zou H, Zhang J, Li WX, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20:348–55.
151. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
152. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22(4):1107–18.
153. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
154. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30(22):3276–8.
155. Xia XH. DAMBE7: new and improved tools for data analysis. Mol Biol Evol. 2018;35(6):1550–2.
156. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.
157. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. Model finder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
158. Stöver BC, Müller KF. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics. 2010;11:7.

Publisher's note