

RESEARCH

Open Access



The chloroplasts genomic analyses of *Rosa laevigata*, *R. rugosa* and *R. canina*

Xianmei Yin^{1†}, Baosheng Liao^{2†}, Shuai Guo², Conglian Liang³, Jin Pei¹, Jiang Xu^{2*} and Shilin Chen²

Abstract

Background: Many species of the genus *Rosa* have been used as ornamental plants and traditional medicines. However, industrial development of roses is hampered due to highly divergent characteristics.

Methods: We analyzed the chloroplast (cp) genomes of *Rosa laevigata*, *R. rugosa* and *R. canina*, including the repeat sequences, inverted-repeat (IR) contractions and expansions, and mutation sites.

Results: The size of the cp genome of *R. laevigata*, *R. rugosa* and *R. canina* was between 156 333 bp and 156 533 bp, and contained 113 genes (30 tRNA genes, 4 rRNA genes and 79 protein-coding genes). The regions with a higher degree of variation were screened out (*trnH-GUU*, *trnS-GCU*, *trnG-GCC*, *psbA-trnH*, *trnC-GCA*, *petN*, *trnT-GGU*, *psbD*, *petA*, *psbJ*, *ndhF*, *rpl32*, *psaC* and *ndhE*). Such higher-resolution loci lay the foundation of barcode-based identification of cp genomes in *Rosa* genus. A phylogenetic tree of the genus *Rosa* was reconstructed using the full sequences of the cp genome. These results were largely in accordance with the current taxonomic status of *Rosa*.

Conclusions: Our data: (i) reveal that cp genomes can be used for the identification and classification of *Rosa* species; (ii) can aid studies on molecular identification, genetic transformation, expression of secondary metabolic pathways and resistant proteins; (iii) can lay a theoretical foundation for the discovery of disease-resistance genes and cultivation of *Rosa* species.

Keywords: *Rosa* species, cp genome, Phylogeny

Background

Rosaceae is a large and diverse family with 100 genera and 3000 species. *Rosa* is a typical genus of the Rosaceae family. *Rosa chinensis* Jacq., *Rosa laevigata* Michx. and *Rosa rugosa* Thunb are documented in the 2015 version of *Chinese Pharmacopoeia* [1].

Plants of the genus *Rosa* are distributed in the temperate and subtropical regions of the Northern hemisphere [2, 3]. The genus *Rosa* has garnered increasing attention as a medicinal agent recently [4–6]. Due to the potential economic and medicinal value of peonies, it is important

to understand the genetic relationships within species for future application of germplasm resources.

In conventional taxonomy, the genus *Rosa* is divided into four subgenera (*Hulthemia*, *Rosa*, *Platyrhodon*, and *Hesperhodos*), and the subgenus of *Rosa* is divided further into 10 sections (*Pimpinellifoliae*, *Galicanae*, *Cani-nae*, *Carolinae*, *Rosa*, *Synstylae*, *Chinenses* [syn. *Indicae*], *Banksianae*, *Laevigatae*, and *Bracteatae* [7, 8].

Despite numerous recent studies examining phylogenetic relationships in the genus *Rosa*, relationships remain obscure because of: (i) hybridization in nature and in the garden, and low levels of chloroplast and nuclear genome variation [9–11]; (ii) phylogenetic analyses only based on a small number of non-coding chloroplast sequences show low internal resolution [12, 13]. *Rosa laevigata*, *R. rugosa* and *R. canina* have been employed in traditional Chinese medicine (TCM)

*Correspondence: jxu@icmm.ac.cn

[†]Xianmei Yin and Baosheng Liao contributed equally to this work

²Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institution of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
Full list of author information is available at the end of the article



formulations. However, several sympatric species of *Rosa* have been used in TCM formulations, and the diversity of medicinal materials can affect the quality and safety of medicinal materials severely.

Chloroplasts are the descendants of ancient bacterial endosymbionts. They are the common organelles of green plants, and have an essential role in photosynthesis [14]. In general, inheritance of the cp genome is patrilineal in gymnosperms, but maternal in angiosperms [15]. The cp genome is conservative in structure, contains a large single-copy (LSC) region, small single-copy (SSC) region, and two inverse repeat (IR) regions. The cp genome is an ideal research model for the study of molecular identification, phylogeny, species conservation, and genome evolution [16, 17]. Over the last decade, researchers have gained more in-depth understanding of chloroplasts, including their origin, structure, evolution, genetic engineering, as well as forward and reverse genetics [18–20]. In addition, the development of sequencing technology has greatly promoted chloroplast study [21, 22], now generating massive chloroplast genome sequence data, helping to overcome the previously unresolved relationships. Moreover, it also provides genomic information such as structure, gene order, content, and mutations in which the critical information of species identification is provided [23–27].

In previous studies, chloroplast genomes provided the effective information for identifying *Rosa* species [28, 29]. The chloroplast genomes of two species from the genus *Rosa*, *R. chinensis* and *R. rugosa*, which have been collected in Chinese Pharmacopoeia 2015 were published. In the present study, the remainder of the recorded species of the genus *Rosa* in the *Chinese Pharmacopoeia*, including two used in TCM (*R. laevigata*, *R. rugosa*) and a traditional medicine used worldwide (*R. canina*) were identified based on the chloroplast genome. The structural characteristics, phylogenetic relationships, inter-specific divergence among *R. laevigata*, *R. rugosa* and *R. canina* were documented.

Materials and methods

DNA sequencing, assembly and validation of the cp genome

The fresh leaves of *R. laevigata* and *R. rugosa* plants were collected in Shennongjia (Hubei Province, China). The dried flowers of *R. canina* were purchased at a medicinal market in Beijing, China. The cetyltrimethylammonium-bromide method was used to extract the whole genomic DNA of tree peonies [30]. The DNA concentration was measured using a ND-2000 spectrometer (NanoDrop Technologies, Wilmington, DE, USA). A shotgun library (250 bp) was constructed according to manufacturer (Vazyme Biotech, Nanjing, China) instructions.

Sequencing was accomplished with the XTM Ten platform (Illumina, San Diego, CA, USA) using the double terminal sequencing method (pair-end 150). The amount of raw data from the sample was 5.0 G, and >34 million paired-end reads were attained.

Raw data were filtered by Skewer-0.2.2 [31]. Chloroplast-like reads were predicted from clean-reads by BLAST [32] searches using the sequences of the reference *Rosa chinensis*. Then, the cp reads were used to assemble sequences by SOAPdenovo-2.04 [33]. Finally, sequences were extended and gaps filled with SSPACE-3.0 and GapCloser-1.12 [34, 35]. To validate the accuracy of junction splicing, random primers were designed to test the four junctions of the sequence by polymerase chain reaction.

Gene annotation and sequence analyses

Sequence annotation was achieved by CpGAVAS [36]. DOGMA (<http://dogma.cccb.utexas.edu/>) and BLAST were used to check the results of annotation [37]. All transfer tRNA genes with default settings were detected by tRNAscanSEv1.21 [38]. The structural features of the cp genome were drawn by OGDRAWv1.2 [39]. MEGA5.2 was used to define relative use of synonymous codons [40].

Comparison of cp genomes

The cp genomes of *Rosa* species were completed by mVISTA [41] (Shuffle-LAGAN mode) using the genome of *R. chinensis* as the reference. Tandem Repeats Finder [42] was used to detect tandem repeats, forward repeats, and palindromic repeats as tested by REPuter [43]. Detection of simple sequence repeats (SSRs) was done by Misa.pl [44] using search parameters of mononucleotides set to ≥ 10 repeat units, dinucleotides ≥ 8 repeat units, trinucleotides and tetranucleotides ≥ 4 repeat units, and pentanucleotides and hexanucleotides ≥ 3 repeat units.

Phylogenetic analyses

Phylogenetic trees were constructed using the genomic sequences of 21 chloroplasts. The sequences were aligned using clustalw2. Construction of an unrooted phylogenetic tree was achieved using the neighbor-joining (NJ) approach with MEGA5.2 [40] with bootstrap replicates of 1000. *Hibiscus rosa-sinensis* was set as the outgroup.

Results

DNA features of the chloroplasts of *R. laevigata*, *R. rugosa* and *R. canina*

The size of the cp genomes ranged from 156 333 bp to 156 533 bp. Among them, the largest cp genome was of *R. rugosa* (156 533 bp) and the smallest cp genome was of *R. laevigata* (156 333 bp). The total guanine + cytosine (G + C) content of the three genomes was 37.3%.

R. laevigata, *R. rugosa* and *R. canina* had a cp genome with a similar structure: LSC region, SSC region, and a pair of inverted repeats (IRA/IRB). For *R. laevigata*, *R. rugosa* and *R. canina*, the length of the LSC region of the cp genome varied from 85 452 bp to 85 657 bp, and the G + C content from 35.2% to 35.3%; the length of SSC-region distribution was from 18 742 bp to 18 785 bp, and the G + C content was from 31.3 to 31.4%. The IR region had a length distribution from 26 048 bp to 26 053 bp, and the G + C content was 42.7% (Table 1). The DNA G + C content is an important indicator of species affinity [45], and *R. laevigata*, *R. rugosa* and *R. canina* have highly similar cpDNA G + C content. The DNA G + C content of the IR regions was higher than that of LSC and SSC regions, which is similar to that seen with other angiosperms [46]. In general, the relatively high DNA G + C content of the IR regions is attributable to rRNA genes and tRNA genes [47, 48]. After annotation, the sequences of the whole cp genome of *R. laevigata*, *R. rugosa* and *R. canina* was submitted to the National Center for Biotechnology Information database (NCBI), the GenBank accession number in Table 1.

A physical map of the cp genomes of *R. laevigata*, *R. rugosa* and *R. canina* was drawn according to annotation results using OGDRAW [39] (Fig. 1). A total of 113 genes were contained in the cp genome of *R. laevigata*, *R. rugosa* and *R. canina*: four rRNA genes, 30 tRNA genes, and 79 protein-coding genes (Table 2). Most genes could be divided crudely into three groups: “self-replication-related”, “photosynthesis-related”, and “other” (Table 2) [49].

Table 1 Summary of complete chloroplast chloroplast genomes for *R. laevigata*, *R. rugosa* and *R. canina*

Species	<i>R. rugosa</i>	<i>R. laevigata</i>	<i>R. canina</i>
Large single copy			
Length (bp)	85,657	85,452	85,653
G + C (%)	35.2	35.3	35.2
Length (%)	54.7	54.7	54.7
Small single copy			
Length (bp)	18,780	18,785	18,742
G + C (%)	31.4	31.3	31.4
Length (%)	12.0	12.0	12.0
Inverse repeat			
Length (bp)	26,048	26,048	26,053
G + C (%)	42.7	42.7	42.7
Length (%)	16.7	16.7	16.7
Total			
Length (bp)	156,533	156,333	156,501
G + C (%)	37.3	37.3	37.3
Accession number	MN661138	MN661139	MN661140

In all anticipated genes of the cp genomes of *R. laevigata*, *R. rugosa* and *R. canina*, introns were discovered in 17 genes: six tRNA genes and 11 protein-encoding genes (Table 3). The tRNA genes with introns were *trnK-UUU*, *trnL-UAA*, *trnV-UAC*, *trnI-GAU*, *trnG-UCC* and *trnA-UGC*. The 11 coding genes with introns were *rps12*, *rps16*, *rpl16*, *rpl2*, *rpoC1*, *ndhA*, *ndhB*, *ycf3*, *petB*, *clpP* and *petD*. Three of the 17 intron-containing genes were inserted by three introns (*rps12*, *ycf3*, *clpP*). The remainder of the genes were inserted by only one intron. Of these, *trnH-UUU* contained the largest intron (2500 bp), which contained the whole *matK*. Similar to other angiosperms, *rps12* of chloroplasts in *R. laevigata*, *R. rugosa* and *R. canina* resulted from trans-splicing activity. The 5' end of *rps12* was in the LSC region, and the 3' end was in the IR region.

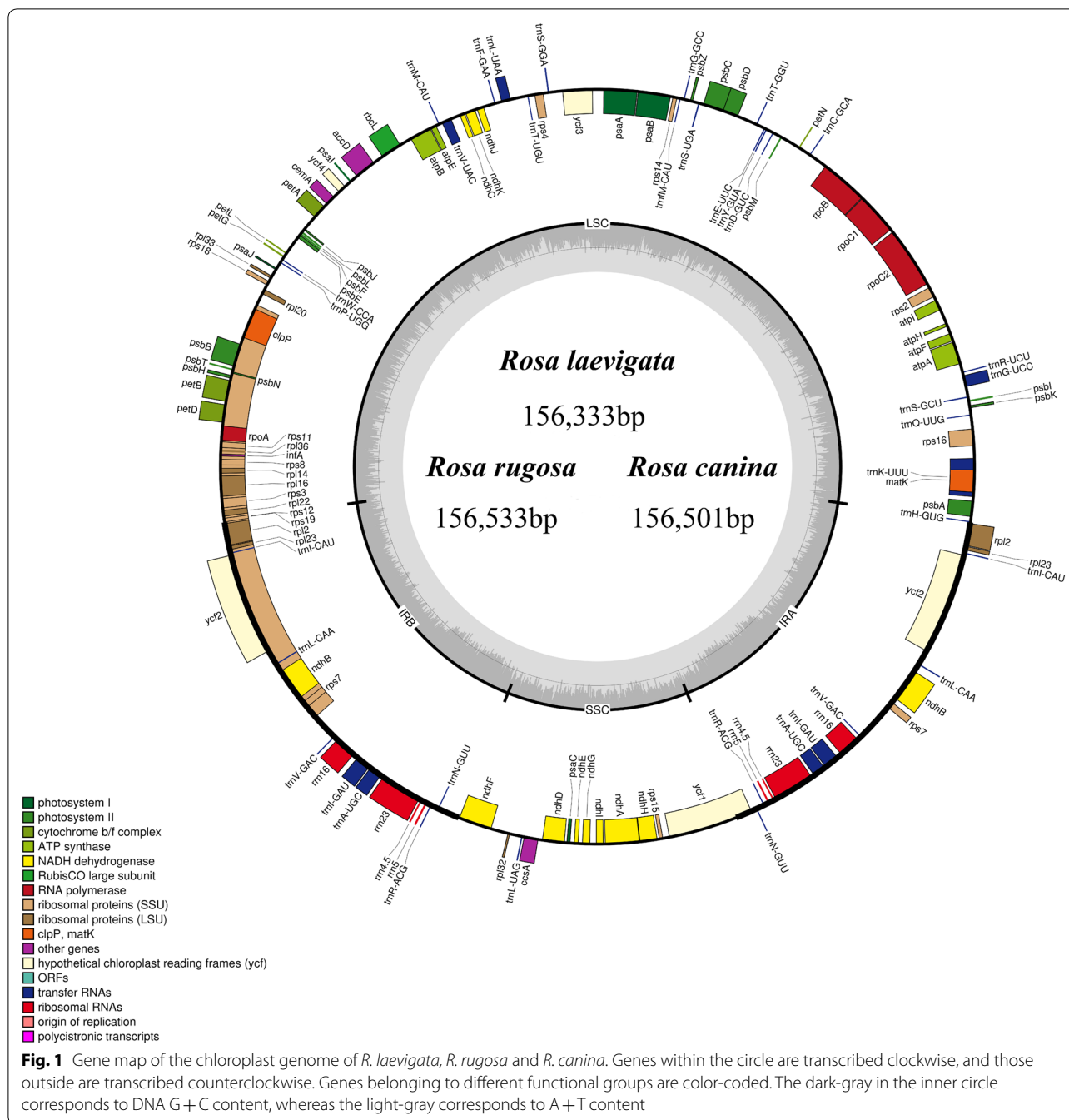
Analyses of long repetitive sequences and SSRs

For *R. laevigata*, *R. rugosa* and *R. canina*, interspersed repeated sequences (IRSs) were evaluated in the cp genomes with a repeat-unit length of ≥ 30 bp. These comprised forward repeats, reverse repeats, complementary repeats, and palindromic repeats. Fifty 50 IRSs were found in *R. rugosa*; 60 IRS in *R. laevigata* and 50 IRS in *R. canina*. Among all types of IRS, the sequence lengths of 20–29 bp occurred most frequently. IRS analyses of the cp genomes of *R. laevigata*, *R. rugosa* and *R. canina* are shown as Fig. 2.

SSRs are disposed to slipped-strand mispairing, which is a key mutational mechanism for generating SSR polymorphisms [50]. SSRs at the intra-specific level in the cp genome are variable, so they are used regularly as genetic markers in studies of evolution and population genetics [51–53]. We found 63 SSRs in *R. rugosa*, 62 SSRs in *R. canina*, and 65 SSRs in *R. laevigata* (Fig. 3).

Genomic sequences

To ascertain differences in the genomic sequences of chloroplasts of *R. laevigata*, *R. rugosa* and *R. canina*, we used the sequence in *R. chinensis* as a reference (Fig. 4). Variability in the IR region of the cp genomes was considerably lower than that of LSC and SSC regions. In addition, most of the protein-coding genes of chloroplasts were highly conserved, except for the large variation in protein-coding genes of some genes (e.g., *rps19*, *petB*, and *ycf2*). Regions with a higher degree of variation among chloroplast genomic sequences were usually located in intergenic regions, such as the spacers for: *trnH-GUU*; *trnS-GCU* and *trnG-GCC*; *psbA-trnH*, *trnC-GCA* and *petN*; *trnT-GGU* and *psbD*; *petA* and *psbJ*; *ndhF* and *rpl32*; *psaC* and *ndhE*. Identification of such higher-resolution loci was necessary for use as barcodes for species identification.



Comparison of IR regions in the cp genomes of *R. laevigata*, *R. rugosa*, *R. canina* and *R. chinensis*

Gene location was relatively conservative in *R. laevigata*, *R. rugosa*, *R. canina* and *R. chinensis*. In these four species, *rps19* was located in the LSC region, *rpl2* in the IRa region, and *ndhF* in the SSC region. However, the coding region of *ycf1* was at the border of SSC/IRb, and spanned the LSC region and IRb region, so the IRa/

SSC boundary (5' end was lost) region created a pseudogene. The region of mutations in the *ycf1* pseudogene in the IRa/SSC region was 1106–1118 bp (Fig. 5). The double-strand break repair theory is considered to be the main mechanism for expansion and contraction of the IR region. Large shrinkages of the IR region are relatively rare.

Table 2 Genes in the chloroplast genome of *R. laevigata*, *R. rugosa* and *R. canina*

Category	Group of genes	Name of genes
Self-replication	Large subunit of ribosomal proteins	<i>rpl2*</i> , a, 14, 16*, 20, 22, 23a, 33, 36
	Small subunit of ribosomal proteins	<i>rps2</i> , 3, 4, 7a, 8, 11, 12*, a, 14, 16*, 18, 19
	DNA-dependent RNA polymerase	<i>rpoA</i> , B, C1*, C2
	rRNA genes	<i>rrn16Sa</i> , <i>rrn23Sa</i> , <i>rrn4.5Sa</i> , <i>rrn5Sa</i>
	tRNA genes	<i>trnA-UGC*</i> , a, <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnfm-CAU</i> , <i>trnG-UCC*</i> , <i>trnG-GCC</i> , <i>trnH-GUG</i> , <i>trnI-CAU</i> , <i>trnI-GAU*</i> , a, <i>trnK-UUU*</i> , <i>trnL-CAA</i> , <i>trnL-UAA*</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> , <i>trnV-UAC*</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>
Photosynthesis	Photosystem I	<i>psaA</i> , B, C, I, J
	Photosystem II	<i>psbA</i> , B, C, D, E, F, H, I, J, K, L, M, N, T, Z,
	NADH oxidoreductase	<i>ndhA*</i> , B*, a, C, D, E, F, G, H, I, J, K
	Cytochrome b6/f complex	<i>petA</i> , B*, D*, G, L, N
	ATP synthase	<i>atpA</i> , B, E, F*, H, I
	Rubisco	<i>rbcl</i>
Other genes	Maturase	<i>matK</i>
	Protease	<i>clpP*</i>
	Envelope membrane protein	<i>cemA</i>
	Subunit acetyl-CoA-carboxylase	<i>accD</i>
	c-Type cytochrome synthesis gene	<i>ccsA</i>
Conserved open-reading frames	<i>ycf1</i> , 2a, 3*, 4, 15	

Phylogenetic analyses

There have been many efforts to reconstruct the phylogenetic trees of plants of the genus *Rosa*. Several scholars have proposed that the extant classification system was artificial [12, 13], and that the interspecies relationships of *Rosa* are still ambiguous. The availability of the complete genomes of chloroplasts can provide further information for reconstruction of robust phylogeny for *Rosa*. A NJ tree was constructed for the cp genomes of 18 species of the Rosaceae family (Fig. 6). Species from the *Rosa* genus were monophyletic clade. Furthermore, *R. laevigata*, *R. rugosa*, *R. canina* and *R. chinensis* could be effectively divided into different sub-clades, and differentiated from each other efficiently. In which, *R. Chinensis* have a closer relationship with *R. rugosa*.

Discussion

We identified the cp genomes of *R. laevigata*, *R. rugosa* and *R. canina* in this study, which are used in TCM formulations. The cp genomes of *R. laevigata*, *R. rugosa* and *R. canina* showed high similarities in terms of genome size, gene classes, gene sequences, codon usage, and distribution of repeat sequences. This is partly because of the extremely low levels of sequence divergence observed across the *Rosa* genus [54, 55]. Some intergenic regions were identified with high degree of variation, which will be used as barcodes for species identification. We also investigated introns in all anticipated genes of three *Rosa*

species. Intron and/or gene losses have been reported for cp genomes [56–58]. Introns have important roles in regulation of gene expression [59], and they can control gene expression temporally and in a tissue-specific manner [60, 61]. Scholars have reported on the regulation mechanisms of introns for gene expression in plants and animals [62–64]. However, the connotations between intron loss and gene expression using the transcriptome for genus *Rosa* have not been published. More experimental work on the roles of introns shall be needed for future work. Comparative analysis of gene location in *R. laevigata*, *R. rugosa*, *R. canina* and *R. chinensis* revealed a pseudogene of *ycf1*, which may provide a basis for studying variations in the cp genomes of higher plants or algae.

Phylogenetic analyses revealed that *Rosa* genus belonged to monophyletic clade (Fig. 6), while their intra-family relationships were almost in agreement with those from a study by Zhang and Marie et al. [13, 65]. However, the exact phylogenetic location of some base taxons needs further verification, such as that the phylogenetic relationship of *R. rugosa* and *R. chinensis* in here contradicts what was previously reported, two *R. rugosa* species were clustered into two different clades. The possible reason is: complicates phylogeny reconstruction in roses was complicated by interspecific hybridization, some studies have suggested that there were frequent interspecific hybridization in the *Rosa* genus [11, 66–69]. Indeed, several contradictions between plastid and nuclear gene

Table 3 Length of exons and introns in genes with introns in the chloroplast genome of three medicinal roses

Species	Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
<i>R. rugosa</i>	<i>trnK-UUU</i>	LSC	35	350	37		
	<i>trnG-UCC</i>	LSC	23	698	48		
	<i>trnL-UAA</i>	LSC	35	549	50		
	<i>trnV-UAC</i>	LSC	35	600	39		
	<i>trnI-GAU</i>	IR	37	956	35		
	<i>trnA-UGC</i>	IR	38	814	35		
	<i>rps12*</i>	LSC	26	538	232		114
	<i>rps16</i>	LSC	230	881	40		
	<i>rpl16</i>	LSC	399	980	9		
	<i>rpl2</i>	IR	434	683	391		
	<i>rpoC1</i>	LSC	1611	769	453		
	<i>ndhA</i>	SSC	540	1216	552		
	<i>ndhB</i>	IR	756	678	777		
	<i>ycf3</i>	SSC	153	768	228	742	126
	<i>petB</i>	LSC	6	791	642		
	<i>clpP</i>	LSC	228	647	292	844	71
	<i>petD</i>	LSC	8	722	475		
	<i>R. laevigata</i>	<i>trnK-UUU</i>	LSC	35	2500	37	
<i>trnG-UCC</i>		LSC	23	697	48		
<i>trnL-UAA</i>		LSC	35	549	50		
<i>trnV-UAC</i>		LSC	35	600	39		
<i>trnI-GAU</i>		IR	37	956	35		
<i>trnA-UGC</i>		IR	38	816	35		
<i>rps12*</i>		LSC	26	538	232		114
<i>rps16</i>		LSC	230	872	40		
<i>rpl16</i>		LSC	399	964	9		
<i>rpl2</i>		IR	434	683	391		
<i>rpoC1</i>		LSC	1611	769	453		
<i>ndhA</i>		SSC	540	1245	522		
<i>ndhB</i>		IR	756	678	777		
<i>ycf3</i>		SSC	153	782	228	742	126
<i>petB</i>		LSC	6	793	642		
<i>clpP</i>		LSC	228	648	292	831	71
<i>petD</i>		LSC	8	722	475		
<i>R. cania</i>		<i>trnK-UUU</i>	LSC	35	2500	37	
	<i>trnG-UCC</i>	LSC	23	697	48		
	<i>trnL-UAA</i>	LSC	35	548	50		
	<i>trnV-UAC</i>	LSC	35	600	39		
	<i>trnI-GAU</i>	IR	37	956	35		
	<i>trnA-UGC</i>	IR	38	814	35		
	<i>rps12*</i>	LSC	26	538	232		114
	<i>rps16</i>	LSC	230	883	40		
	<i>rpl16</i>	LSC	399	964	9		
	<i>rpl2</i>	IR	434	683	391		
	<i>rpoC1</i>	LSC	1611	700	453		
	<i>ndhA</i>	SSC	540	1205	552		
	<i>ndhB</i>	IR	756	678	777		
	<i>ycf3</i>	SSC	153	785	228	740	126
	<i>petB</i>	LSC	6	792	642		
	<i>clpP</i>	LSC	228	650	292	839	71
	<i>petD</i>	LSC	8	728	475		

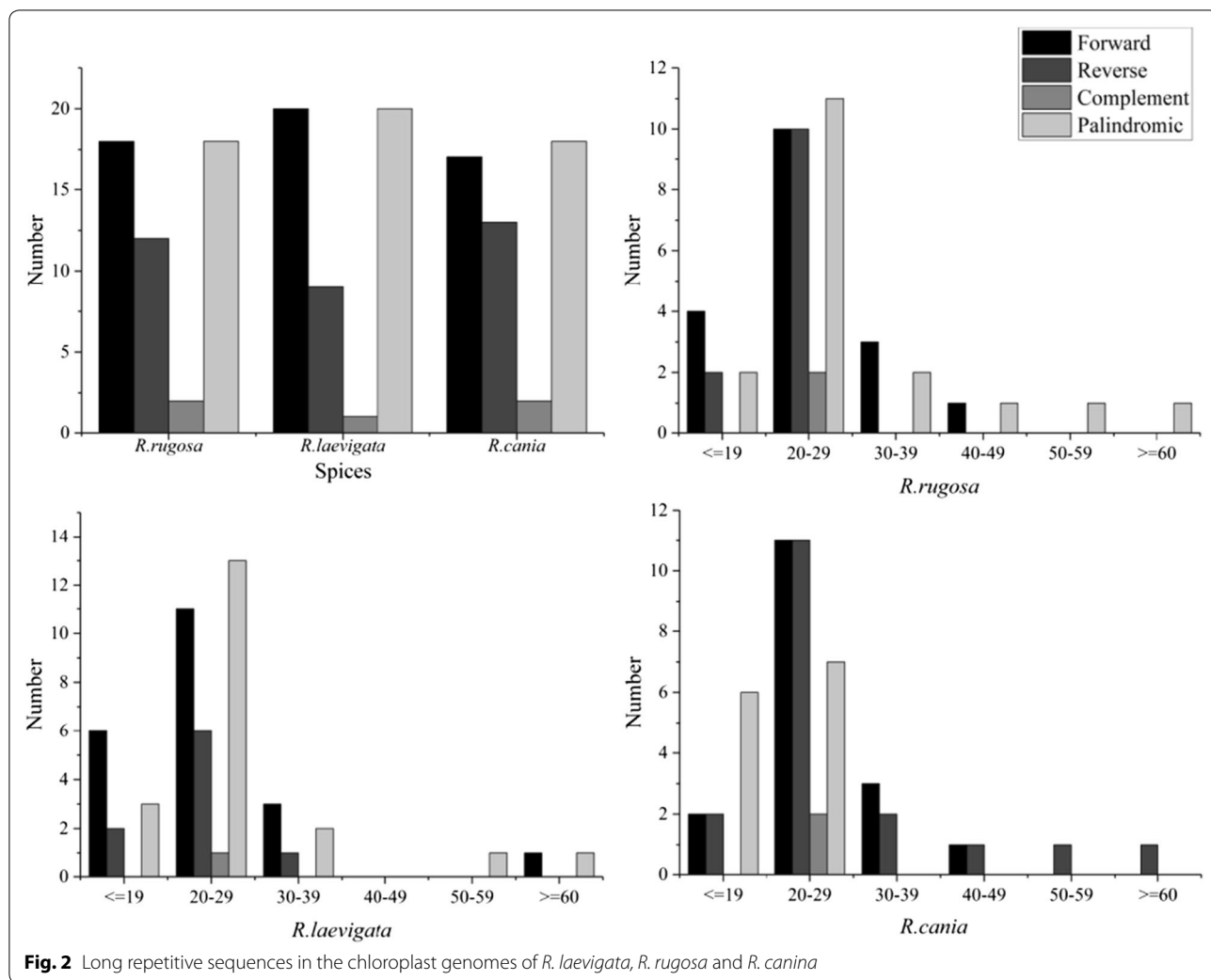


Fig. 2 Long repetitive sequences in the chloroplast genomes of *R. laevigata*, *R. rugosa* and *R. canina*

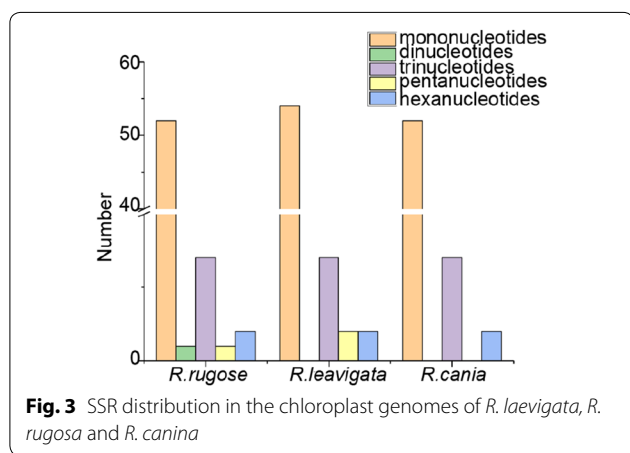


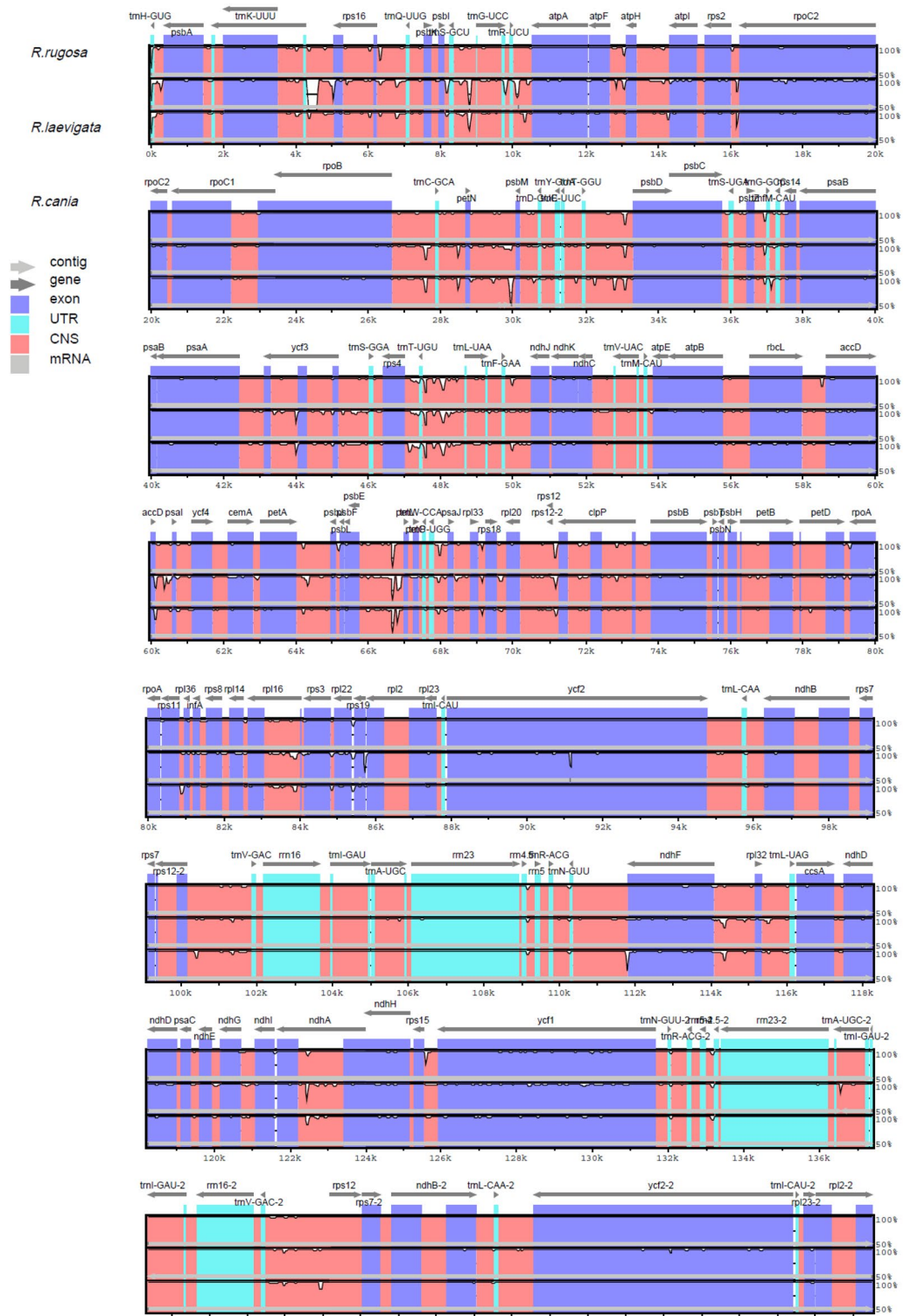
Fig. 3 SSR distribution in the chloroplast genomes of *R. laevigata*, *R. rugosa* and *R. canina*

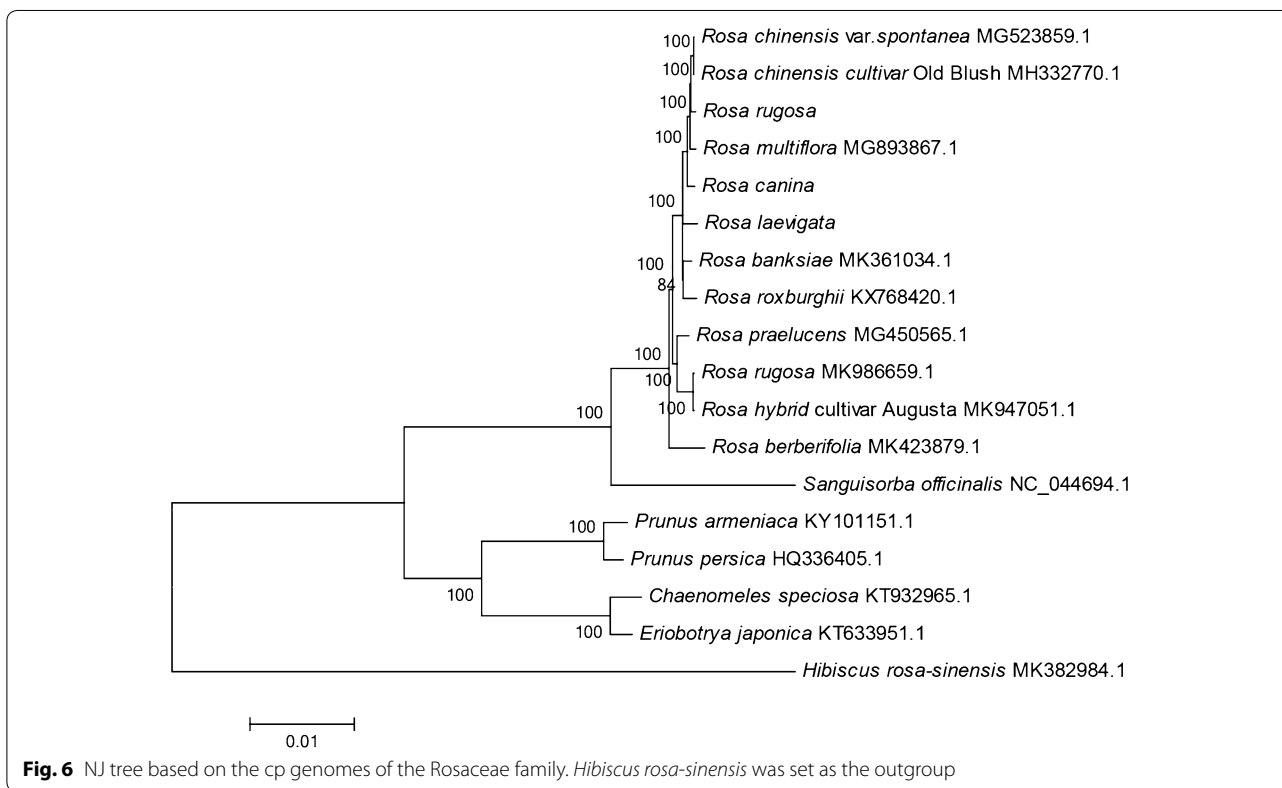
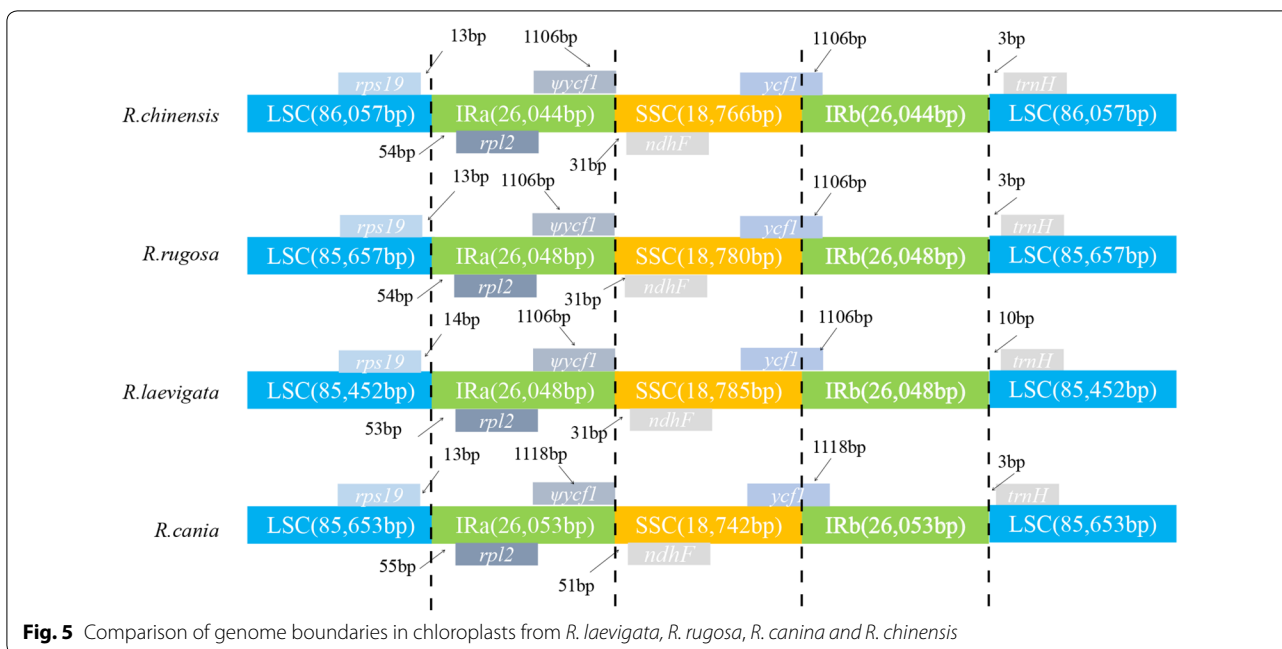
phylogenies of *Rosa* genus were discovered in previous study [55]. In addition, publications of numerous names given to morphological variants and hybrids, result in

Rosa taxonomy further complication [70]. Further identification of plant material or sequencing of those hybrids could explain why conspecific samples sometimes fall into distinct clades [12].

Conclusions

The whole cp genomes of *R. laevigata*, *R. rugosa* and *R. canina* was sequencing and analysis in this study. The status of the major taxa within the genus *Rosa* was consistent with our results for sequencing of cp genomes. *R. laevigata*, *R. rugosa*, *R. canina* and *R. chinensis* could be differentiated from other *Rosa* species efficiently. Our data reveal that cp genomes can be used for the identification and classification of *Rosa* species. Our results can aid studies on molecular identification, genetic transformation, and lay a theoretical foundation for the discovery of disease-resistance genes and cultivation of *Rosa*





species. Our observations complement the database of herbgenomics [71].

Abbreviations

LSC: Large single-copy; SSC: Small single-copy region; IR: Inverse repeat; cp: Chloroplast; NCBI: National Center for Biotechnology Information database; TCM: Traditional Chinese medicine.

Acknowledgements

I sincerely thank all those who provided an abundance of helpful on my data analysis and tried their best to improve my paper.

Authors' contributions

JX conceived and designed the research framework; JP collected and identified the sample; XY, SG and BL performed the experiments; XY, BL and CL analyzed the data; and XY and BL wrote the paper. SC made revisions to the final manuscript. All the authors read and approved the final manuscript.

Funding

The Major Project of "Research on modernization of traditional Chinese medicine", under Grant #2017YFC1702100.

Availability of data and materials

The datasets generated during the current study are available in the National Center for Biotechnology Information database (NCBI). <https://www.ncbi.nlm.nih.gov/MN661138>, MN661139, MN661140.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ College of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611130, China. ² Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institution of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China. ³ College of Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan 250355, China.

Received: 10 December 2019 Accepted: 5 February 2020

Published online: 13 February 2020

References

- Anonymous. Pharmacopoeia of People's Republic of China. Part 2. Chinese Medicine. Edited by Committee NP. Beijing: Science Press., 2015;75:200, 221.
- Wissemann V, Ritz CM. The genus *Rosa* (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.
- Christenhusz MJM, Fay MF, Chase MW. Plants of the world: an illustrated encyclopedia of vascular plants. Madroño. 2018;65(2):101–2.
- Uzunçakmak T, Ayaz Alkaya S. Effect of aromatherapy on coping with premenstrual syndrome: a randomized controlled trial. *Complement Ther Med.* 2017;36:63–7.
- Baiyisaiti A, Wang Y, Zhang X, et al. *Rosa rugosa* flavonoids exhibited PPAR α agonist-like effects on genetic severe hypertriglyceridemia of mice. *J Ethnopharmacol.* 2019;240:111952.
- Liu Y, Zhi D, Wang X, Fei D, et al. Rose (*R. Setate* x *R. Rugosa*) decoction exerts antitumor effects in *C. elegans* by down regulating Ras/MAPK pathway and resisting oxidative stress. *Int J Mol Med.* 2018;42(3):1–8.
- Rehder A. Bibliography of cultivated trees and shrubs hardy in the cooler temperate regions of the Northern Hemisphere, vol. 1. Jamaica Plain: Arnold Arboretum of Harvard University; 1949. p. 296–317.
- 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;87:50–64.
- Matthews JR. Hybridism and classification in the genus *rosa*. *New Phytol.* 1920;19(7–8):153–71.
- Rowley G. Some naming problems in *rosa*. *Bulletin du Jardin botanique de l'Etat Bruxelles.* 1959;29(3):205–11.
- Ritz CM, Schmutz H, Wissemann V. Evolution by reticulation: European *Dogroses* originated by multiple hybridization across the genus *Rosa*. *J Hered.* 2005;96(1):4–14.
- Joly S, Starr JR, Bruneau A. 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.
- Fougère-Danezan M, Simon J, Anne B, et al. Phylogeny and biogeography of wild roses with specific attention to polyploids. *Ann Bot.* 2015;115(2):275–91.
- Brunkard JO, Runkel AM, Zambryski PC. Chloroplasts extend stomules independently and in response to internal redox signals. *Proc Natl Acad Sci.* 2015;112(32):10044–9.
- Hu Y, Zhang Q, Rao G. Occurrence of plastids in the sperm cells of *Capri-foliaceae*: biparental plastid inheritance in angiosperms is unilaterally derived from maternal inheritance. *Plant Cell Physio.* 2008;49(6):958–68.
- Daniell H, Kumar S, Dufourmantel N. Breakthrough in chloroplast genetic engineering of agronomically important crops. *Trends Biotechnol.* 2005;23(5):238–45.
- Sigmon BA, Adams RP, Mower JP. Complete chloroplast genome sequencing of vetiver grass (*Chrysopogon zizanioides*) identifies markers that distinguish the non-fertile 'Sunshine' cultivar from other accessions. *Ind Crops Prod.* 2017;108:629–35.
- Kim KJ, Lee HL. Complete chloroplast genome sequence from Korean Ginseng (*Panax schiseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* 2004;11(4):247–61.
- Li P, Zhang S, Li F, et al. A phylogenetic analysis of chloroplast genomes elucidates the relationships of the six economically important brassica species comprising the triangle of U. *Front Plant Sci.* 2017;8:111.
- Shen X, Wu M, Liao B, et al. Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. *Molecules.* 2017;22(8):1330.
- Li R, Ma PF, Wen J, Yi TS. Complete sequencing of five *Araliaceae* chloroplast genomes and the phylogenetic implications. *PLoS ONE.* 2013;8(10):e78568.
- Yang JB, Li DZ, Li HT, et al. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour.* 2014;14(5):1024–31.
- Kane N, Sveinsson S, Dempewolf H, et al. Ultra-barcoding in cacao (*Theobroma* spp; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Am J Bot.* 2012;99(2):320–9.
- Dodsworth Steven. Genome skimming for next-generation biodiversity analysis. *Trends Plant Sci.* 2015;20(9):525–7.
- Wang A, Wu H, Zhu X, et al. Species identification of *Conyza bonariensis* assisted by chloroplast genome sequencing. *Front Genet.* 2018;9:374.
- Luo H, Shi J, Arndt W, et al. Gene Order Phylogeny of the genus *Prochlorococcus*. *PLoS ONE.* 2008;3(12):e3837.
- Luo H, Sun Z, Arndt W, et al. Gene order phylogeny and the evolution of methanogens. *PLoS ONE.* 2009;4(6):e6069.
- 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.
- Jiang HY, Zhang YH, Yan HJ, et al. The complete chloroplast genome of a key ancestor of modern roses, *rosa chinensis* var. *spontanea*, and a comparison with congeneric species. *Molecules.* 2018;23(2):389.
- Guo Q, Guo LL, Zhang L, et al. Construction of a genetic linkage map in tree peony (*Paeonia* Sect. *Moutan*) using simple sequence repeat (SSR) markers. *Sci Hortic.* 2017;219:294–301.
- Jiang H, Lei R, Ding SW, et al. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinform.* 2014;15:182.
- Deng P, Wang L, Cui L, et al. Global identification of microRNAs and their targets in Barley under salinity stress. *PLoS ONE.* 2015;10(9):e0137990.
- Gogniashvili M, Naskidashvili P, Bedoshvili D, et al. Complete chloroplast DNA sequences of Zanduri wheat (*Triticum* s). *Genet Resour Crop Evol.* 2015;62:1269–77.
- Acemel RD, Tena JJ, Irastorza-Azcarate I, et al. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate hox bimodal regulation. *Nat Genet.* 2016;48(3):336–41.

35. Boetzer M, Henkel CV, Jansen HJ, et al. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. 2010;27(4):578–9.
36. Liu C, Shi L, Zhu Y, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics*. 2012;13(1):715–22.
37. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004;20(17):3252–5.
38. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25(5):955–64.
39. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*. 2007;52(5–6):267–74.
40. Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28(10):2731–9.
41. Frazer KA, Pachter L, Poliakov A, et al. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32(Web Server issue):W273–9.
42. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573–80.
43. Kurtz S, Choudhuri JV, Ohlebusch E, et al. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29(22):4633–42.
44. Beier S, Thiel T, Münch T, et al. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33(16):2583–5.
45. Choi KS, Park S. The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae); genomic features and relationship with Asteraceae. *Gene*. 2015;572(2):214–21.
46. Guo S, Guo L, Zhao W, et al. Complete chloroplast genome sequence and phylogenetic analysis of *Aster tataricus*. *Molecules*. 2018;23(10):246.
47. Doorduyn L, Gravendeel B, Lammers Y, et al. The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Res*. 2011;18(2):93–105.
48. Lee SB, Kaittanis C, Jansen RK, et al. The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms. *BMC Genomics*. 2006;7:61.
49. Saski C, Lee S, Daniell H, et al. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plt Mol Biol*. 2005;59(2):309–22.
50. Asaf S, Khan AL, Khan MA, et al. chloroplast genomes of *Arabidopsis halleri* ssp. *gemmaifera* and *Arabidopsis lyrata* ssp. *petraea*: structures and comparative analysis. *Sci Rep*. 2017;7(1):7556.
51. Dong W, Xu C, Cheng T, et al. Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biol Evol*. 2013;5(5):989–97.
52. Yang Y, Zhou T, Duan D, et al. Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front Plant Sci*. 2016;7:959.
53. Suo Z, Li W, Jin X, Zhang H. A new nuclear DNA marker revealing both microsatellite variations and single nucleotide polymorphic loci: a case study on classification of cultivars in *Lagerstroemia indica* L. *J Microb Biochem Technol*. 2016;8:266–71.
54. Matsumoto S, Kouchi M, Yabuki J, et al. Phylogenetic analyses of the genus *Rosa* using the matK sequence: molecular evidence for the narrow genetic background of modern roses. *Scientia Hort*. 1998;77(1–2):73–82.
55. Wissemann V, Ritz CM. The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nr ITS-1 and atpB-rbcL intergenic spacer (IGS) versus conventional taxonomy. *Botanical J Linn Soc*. 2005;147(3):275–90.
56. Downie SR, Llanas E, Katz-Downie DS. Multiple independent losses of the rpoC1 intron in angiosperm chloroplast DNAs. *Syst Bot*. 1996;21(2):135–51.
57. Graveley BR. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet*. 2001;17(2):100–7.
58. Ueda M, Fujimoto M, Arimura SI, et al. Loss of the rpl32 gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. *Gene*. 2007;402(1–2):51–6.
59. Xu J, Chu Y, Liao B, et al. *Panax ginseng* genome examination for ginsenoside biosynthesis. *Gigascience*. 2017;6(11):1–15.
60. Le Hir H, Nott A, Moore MJ. How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci*. 2003;28(4):215–20.
61. Niu DK, Yang YF. Why eukaryotic cells use introns to enhance gene expression: splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity. *Biol Direct*. 2011;6:24.
62. Callis J, Fromm M, Walbot V. Introns increase gene expression in cultured maize cells. *Genes Dev*. 1987;1(10):1183–200.
63. Emami S, Arumainayagam D, Korf I, et al. The effects of a stimulating intron on the expression of heterologous genes in *Arabidopsis thaliana*. *Plant Biotechnol J*. 2013;11(5):555–63.
64. Ted C, Manley H, Cornelia G, et al. A generic intron increases gene expression in transgenic mice. *Mol Cell Biol*. 1991;11(6):3070–4.
65. Zhang SD, Jin JJ, Chen SY, et al. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytol*. 2017;214(3):1355–67.
66. Joly S, Starr JR, Lewis WH, et al. Polyploid and hybrid evolution in roses east of the Rocky Mountains. *Am J Bot*. 2006;93(3):412–25.
67. Mercure M, Bruneau A. Hybridization between the escaped *Rosa rugosa* (Rosaceae) and native *R. blanda* in Eastern North America. *Am J Bot*. 2008;95(5):597–607.
68. Ritz CM, Koehnen I, Groth M, et al. To be or not to be the odd one out—allele-specific transcription in pentaploid dogroses (*Rosa* L. sect. *Caninae* (DC) Ser). *BMC Plant Biol*. 2011;11(1):37.
69. Qiu XQ, Zhang H, Wang QG, et al. Phylogenetic relationships of wild roses in China based on nrDNA and matK data. *Sci Hort*. 2012;140:45–51.
70. Wissemann V. Conventional taxonomy (wild roses). In: Roberts AV, Debener T, Gudin S, editors. *Encyclopedia of rose science*. Amsterdam: Elsevier; 2003. p. 111–7.
71. Hu H, Shen X, Liao B, et al. Herbgonomics: a stepping stone for research into herbal medicine. *Sci China Life Sci*. 2019;62(7):913–20.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

