BMC Plant Biology

RESEARCH Open Access

Comparative analysis of the chloroplast genomes of *Rosa* species and RNA editing analysis

Chengwen Gao[1*], Teng Li[1,2], Xia Zhao[1], Chuanhong Wu[1], Qian Zhang[1], Xiangzhong Zhao[1], Mingxuan Wu[1], Yihong Lian[1] and Zhiqiang Li[1*]

Abstract

Background The genus *Rosa* (Rosaceae) contains approximately 200 species, most of which have high ecological and economic values. Chloroplast genome sequences are important for studying species differentiation, phylogeny, and RNA editing.

Results In this study, the chloroplast genomes of three *Rosa* species, *Rosa hybrida*, *Rosa acicularis*, and *Rosa rubiginosa*, were assembled and compared with other reported *Rosa* chloroplast genomes. To investigate the RNA editing sites in *R. hybrida* (commercial rose cultivar), we mapped RNA-sequencing data to the chloroplast genome and analyzed their post-transcriptional features. *Rosa* chloroplast genomes presented a quadripartite structure and had highly conserved gene order and gene content. We identified four mutation hotspots (*ycf3-trnS, trnT-trnL, psbE-petL*, and *ycf1*) as candidate molecular markers for differentiation in the *Rosa* species. Additionally, 22 chloroplast genomic fragments with a total length of 6,192 bp and > 90% sequence similarity with their counterparts were identified in the mitochondrial genome, representing 3.96% of the chloroplast genome. Phylogenetic analysis including all sections and all subgenera revealed that the earliest divergence in the chloroplast phylogeny roughly distinguished species of sections *Pimpinellifoliae* and *Rosa* and subgenera *Hulthemia*. Moreover, DNA- and RNA-sequencing data revealed 19 RNA editing sites, including three synonymous and 16 nonsynonymous, in the chloroplast genome of *R. hybrida* that were distributed among 13 genes.

Conclusions The genome structure and gene content of *Rosa* chloroplast genomes are similar across various species. Phylogenetic analysis based on the *Rosa* chloroplast genomes has high resolution. Additionally, a total of 19 RNA editing sites were validated by RNA-Seq mapping in *R. hybrida*. The results provide valuable information for RNA editing and evolutionary studies of *Rosa* and a basis for further studies on genomic breeding of *Rosa* species.

Keywords *Rosa*, Chloroplast genome, RNA editing, RNA-Seq, Phylogenetic analysis

*Correspondence:
Chengwen Gao
gaochengwen@qdu.edu.cn
Zhiqiang Li
lizqsjtu@qdu.edu.cn
[1]Medical Research Center, The Affiliated Hospital of Qingdao University, Qingdao University, Qingdao 266000, Shangdong, China
[2]School of Public Health, Qingdao University, Qingdao 266000, Shangdong, China

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

As a vital post-transcriptional regulation mechanism, RNA editing is pervasive in gene expression across chloroplast genomes of terrestrial plants [1, 2]. RNA editing typically involves conversion of cytidine (C) to uridine (U) within RNA molecules in the chloroplast genomes of higher plants [3, 4]. Reverse U-to-C editing has also been reported in plant organelle genomes, whereas U-to-C editing has been virtually absent in gymnosperms and angiosperms [1]. Most flowering plant chloroplast genomes have 20–60 RNA editing sites [5]. Chloroplast RNA editing sites decreased during angiosperm evolution [6–8]. Most RNA editing sites have been found in protein-coding regions, with a few sites located in untranslated regions, structural RNAs, and intronic regions [9]. Although the molecular mechanisms of RNA editing have been extensively studied [10], how RNA editing evolved in different species and about the mechanisms underlying the diversity of editing frequencies remain unclear. To date, relevant studies on detection of RNA editing sites via RNA-sequencing (RNA-Seq) read mapping and variant calling is lacking in the genus *Rosa*.

The genus *Rosa* L. (Rosaceae) contains approximately 200 species and grows in the subtropical and temperate regions of the northern hemisphere [11, 12]. Conventional taxonomy divided the genus *Rosa* into four subgenera (*Rosa*, *Hesperhodos*, *Hulthemia*, and *Platyrhodon*), while species of the subgenus *Rosa* are further divided into ten sections (*Rosa*, *Banksianae*, *Bracteatae*, *Caninae*, *Carolinae*, *Chinenses*, *Gallicanae*, *Pimpinellifoliae*, *Laevigatae*, and *Synstylae*) [13, 14]. *Rosa* species have extensive morphological variation and complex taxonomic profiles. In addition, reconstruction of the phylogeny of *Rosa* species has been difficult due to hybridization, incomplete lineage sorting, and low differentiation among the genus *Rosa* [15].

Chloroplasts are specialized plastids that contain chlorophyll to absorb light energy [16, 17]. Plant chloroplast genomes provide important information for exploring genetic diversity, understanding evolutionary differences, and generating high-resolution phylogenies, especially at low/complex taxonomic levels [18–20].

The chloroplast genome phylogenetic relationships of the genus *Rosa* still unclear because of the failure of species division, low resolution, limited samples, and low support values [15]. In the present study, the chloroplast genomes of three *Rosa* species, namely *R. hybrida* (Sect. *Chinenses*), *R. acicularis* (Sect. *Rosa*), and *R. rubiginosa* (Sect. *Caninae*), were assembled and compared. Among these three species, *R. acicularis* and *R. rubiginosa* have great medicinal importance, while *R. hybrida* is a commercial rose cultivar [21, 22]. Combined with the previously reported 41 chloroplast genomes of *Rosa*, we performed a comprehensive chloroplast genome analysis of this taxonomically difficult plant taxon. Furthermore, to our knowledge, we have for the first time determined RNA editing sites in the whole chloroplast genome of *R. hybrida* (commercial rose cultivar) using RNA-Seq data. This study aimed to (1) perform a comparative analysis of the chloroplast genomes of *Rosa* species; (2) ascertain highly variable regions in the *Rosa* chloroplast genome sequences; (3) identify chloroplast gene insertion in mitochondria; (4) obtain early evolutionary information on the chloroplast genomes of Rosa species and analyze molecular phylogeny by comparing chloroplast genomes; and (5) identify RNA editing sites of *R. hybrida* using RNA-Seq data. This study will provide a better understanding of the interspecific differences in the genus *Rosa* and will be valuable for further research on RNA editing in *Rosa* species.

Results

Characteristics of the *Rosa* chloroplast genomes

Raw sequence data of *R. rubiginosa*, *R. hybrida*, and *R. acicularis* were obtained, and the chloroplast genomes were 156,553 bp, 156,600 bp, and 157,219 bp long, respectively (Fig. 1). The three newly assembled *Rosa* chloroplast genomes were deposited in the GenBank database (OP032236, OP032237, and OP032238). They exhibited a quadripartite structure with a large single-copy (LSC) region (85,820–86,462 bp), dual inverted repeat (IR) regions (25,981–25,985 bp), and a small single-copy (SSC) region (18,763–18,787 bp), as shown in Fig. 2a.

The chloroplast genomes of *R. rubiginosa*, *R. hybrida*, and *R. acicularis* were conserved and contained 115 unique genes, of which 80 were protein-coding genes, 31 were transfer RNA (tRNA) genes, and 4 were ribosomal RNA (rRNA) genes. Seventeen genes had introns, of which eight protein-coding genes (*rpl16, rpl2, rps16, rpoC1, petB, petD, ndhA*, and *ndhB*) and six tRNAs (*trnG-UCC, trnI-GAU, trnK-UUU, trnL-UAA, trnA-UGC*, and *trnV-UAC*) contained one intron while the other three genes (*pafI, clpP1*, and *rps12*) had two introns.

Comparative analysis of *Rosa* chloroplast genomes

The *Rosa* chloroplast genomes had high sequence similarity. By comparison of the expansion and contraction of the IR/SC boundary between the chloroplast genomes of *Rosa*, it can be seen that the *Rosa* chloroplast genomes shows high similarity at the IR/SC boundary (Fig. 2a). The *rpl2* gene contained all LSC/IRb junctions, and the boundary gene between SSC and IRa/IRb is *ycf1*. Overall, the *Rosa* IR regions are similar in length and structure, which is consistent with previous findings [23, 24]. Similar to most terrestrial plants, the IR regions of chloroplast genome were more conserved than the LSC and SSC regions, and noncoding regions exhibited relatively

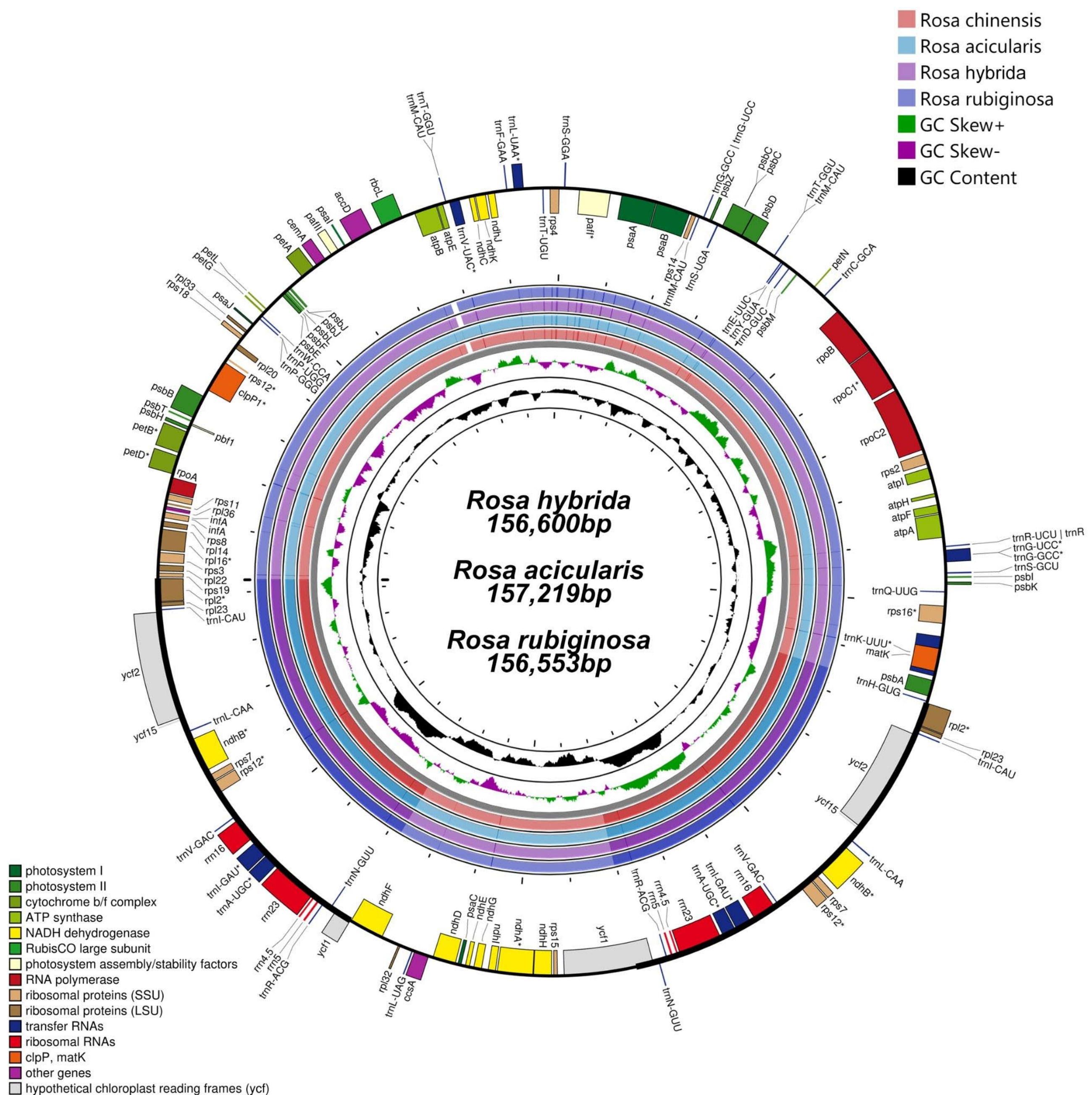

Fig. 1 Map of aligned *Rosa* chloroplast genomes. Gene map of the *Rosa* chloroplast genomes, sequence alignment of *Rosa* species chloroplast genomes with *R. rugosa* as the reference, GC skew, and GC content from outside to inside. The circular map was drawn using OGDraw

higher sequence differentiation than gene-coding regions (Figs. 1 and 3) [25]. Additionally, there were no gene rearrangements, inversions, or losses among the chloroplast genomes of the five *Rosa* species (Fig. 2b). There were some highly variable regions in the chloroplast genome sequences that were often clustered together and were referred to as "hotspots" [26]. Next, nucleotide substitution and nucleotide diversity (Pi) values for 24 *Rosa* chloroplast genomes (Table S1) were calculated to identify sequence divergence hotspots (Figs. 3 and 4). A nucleotide substitution search of 24 *Rosa* chloroplast genomes identified 3,173 (1.95%) variable sites, including 1,426 (0.88%) parsimony-informative sites. The Pi values were in the range of 0–0.016, with high values (Pi > 0.013) in the following regions: *ycf3-trnS, trnT-trnL, psbE-petL*, and *ycf1*. The hotspot regions could be used as molecular markers for differentiation in *Rosa* species.

Fig. 2 Comparative analysis of five *Rosa* chloroplast genomes. (**a**) Comparison of the borders of large single-copy, inverted repeat, and small single-copy regions among the five *Rosa* genomes. Colored boxes indicate the genes across the junctions. (**b**) *Rosa* chloroplast genome collinearity comparison plot. Local co-linear blocks (LCB) were colored to indicate regions of commonality. The histogram within each block indicates the degree of sequence similarity. The results were visualized by IRscope and Mauve

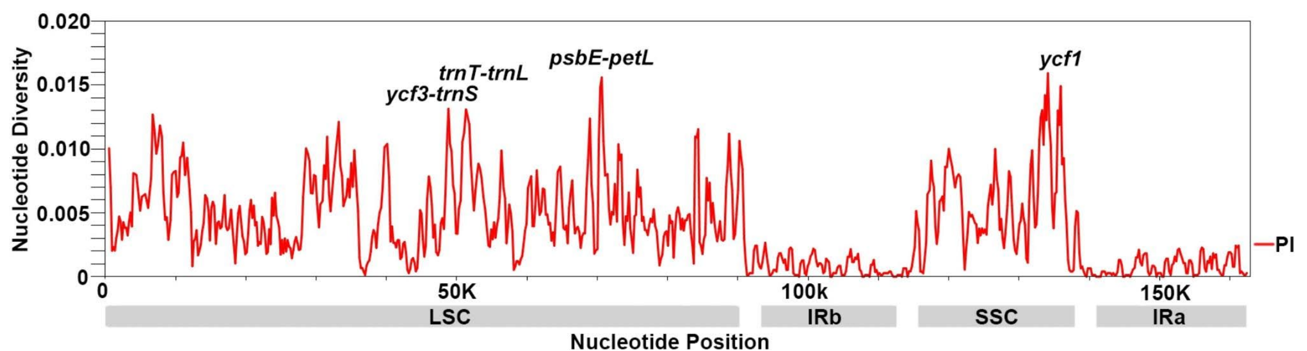

Fig. 3 Sliding window analysis of the *Rosa* chloroplast genomes using the DnaSP program. Window length: 600 bp; step size: 200 bp. X-axis, Position of a window; Y-axis, Nucleotide diversity per window

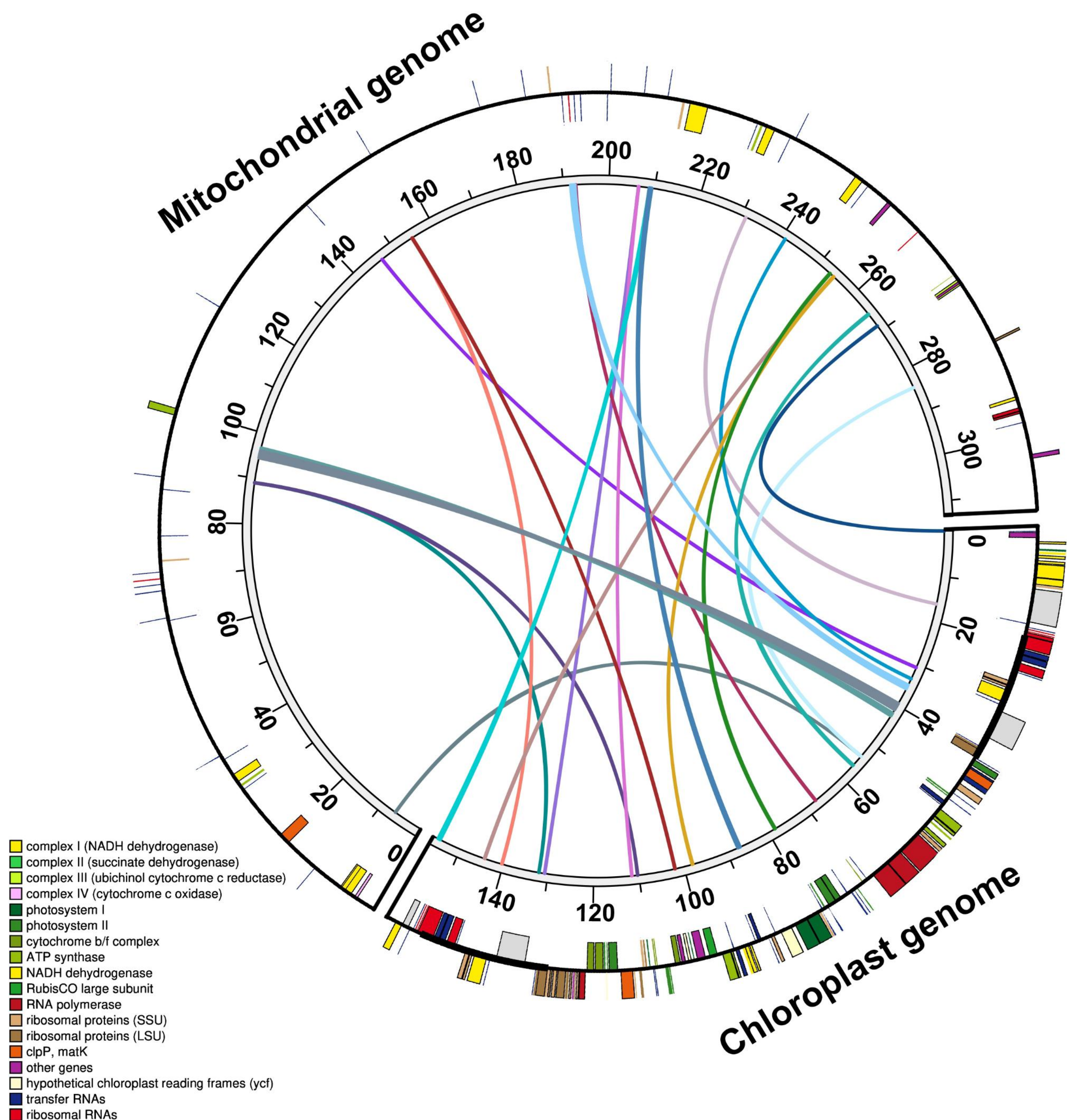

Fig. 4 Schematic diagram of gene transfer between mitochondrial and chloroplast genomes in *R. chinensis*. Colored lines within the circle show where the chloroplast genome segment entering the mitochondrial genome. Genes within a circle are transcribed clockwise, while those outside the circle are transcribed counterclockwise. The gene transfer results were visualized using Circos

Gene transfer between the chloroplast and mitochondrial genomes

The length of the mitochondrial genome sequence for *R. chinensis* in GenBank was found to be approximately twice as large as the chloroplast genome. Additionally, 22 chloroplast genomic fragments with a total length of 6,192 bp and >90% sequence similarity with their counterparts were identified in the mitochondrial genome, representing 3.96% of the chloroplast genome (Fig. 4 and Table S2). Two complete mitochondrial protein-coding genes (*psbC* and *rpl23*) and four tRNAs genes (*trnW-CCA, trnN-GUU, trnH-GUG*, and *trnM-CAU*) were identified.

Phylogenetic relationship based on chloroplast genomes

The chloroplast genomes of the 44 *Rosa* species were used to infer their phylogenetic location, except for the three newly assembled chloroplast genomes, the complete chloroplast genome sequences of 41 *Rosa* species were obtained from the National Center for Biotechnology Information (NCBI) database. Most Maximum Likelihood (ML) tree nodes had bootstrap support values of 100% (Fig. 5). Four well-supported clades (C1, C2, C3, and C4 Clade) were recovered within *Rosa*. C1 Clade included sections *Rosa*, *Carolinae*, *Hesperhodos*, and two species from section *Pimpinellifoliae, Rosa glomerata* (Sect. *Synstylae*) and *Rosa praelucens* (Subg. *Platyrhodon*) were nested in C1 Clade. C2 Clade included most samples from section *Synstylae*, all samples from sections *Bracteatae, Laevigatae, Banksianae, Chinenses, Caninae, Gallicanae*, as well as one species from subgenus *Platyrhodon* (*R. roxburghii*). The *Hulthemia* species formed C3 Clade. C4 Clade includes three species from section *Pimpinellifoliae* (*R. omeiensis*, *R. sericea*, and *R. xanthina*).

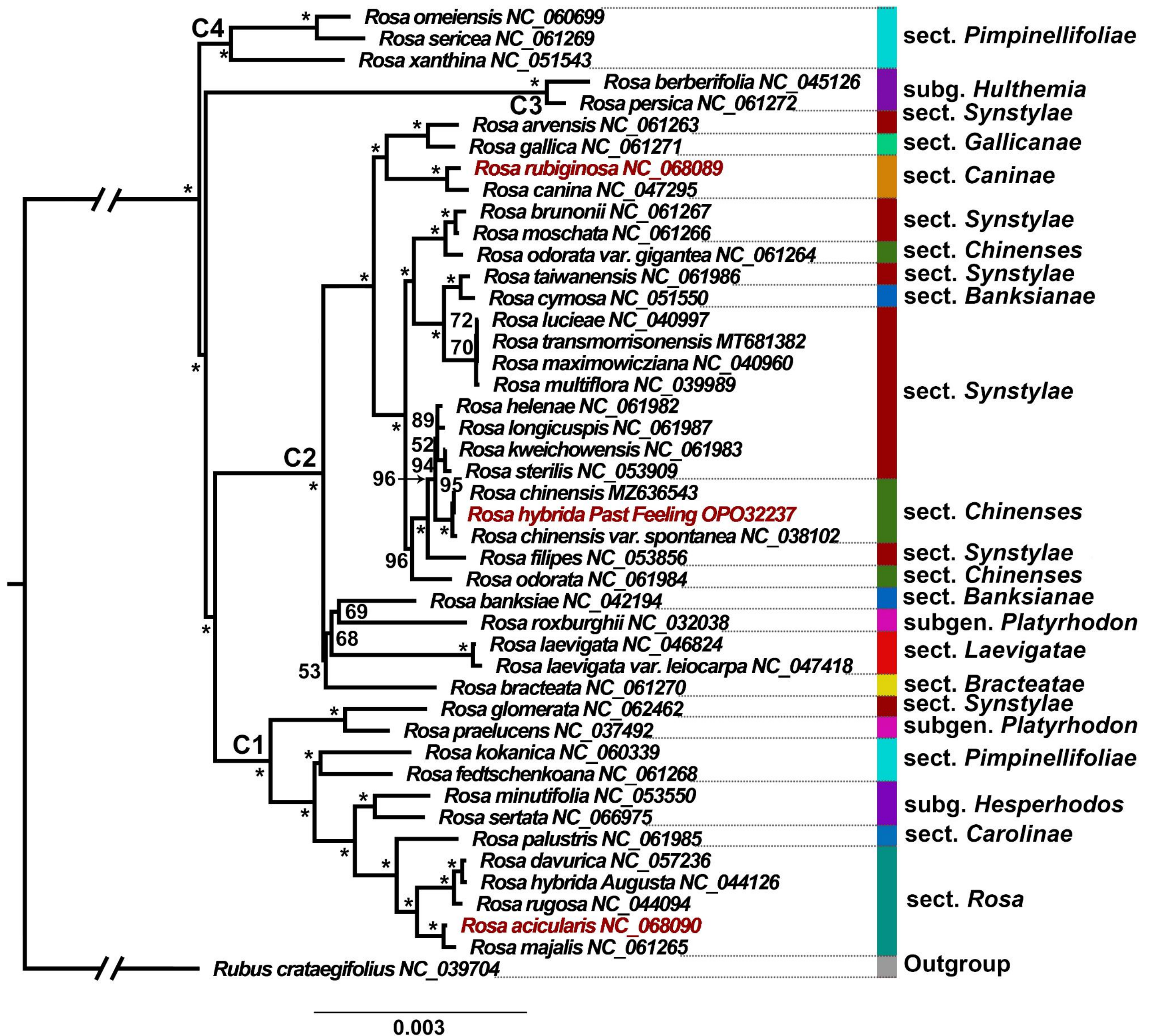

Fig. 5 Maximum Likelihood (ML) phylogenetic tree reconstruction of 44 *Rosa* species based on whole chloroplast genome sequences using IQ-TREE. The best-fit substitution model (TVM + F + I + G4) was used to build phylogenetic tree. Bootstrap resampling with 1,000 replicates was employed to assess branching support. Numbers with branches indicate ML bootstrap values, asterisk denotes 100% ML bootstrap support. *Rubus crataegifolius* was used as the outgroup. The GenBank numbers of all species are shown in the figure. Different colors correspond to the section names

Identification of RNA editing sites using RNA-Seq data

RNA editing sites in *R. hybrida* 'Past Feeling' were identified via RNA-Seq data mapping. A 99% region of the organellar transcripts were covered by reads, and the average sequencing depth was over 52x. In addition, the distribution of reads was uneven. The genome coverage maps are shown in Figure S1. Using a stringent screening procedure described in Materials and Methods (Fig. 6), we identified a total of 19 RNA editing sites in the chloroplast genome (Table 1). All of the editing sites were C-to-U conversions and were located in protein-coding regions. The 19 RNA editing sites in the chloroplast genome were distributed among 13 genes and included three synonymous and 16 nonsynonymous RNA editing sites. Most RNA editing sites occurred at the second codon position. RNA editing at the first and second codon positions resulted in amino acid conversion, whereas that at the third codon position resulted in silent changes, e.g. proline (CCC) to proline (CCU). However, silent codon changes only accounted for 15.79% of the total number of RNA editing sites in the chloroplast genome. The RNA editing efficiency ranged from 38.89 to 100% with a mean of 82.96%. Compared with the RNA editing of the *Arabidopsis* chloroplast genome [27], six conserved RNA editing sites (*rps14*-27, *rps14*-50, *accD*-264, *clpP1*-187, *rpoA*-277, and *ndhD*-128) were identified in the *R. hybrida* chloroplast genome, accounting for 31.58% of the total number of RNA editing sites.

Discussion

The chloroplast genomes of the *Rosa* species were generally consistent in terms of genomic structure, gene number, type, and order, with the exception of some single nucleotide polymorphisms (SNPs) and insertion and deletion variations [15, 23, 28, 29]. There is gene loss in the evolution of plant chloroplast genome [30], while there is a high level of conservation in the genus *Rosa* suggests evolutionary constraint in the chloroplast genome, which is prevalent in higher plants [25].

Unlike the nuclear genome, the chloroplast genome has multiple copies in the cell and is smaller in size. In addition, chloroplast genomes have sufficient interspecific differentiation. Therefore, the use of chloroplast genome sequences is one of the best approaches for species identification at present [31]. In this study, based on the results of the alignment of *Rosa* chloroplast genomes and SNP analysis, we found an increased number of variable sites in the four specific regions, namely *ycf3-trnS, trnT-trnL, psbE-petL*, and *ycf1*. Thus, using these regions as novel candidate segments may provide useful information for *Rosa* species identification. However, further experiments are needed to support these results.

Intracellular gene transfer occurs between the nucleus, mitochondria, and chloroplast [32, 33]. Gene transfer among mitochondrial and chloroplast genomes is common during the long-term evolution of plants [32, 34]. Intracellular gene transfer may be responsible for the high rearrangement of the mitochondrial genome, because the chloroplast genome segment entering the mitochondria

Table 1 RNA editing sites in the chloroplast genome of *R. hybrida* identified using RNA-sequencing data

Gene name	Editing position in genome	Editing position in gene	Editing position in codon	Editing type	Codon Change	Amino Acid Change	Coverage Depth	Editing efficiency
atpA	10,866	1263	3	C->U	CUC->CUU	L->L	9	88.89%
rps14	37,798	80	2	C->U	UCA->UUA	S->L	50	92.00%
rps14	37,729	149	2	C->U	CCA->CUA	P->L	37	97.30%
accD	59,505	791	2	C->U	UCG->UUG	S->L	18	38.89%
psaI	60,728	83	2	C->U	UCU->UUU	S->F	13	84.62%
psaI	60,733	88	1	C->U	CAU->UAU	H->Y	13	69.23%
petL	67,058	5	2	C->U	CCU->CUU	P->L	42	80.95%
rps18	69,519	221	2	C->U	CCG->CUG	P->L	27	74.07%
clpP1	71,381	559	1	C->U	CAU->UAU	H->Y	14	100.00%
psbB	74,290	414	3	C->U	AUC->AUU	I->I	117	69.23%
petB	77,760	611	2	C->U	CCA->CUA	P->L	53	73.58%
rpoA	80,192	200	2	C->U	UCU->UUU	S->F	9	77.78%
rpoA	79,562	830	2	C->U	UCA->UUA	S->L	14	71.43%
rpl16	83,113	12	3	C->U	CCC->CCU	P->P	24	75.00%
ndhD	118,698	383	2	C->U	CCA->CUA	P->L	5	100.00%
ndhA	121,876	961	1	C->U	CCU->UCU	P->S	10	100.00%
ndhA	121,764	1073	2	C->U	UCC->UUC	S->F	12	91.67%
ndhA	121,876	2053	1	C->U	CCU->UCU	P->S	10	100.00%
ndhA	121,764	2165	2	C->U	UCC->UUC	S->F	12	91.67%

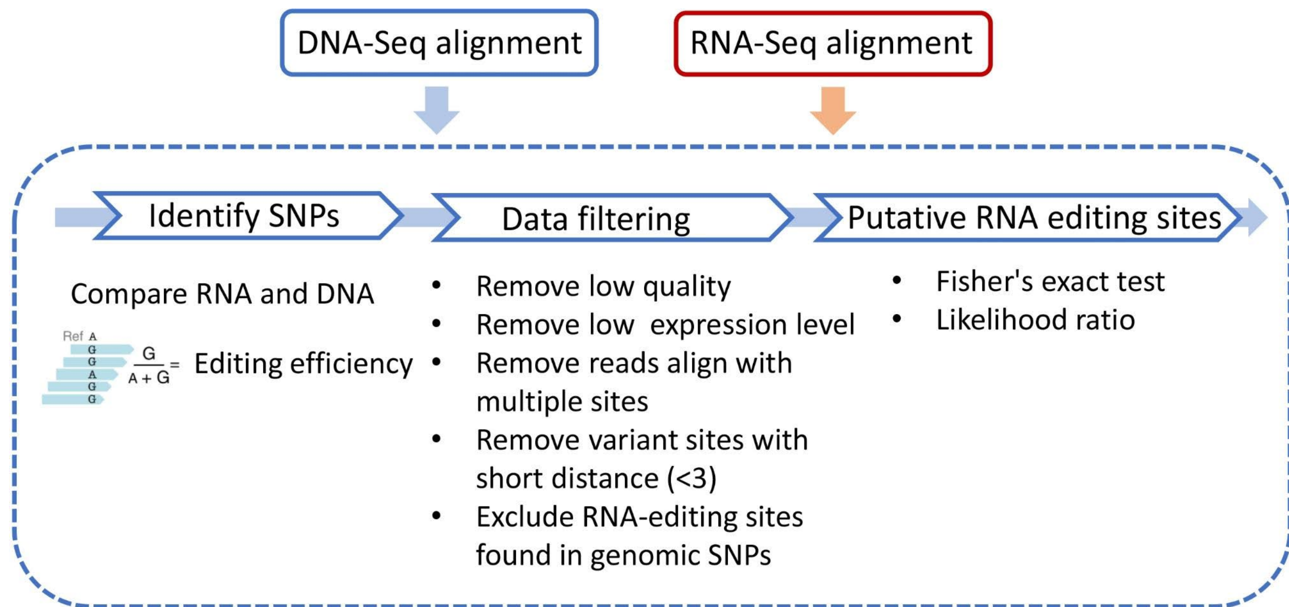

Fig. 6 Overview of the RNA editing site identification and analysis pipeline

was highly aligned with the original chloroplast genome sequences and the insertion position of the segments were randomly located [35]. The total length of these transferred fragments in Rosa mitochondrial genome was 6,192 bp, this is much shorter than the transfer fragments we found in other genera [36], this may be one of the reasons why the mitochondrial genome of Rosa is relatively small.

In this study, a phylogenetic tree based on chloroplast genome sequences was constructed to explore the evolutionary relationship in the genus *Rosa* and was found to be generally consistent with previously reported results [13, 29, 37]. There were several inconsistencies between the nuclear and chloroplast phylogenetic topology, particularly the position of section *Rosa*, which may be due to incomplete lineage, differences in the evolutionary rates of chloroplast and nuclear genes, or introgressive hybridization [37]. The earliest divergence in the chloroplast phylogeny roughly distinguished species of sections *Pimpinellifoliae* and *Rosa* and subgenera *Hulthemia*, *Platyrhodon*, and *Hesperhodos* from species of sections *Synstyale*, *Laevigatae*, *Banksianae*, *Caninae*, and *Chinenses*, which is consistent with previous studies [37, 38].

RNA editing of the *Rosa* chloroplast genome is one of the focal points of this study. As a vital post-transcriptional regulation mechanism, it has been generally accepted that 20–60 RNA editing sites are present in most chloroplast genomes [1, 39]. Previously, a software was used to predict RNA editing sites; however, its accuracy rate was generally low, and synonymous mutation sites could not be predicted. The advent of next-generation sequencing (NGS) has improved the sensitivity and accuracy of RNA editing site identification [40, 41]. In this study, similar to many plant organellar genome RNA editing studies [41–43], the data was obtained through the polyA RNA protocol. Since plant organellar transcripts generally do not have poly-A tail [44], the editing efficiency can be biased. Nonetheless, RNA-seq data obtained by polyA RNA protocol have implications in RNA editing studies of organelle genome [44]. In the present study, all editing sites found were C-to-U conversions. Furthermore, no editing sites were observed in tRNA and rRNA genes. These may be due to the stringent filtering process in our identification pipeline. Each species has its own unique RNA editing sites in comparison with other species, which indicates that RNA editing sites are independently lost after species divergence. Overall, the codon preference of targets for RNA editing, the tendency of increased protein hydrophobicity, and site distribution showed similar trends across species.

Conclusions

In conclusion, we assembled and compared the chloroplast genomes of *Rosa* species and found that the genome structure and gene content of *Rosa* chloroplast genomes are similar across various species. We also identified 22 chloroplast fragments in the mitochondrial genome. Phylogenetic analysis based on the *Rosa* chloroplast genomes has high resolution. Additionally, a total of 19 RNA editing sites in 13 genes were validated by RNA-Seq mapping in *R. hybrida*. The findings of this study provide valuable genetic resources for further research on *Rosa* species.

Materials and methods

Plant material and sequencing

The *Rosa* accessions were from the *Rosa* nuclear genome and transcriptome sequencing projects (Table S3). Total genomic DNA was extracted from herbarium (*R. acicularis, R. rubiginosa*) or petals (*R. hybrida*) using the CTAB method. The voucher specimens of *R. acicularis* (TROM_V_91069) and *R. rubiginosa* (TROM_V_148853) and leaves were used for DNA extraction. Petals were provided by Kunming Yangyueji Company. Paired−end (2×100 bp) genomic libraries were constructed using Illumina kit for sequencing on BGISEQ−500 and Illumina hiseq 2500 sequencers with an average insertion size of 300 bp. Total RNA was extracted from petals using the SV total RNA Isolation Kit (Promega, WI, USA). The method of rRNA depletion is poly-A selection, which relies on the use of Oligo (dT)-attached magnetic beads to isolate protein-coding polyadenylated RNA transcripts. A NEBNext® UltraTM RNA Library Prep Kit (New England Biolabs, MA, USA) was used to generate libraries and sequenced on an Illumina HiSeqTM 2000 instrument at Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). The raw chloroplast genomes and transcriptome sequencing data were uploaded in the NCBI sequence read archive with accession numbers SRR21561260–SRR21561263.

Chloroplast genome assembly and annotation

Raw sequencing data were filtered using Trimmomatic v0.38 [45]. *De novo* assembly was then performed using SPAdes version 3.61 with different k-mer parameters [46]. Next, the Geneious Prime software v2022.2 [47] was used to order *de novo* scaffolds that were positively correlated with chloroplasts on to the reference chloroplast genome of *R. rugosa* (NC_044094).

GeSeq was used to perform chloroplast genome annotation to predict gene-coding proteins, rRNAs, and tRNAs, with manual curation as needed [48]. Subsequently, the circular map of the *Rosa* chloroplast genome was drawn using OGDraw v1.3.1 [49].

Genome comparative analysis and hotspots regions screening

Rosa chloroplast genome sequences were aligned using MAFFT v7.221 [50]. Comparison of the borders of LSC, IR and SSC regions among the five *Rosa* genomes (OP032236, OP032237, OP032238, MK986659, and NC_038102) was visualized by IRscope [51]. The Mauve multiple genome alignment method was used to detect rearrangements and co-linearities in the chloroplast genomes of the five *Rosa* species [52]. To examine the rapidly evolving molecular markers among *Rosa* species, we used 24 *Rosa* chloroplast genomes (Table S1) for the sliding window analysis with a window size of 600 bp and a step length of 200 bp using DnaSP v6.12 [53].

Identification of chloroplast gene insertion in mitochondria

The mitochondrial and chloroplast genomes of *R. chinensis* were retrieved from GenBank (CM009589 and CM009590, respectively). The genes transferred between the mitochondrial and chloroplast genomes were then identified via homology searches using Basic Local Alignment Search Tool. Chloroplast and mitochondrial maps of *Rosa* and fragments of gene transfer were visualized using Circos [54].

Phylogenetic analysis

Phylogenetic trees were constructed using the whole chloroplast genome sequences of 44 *Rosa* species to identify their genetic relationship. *Rubus crataegifolius* was used as the outgroup. Genome sequences were aligned using MAFFT v7.221 [50], and all alignments were manually inspected and adjusted. IQ-TREE v 1.6.12 [55]was used to build an ML phylogenetic tree with the best-fit substitution model (TVM+F+I+G4) determined by ModelFinder v3.7 [56]. Bootstrap resampling with 1,000 replicates was employed to assess branching support.

Identification of RNA editing sites using RNA-Seq data

The clean RNA-Seq reads were aligned to the chloroplast genome of *R. hybrida* 'Past Feeling' using the Hisat2 v2.1.0 tool [57]. To convert sequence alignment map to binary alignment map, the samtools v1.9 view command was used [58]. Potential RNA editing sites were extracted using the SNP calling method in bcftools v1.9 [58]. Extracted SNPs were then processed with REDO v1.0 to provide annotation information for editing sites [59]. To eliminate the false positive RNA editing sites, DNA-Seq reads of *R. hybrida* 'Past Feeling' were aligned to the chloroplast genome using Bowtie 2 v2.3.5 [60]. Genomic SNP-calling was performed using bcftools v1.9 [58]. RNA editing sites that were found in genomic SNPs were then excluded (Fig. 6).

Abbreviations

RNA-Seq	RNA-sequencing
SSC	Small single copy
LSC	Large single copy
IR	Inverted repeat
tRNA	Transfer RNA
rRNA	Ribosomal RNA
ML	Maximum Likelihood
SNPs	Single nucleotide polymorphisms
NGS	Next-generation sequencing

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-023-04338-0.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
We would like to thank Editage (www.editage.com) for English language editing. We finally thank the anonymous reviewers for reviewing the manuscript and providing valuable comments.

Authors' contributions
ZQL and CWG conceived the study and acquired the funding. CWG and TL performed the data analyses and drafted the earlier version of manuscript. XZ, CHW, QZ, XZZ, MXW, and YHL participate in project management and assistance, all authors approved the final manuscript.

Funding
This work was supported by the Natural Science Foundation of Shandong Province (No. ZR2020QH321).

Data availability
The data supporting the findings of this study are freely available in GenBank on the NCBI website at https://www.ncbi.nlm.nih.gov, using the accession number OP032236, OP032237, and OP032238. Raw sequencing data have been deposited at the NCBI Sequence Read Archive (SRA) under accession SRR21561260–SRR21561263.

Declarations

Ethics approval and consent to participate
The authors confirm that all methods comply with local and national regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 8 October 2022 / Accepted: 8 June 2023
Published online: 14 June 2023

References

1. Small ID, Schallenberg-Rudinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. Plant J. 2020;101(5):1040–56.
2. Tsudzuki T, Wakasugi T, Sugiura M. Comparative analysis of RNA editing sites in higher plant chloroplasts. J Mol Evol. 2001;53(4–5):327–32.
3. Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T, Yoshinaga K. The complete nucleotide sequence of the hornwort (Anthoceros formosae) chloroplast genome: insight into the earliest land plants. Nucleic Acids Res. 2003;31(2):716–21.
4. Chu D, Wei L. The chloroplast and mitochondrial C-to-U RNA editing in Arabidopsis thaliana shows signals of adaptation. Plant Direct 2019, 3(9).
5. Ichinose M, Sugita M. RNA editing and its molecular mechanism in Plant Organelles. Genes 2017, 8(1).
6. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW, et al. Horizontal transfer of entire genomes via mitochondrial Fusion in the Angiosperm Amborella. Science. 2013;342(6165):1468–73.
7. Ishibashi K, Small I, Shikanai T. Evolutionary model of Plastidial RNA editing in Angiosperms presumed from genome-wide analysis of Amborella trichopoda. Plant Cell Physiol. 2019;60(10):2141–51.
8. Freyer R, KieferMeyer MC, Kossel H. Occurrence of plastid RNA editing in all major lineages of land plants. Proc Natl Acad Sci USA. 1997;94(12):6285–90.
9. Gerke P, Szovenyi P, Neubauer A, Lenz H, Gutmann B, McDowell R, Small I, Schallenberg-Rudinger M, Knoop V. Towards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort Anthoceros agrestis. New Phytol. 2020;225(5):1974–92.
10. Barkan A, Small I. Pentatricopeptide Repeat Proteins in Plants. In: *Annual Review of Plant Biology, Vol 65* Edited by Merchant SS, vol. 65; 2014: 415-+.
11. 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.
12. Ayati Z, Ramezani M, Amiri MS, Sahebkar A, Emami SA. Genus Rosa: A Review of Ethnobotany, Phytochemistry and Traditional Aspects According to Islamic Traditional Medicine (ITM). In: *Pharmacological Properties of Plant-Derived Natural Products and Implications for Human Health* Edited by Barreto GE, Sahebkar A, vol. 1308; 2021: 353–401.
13. Zhu ZM, Gao XF, Fougere-Danezan M. Phylogeny of Rosa sections Chinenses and Synstylae (Rosaceae) based on chloroplast and nuclear markers. Mol Phylogenet Evol. 2015;87:50–64.
14. AlfredRehder. Manual of cultivated trees and shrubs. Manual of cultivated trees and shrubs; 1951.
15. 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).
16. Brunkard JO, Runkel AM, Zambryski PC. Chloroplasts extend stromules independently and in response to internal redox signals. Proc Natl Acad Sci USA. 2015;112(32):10044–9.
17. Douglas SE. Plastid evolution: origins, diversity, trends. Curr Opin Genet Dev. 1998;8(6):655–61.
18. Hu H, Hu QJ, Al-Shehbaz IA, Luo X, Zeng TT, Guo XY, Liu JQ. Species Delimitation and Interspecific Relationships of the Genus Orychophragmus (Brassicaceae) inferred from whole chloroplast genomes. Front Plant Sci. 2016;7:10.
19. Xu WQ, Losh J, Chen C, Li P, Wang RH, Zhao YP, Qiu YX, Fu CX. Comparative genomics of figworts (Scrophularia, Scrophulariaceae), with implications for the evolution of Scrophularia and Lamiales. J Syst Evol. 2019;57(1):55–65.
20. Huang J, Yu Y, Liu YM, Xie DF, He XJ, Zhou SD. Comparative Chloroplast Genomics of Fritillaria (Liliaceae), Inferences for phylogenetic Relationships between Fritillaria and Lilium and Plastome Evolution. Plants-Basel. 2020;9(2):15.
21. Olennikov DN, Chemposov VV, Chirikova NK. Metabolites of Prickly Rose: Chemodiversity and Digestive-Enzyme-inhibiting potential of Rosa acicularis and the Main Ellagitannin Rugosin D. Plants-Basel 2021, 10(11).
22. Jimenez-Lopez J, Ruiz-Medina A, Ortega-Barrales P, Llorent-Martinez EJ. Rosa rubiginosa and Fraxinus oxycarpa herbal teas: characterization of phytochemical profiles by liquid chromatography-mass spectrometry, and evaluation of the antioxidant activity. New J Chem. 2017;41(15):7681–8.
23. 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).
24. 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).
25. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 2016, 17.
26. Liu L, Wang Y, He P, Li P, Lee J, Soltis DE, Fu C. Chloroplast genome analyses and genomic resource development for epilithic sister genera Oresitrophe and Mukdenia (Saxifragaceae), using genome skimming data. BMC Genomics. 2018;19(1):235.
27. Ruwe H, Castandet B, Schmitz-Linneweber C, Stern DB. Arabidopsis chloroplast quantitative editotype. FEBS Lett. 2013;587(9):1429–33.
28. 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(1).
29. 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.
30. Mohanta TK, Mishra AK, Khan A, Hashem A, Abd Allah EF, Al-Harrasi A. Gene loss and evolution of the Plastome. Genes 2020, 11(10).
31. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLoS ONE. 2011;6(5):13.
32. Nguyen VB, Giang VNL, Waminal NE, Park HS, Kim NH, Jang W, Lee J, Yang TJ. Comprehensive comparative analysis of chloroplast genomes from seven Panax species and development of an authentication system based on species-unique single nucleotide polymorphism markers. J Ginseng Res. 2020;44(1):135–44.

33. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 2004;5(2):123–U116.
34. Gui ST, Wu ZH, Zhang HY, Zheng YZ, Zhu ZX, Liang DQ, Ding Y. The mitochondrial genome map of Nelumbo nucifera reveals ancient evolutionary features. Sci Rep. 2016;6:11.
35. Shidhi PR, Biju VC, Anu S, Vipin CL, Deelip KR, Achuthsankar SN. Genome characterization, comparison and phylogenetic analysis of complete mitochondrial genome of Evolvulus alsinoides reveals highly rearranged Gene Order in Solanales. Life-Basel 2021, 11(8).
36. Gao C, Wu C, Zhang Q, Zhao X, Wu M, Chen R, Zhao Y, Li Z. Characterization of Chloroplast Genomes from two Salvia Medicinal plants and gene transfer among their mitochondrial and chloroplast genomes. Front Genet. 2020;11:574962.
37. Debray K, Le Paslier MC, Berard A, Thouroude T, Michel G, Marie-Magdelaine J, Bruneau A, Foucher F, Malecot V. Unveiling the patterns of reticulated evolutionary processes with Phylogenomics: hybridization and polyploidy in the Genus Rosa. Systemic Biology. 2022;71(3):547–69.
38. Fougere-Danezan M, Joly S, Bruneau A, Gao XF, Zhang LB. Phylogeny and biogeography of wild roses with specific attention to polyploids. Ann Botany. 2015;115(2):275–91.
39. Kugita M, Yamamoto Y, Fujikawa T, Matsumoto T, Yoshinaga K. RNA editing in hornwort chloroplasts makes more than half the genes functional. Nucleic Acids Res. 2003;31(9):2417–23.
40. Hao W, Liu GX, Wang WP, Shen W, Zhao YP, Sun JL, Yang QY, Zhang YX, Fan WJ, Pei SS et al. RNA editing and its roles in Plant Organelles. Front Genet 2021, 12.
41. Wang S, Yang CP, Zhao XY, Chen S, Qu GZ. Complete chloroplast genome sequence of Betula platyphylla: gene organization, RNA editing, and comparative and phylogenetic analyses. BMC Genomics 2018, 19.
42. Wu B, Chen HM, Shao JJ, Zhang H, Wu K, Liu C. Identification of symmetrical RNA editing events in the Mitochondria of Salvia miltiorrhiza by strand-specific RNA sequencing. Sci Rep. 2017;7:11.
43. Fang B, Li JL, Zhao Q, Liang YP, Yu J. Assembly of the complete mitochondrial genome of Chinese Plum (Prunus salicina): characterization of genome recombination and RNA editing Sites. Genes 2021, 12(12).
44. He ZS, Zhu AD, Yang JB, Fan WS, Li DZ. Organelle genomes and transcriptomes of Nymphaea reveal the interplay between Intron Splicing and RNA editing. Int J Mol Sci 2021, 22(18).
45. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
46. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a New Genome Assembly Algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
47. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
48. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–W11.
49. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–W64.
50. Katoh K, Standley DM. MAFFT multiple sequence alignment Software Version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
51. Amiryousefi A, Hyvonen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1.
52. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft genomes using the Mauve Aligner. Bioinformatics. 2009;25(16):2071–3.
53. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.
54. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.
55. 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.
56. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–.
57. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–.
58. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM et al. Twelve years of SAMtools and BCFtools. *GigaScience* 2021, 10(2).
59. Wu SY, Liu WF, Aljohi HA, Alromaih SA, AlAnazi IO, Lin Q, Yu J, Hu SN. REDO: RNA editing detection in Plant Organelles based on variant calling results. J Comput Biol. 2018;25(5):509–16.
60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–U354.

Publisher's Note

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