Physiology and Molecular Biology of Plants (June 2023) 29(6):855–869
https://doi.org/10.1007/s12298-023-01331-7

RESEARCH ARTICLE

Uncovering the first complete chloroplast genomics, comparative analysis, and phylogenetic relationships of the medicinal plants *Rhamnus cathartica* and *Frangula alnus* (*Rhamnaceae*)

Wenbo Shi[1] · Siqi Hu[1] · Weicai Song[1] · Yahui Huang[1] · Chao Shi[1,2] · Shuo Wang[1]

Received: 2 April 2023 / Revised: 4 June 2023 / Accepted: 5 July 2023 / Published online: 15 July 2023

Abstract

Rhamnus cathartica and *Frangula alnus* are economically valuable medicinal plants from the *Rhamnaceae* family. However, their chloroplast genome structure, phylogenetic position, relationships, and evolution remain poorly understood. Herein, the complete chloroplast genome resources of *R. cathartica* and *F. alnus* have been added. The first comparative analysis of the *Rhamnus* and *Frangula* species based on complete chloroplast genomes was provided. The chloroplast genomes of *R. cathartica* and *F. alnus* exhibited a quadripartite structure, with total lengths of 161,149 bp and 161,255 bp, respectively. The lack of the *infA* and *psbL* genes does not negatively impact the normal functioning of *Rhamnus* and *Frangula* species. The *rpl20* and *rpl33* genes are undergoing rapid evolution. *Rhamnus* and *Frangula* species prefer amino acids with A/U-terminal codons. There were between 100 and 126 simple sequence repeats and between 38 and 100 long repeats. Several highly divergent intergenic regions (*trnK*-UUU–*trnQ*-UUG, *atpH–atpI*, *trnY*-GUA–*trnE*-UUC, *trnG*-GCC–*trnfM*-CAU, *trnT*-UGU–*trnF*-GAA, *rpl20–rps12*, and *rpl22–rps19*) and highly divergent genes (*ycf3*, *ndhA*, *rpl32*, and *ycf1*) were identified, which could serve as potential phylogenetic markers due to their variability. We reconstructed the phylogenetic relationships among *Rhamnus* species and *F. alnus* using complete chloroplast genomes. There is no significant correlation between the medicinal value of the species analyzed and their phylogenetic relationships. These results provide valuable insights for understanding the phylogenetic relationship and evolution of *Rhamnus* and *Frangula* species. These findings could serve as a foundation for future studies on the *Rhamnaceae*.

Keywords *Rhamnus* · *Frangula alnus* · Chloroplast genome · Comparative analysis · Divergence · Phylogenetic analysis

Introduction

The *Rhamnaceae* family is a cosmopolitan group of flowering plants that includes approximately 55 genera and 950 species. Most *Rhamnaceae* plants are trees, but there are also shrubs, lianas, and herbs (Richardson et al. 2004; Christenhusz and Byng 2016). The *Rhamnaceae* family is rich in morphological and genetic diversity, making it an ideal subject for study into flowering plant origin and evolution (Kang et al. 2019). Furthermore, the earliest known fossils of *Rhamnaceae* plants (*Phylica*) are from the Cretaceous period (Shi et al. 2022). This discovery shifts the focus of the study of the rapid diversification of flowering plants to *Rhamnaceae* plants (He and Lamont 2022). In the *Rhamnaceae* family, the *Rhamnus* genus is a representative member (Richardson et al. 2000). *Rhamnus* species are found mainly in Asia, Europe, and Africa (Hauenschild et al. 2016). They are shrubs or trees with medicinal, ornamental, and economic value (especially *R. cathartica*) (Whitfeld et al. 2014; Nigussie et al. 2021). Therefore, it is very meaningful to study the *Rhamnus* species. The *Rhamnus* genus has been divided into two subgenera, *Rhamnus* and *Frangula*, based on their morphological characteristics (Richardson et al. 2000). Grubov (1949) proposed that *Frangula*

✉ Chao Shi
chsh1111@aliyun.com

✉ Shuo Wang
shuowang@qust.edu.cn

[1] College of Marine Science and Biological Engineering, Qingdao University of Science and Technology, Qingdao 266042, China

[2] Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, The Chinese Academy of Sciences, Kunming 650204, China

be classified as a separate genus based on the number of petals present in the hermaphroditic flower. Grubov's suggestion is supported by subsequent molecular-based phylogenetic analyses, which indicate that the *Frangula* genus is monophyletic (Bolmgren and Oxelman 2004; Hauenschild et al. 2016). Recently, phylogenetic genomic analyses using the complete chloroplast genome have emerged (Song et al. 2022; Shi et al. 2023a). The comparative analysis of chloroplast genomes may help us better understand the relationship between *Rhamnus* and *Frangula* species.

The chloroplast is one of the essential cellular organelles in plants, serving as an energy converter (Kirchhoff 2019). In addition, chloroplasts participate in metabolic pathways such as starch storage, sugar biosynthesis, and lipid production (Nielsen et al. 2016). The chloroplast genome is usually a circular, double-helix structure that ranges in size from 120 to 180 kb (Wicke et al. 2011). A large single-copy region (LSC), a small single-copy region (SSC), and two inverted repeat regions (IRs: IRa and IRb) make up its quadripartite structure (Wicke et al. 2011). The gene content, number, and structure of the plant chloroplast genomes are relatively conserved (Zurawski and Clegg 1987). Nevertheless, gene loss, pseudogenization, and genome size changes remain present in chloroplast genomes (Henriquez et al. 2020). These characteristics can be used to understand interspecific relationships and the evolution of plants through comparative, phylogenetic, and evolutionary studies (Jheng et al. 2012; Shi et al. 2023b). The advancement of high-throughput sequencing techniques has made it possible to sequence the complete chloroplast genomes of plant species (Dodsworth 2015). This has resulted in the widespread use of chloroplast genomes in plant evolutionary studies to improve phylogenetic resolution at the genus and species levels (Guo et al. 2023; Wanichthanarak et al. 2023).

To date, the phylogenetic analysis of the *Rhamnus* genus has focused mainly on molecular fragments such as *trnL–trnF* and internal transcribed spacer (ITS) regions (Bolmgren and Oxelman 2004; Hauenschild et al. 2016), resulting in advancements in the taxonomy of the *Rhamnus* genus. However, due to the absence of polymorphic sites, molecular fragments might not accurately reflect the entire genome (Guo et al. 2023). Additionally, phylogenetic trees constructed with these molecular fragments usually have nodes with poor bootstrap values. The complete chloroplast genome sequence provides sufficient informative sites, making it better usable for phylogenetic analyses (Zhang et al. 2022; Shi et al. 2023b). Herein, we sequenced the complete chloroplast genome of *R. cathartica*, an economically valuable species. Furthermore, we present the first complete chloroplast genome of the *Frangula* species (*F. alnus*). The complete chloroplast genomes of 11 *Rhamnus* (*R. crenata*, *R. davurica*, *R. diamantiaca*, *R. globosa*, *R. heterophylla*, *R. lamprophylla*, *R. leptophylla*, *R. subapetala*, *R. taquetii*, *R. ussuriensis*, and *R. wilsonii*) species were downloaded from the NCBI database for comparative analysis and phylogeny. The goals of this study were to (1) investigate the divergence of the chloroplast genomes of *Rhamnus* and *Frangula* species and (2) elucidate the phylogenetic relationships between *Rhamnus* species and *Frangula alnus* based on the complete chloroplast genome.

Materials and methods

Plant materials, DNA extraction from samples, and sequencing

To enrich the complete chloroplast genome resource. *Rhamnus cathartica* and *Frangula alnus*, two plants with a distribution in Xinjiang Province, China, were chosen (Gassmann et al. 2008). These two species have some medicinal value and should be studied further (Kremer et al. 2012; Whitfeld et al. 2014). The plant materials of *R. cathartica* and *F. alnus* were collected from the Plant Germplasm and Genomics Center, Kunming Institute of Botany, and the Chinese Academy of Sciences. Two specimens of each species were chosen in accordance with the pertinent laws and regulations. The authors identified the botanical specimens and subsequently deposited them at the Qingdao University of Science and Technology. Chloroplast DNA was isolated from mature leaf samples of *R. cathartica* and *F. alnus* using a modified high-salinity approach, as previously reported (Shi et al. 2012). The genomic DNA isolated was quantified using the dsDNA HS kit on a Qubit instrument (Invitrogen, Carlsbad, California, USA) based on fluorescence. Additionally, agarose gel electrophoresis (Green and Sambrook 1972) with a concentration of 1% was performed to evaluate the quality of the isolated DNA from *R. cathartica* and *F. alnus*. Sequencing of the chloroplast genomes of *R. cathartica* and *F. alnus* was carried out on the Illumina Hiseq 4000 platform (Illumina, San Diego, CA, USA) in accordance with the manufacturer's protocol using high-quality DNA that was provided to Novogene (Beijing, China). Two sets of 151 bp paired-end raw reads have been acquired at the end.

Chloroplast genome de novo assembly and annotation

Trimmomatic v0.40 (Bolger et al. 2014) was used to preprocess the paired-end raw reads, including the removal of low-quality and over-N-base reads. Subsequently, the high quality of newly generated clean short reads was assessed via FastQC v0.11.9 (Brown et al. 2017). High quality short reads with average Phred scores above 35 were screened for the assembly of the chloroplast genomes. The complete chloroplast genomes of *R. cathartica* and *F. alnus* were

assembled through NOVOplasty v4.3.1 (Dierckxsens et al. 2017). The *rbcL* genes (KM360955 and KM360790) of *R. cathartica* and *F. alnus* were used as seed sequences. To produce the best assembly results, various k-mer values were modified. To assess the assembly outcomes, Minimap v2.17 (Li 2018) was employed to map raw reads to the assembled chloroplast genome sequences. Chloroplast genomes that displayed consistent coverage and good continuity were regarded as the best assembly results. For the annotation of the assembled chloroplast genomes of *R. cathartica* and *F. alnus*, GeSeq v1.42 (Tillich et al. 2017) and CPGAVAS v2 (Shi et al. 2019) were used. The results produced from the annotations were compared, and the best results were chosen. Finally, the codons and gene boundaries inside the chloroplast genomes were artificially corrected using GB2Sequin v16 (Lehwark and Greiner 2018). In addition, we used the same approach to annotate the complete chloroplast genomes of eight *Rhamnus* species (*R. crenata* (MW801148), *R. davurica* (ON881434), *R. diamantiaca* (ON881431), *R. lamprophylla* (ON881513), *R. leptophylla* (MW800935), *R. subapetala* (ON881484), *R. ussuriensis* (ON881419), and *R. wilsonii* (MW801162)) tagged as unverified in the NCBI database. The de novo assembled chloroplast genomes of *R. cathartica* and *F. alnus* have been submitted to GenBank and assigned the accession numbers NC_068506 and NC_068507.

Statistical analysis of characteristics in chloroplast genomes

Geneious v9.0.2 (Kearse et al. 2012) was applied to perform a comparative analysis of the fundamental characteristics of the chloroplast genomes in various *Rhamnus* species and *Frangula alnus*. The length of the sequences in each region, the proportion of distinct chloroplast sequences, and the percentage of GC in various regions were all determined as part of the investigation. The annotated Genbank files for these species were used to identify the type and number of genes present. Chloroplot v0.2.4 (Zheng et al. 2020) was used to make a circular map that visually displayed the chloroplast genomes.

Structural analysis of chloroplast genomes

Based on respective Genbank files, CPJSdraw (Li et al. 2023) was employed to analyze and display the expansion and contraction of the IR regions in the chloroplast genomes of *Rhamnus* species and *Frangula alnus*. MISA-web (Beier et al. 2017) was used to identify simple sequence repeats (SSRs) in the chloroplast genomes. For mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, the fundamental parameters were set to 10, 5, 4, 3, 3, and 3. The REPuter v1 (Kurtz et al. 2001) was used to examine repeats that were either forward (F), palindromic (P), reverse (R), or complement (C), with a hamming distance of 3, a minimum repeat length of 30 bp, and a maximum repeat length of 100 bp.

Evolutionary analysis of chloroplast genomes

This study's evolutionary analysis of the chloroplast genomes included selection pressure and codon preference analyses. To begin, the protein-coding genes should be extracted from the complete chloroplast genomes using Geneious v9.0.2. The protein-coding genes of several *Rhamnus* species and *Frangula alnus* were aligned using MACSE v2 (Ranwez et al. 2018). MACSE v2 is well suited for coding sequence alignment. Then, *R. crenata* (MW801148) was used as the reference species to calculate the ratio of non-synonymous (Ka) to synonymous (Ks) substitution rates by the KaKs_Calculator v2 (Wang et al. 2010). On the online platform Chiplot (https://www.chiplot.online), a clustering heat map that visualizes Ka/Ks values was created. The value of the relative synonymous codon usage (RSCU) can be used to demonstrate genomic codon usage preferences. MEGA v11 (Tamura et al. 2021) was used to calculate the RSCU values in the protein-coding gene sequences of the chloroplast genomes. For RSCU value visualization, Tbtools v1.1 (Chen et al. 2020) was utilized.

Chloroplast genome comparison

Using *R. crenata* (MW801148) as the reference, a comparison of the chloroplast genomes of *Rhamnus* species and *Frangula alnus* was performed using mVISTA (Frazer et al. 2004). The LAGAN approach was used in the alignment program. Sequence alignment is necessary in the early stages of the analysis of nucleotide variability (Pi). The complete chloroplast genomes of *Rhamnus* species and *Frangula alnus* were aligned using MAFFT v7.308 (Katoh and Standley 2013). DnaSP v6.12 (Rozas et al. 2017) was used to evaluate the Pi values of relevant species' chloroplast genome sequences. The sliding window was set up using an 800 bp configuration and a 200 bp step size.

Phylogenetic analysis

For phylogenetic analysis, we downloaded the complete chloroplast genomes of 11 *Rhamnus* species (Table 1) from the NCBI database in addition to the newly assembled complete chloroplast genomes of *R. cathartica* and *Frangula alnus*. Two *Berchemia* species, *B. berchemiifolia* (NC_037477) and *B. racemose* (NC_066653), were used as outgroups. The *Berchemia* genus is closely related to *Frangula* and *Rhamnus* (Richardson et al. 2000; Wanichthanarak et al. 2023), and *Berchemia* species have frequently been used as outgroups in phylogenetic analyses of these two

Table 1 Summary of complete chloroplast genomes of *Frangula alnus* and various *Rhamnus* species

Genome features	Accession number	Length (bp)				Gene number				GC cotent (%)			
		Full	LSC	SSC	IR	Full[unique]	PCG[unique]	tRNA[unique]	rRNA[unique]	Full	LSC	SSC	IR
Frangula alnus	NC_068506	1,61,255	89,266	19,153	26,418	130 (112)	85 (78)	37 (30)	8 (4)	37.0	34.8	31.1	42.8
Rhamnus cathartica	NC_068507	1,61,149	89,755	18,887	26,248	129 (111)	84 (77)	37 (30)	8 (4)	37.1	35.0	31.4	42.9
Rhamnus crenata	MW801148	1,60,913	89,151	19,046	26,358	131 (113)	86 (79)	37 (30)	8 (4)	36.9	34.7	31.0	42.9
Rhamnus davurica	ON881434	1,60,575	88,777	18,902	26,448	131 (113)	86 (79)	37 (30)	8 (4)	37.2	35.1	31.4	42.7
Rhamnus diamantiaca	ON881431	1,61,039	89,310	18,893	26,418	131 (113)	86 (79)	37 (30)	8 (4)	37.1	35.0	31.4	42.8
Rhamnus globosa	NC_057506	1,60,642	88,889	18,897	26,428	130 (111)	84 (77)	37 (30)	8 (4)	37.1	35.0	31.4	42.8
Rhamnus heterophylla	NC_057481	1,56,514	86,023	19,879	25,306	130 (112)	84 (78)	37 (30)	8 (4)	37.3	35.3	31.7	42.9
Rhamnus lamprophylla	ON881513	1,61,109	89,371	18,902	26,418	131 (113)	86 (79)	37 (30)	8 (4)	37.1	35.0	31.4	42.8
Rhamnus leptophylla	MW800935	1,61,124	89,379	18,909	26,418	131 (113)	86 (79)	37 (30)	8 (4)	37.1	35.0	31.4	42.8
Rhamnus subapetala	ON881484	1,61,426	89,570	18,990	26,433	131 (113)	86 (79)	37 (30)	8 (4)	37.1	35.0	31.4	42.7
Rhamnus taquetii	NC_045855	1,61,205	89,373	18,936	26,448	129 (112)	84 (78)	37 (30)	8 (4)	37.1	35.0	31.4	42.7
Rhamnus ussuriensis	ON881419	1,61,091	89,344	18,901	26,423	131 (113)	86 (79)	37 (30)	8 (4)	37.1	35.0	31.4	42.8
Rhamnus wilsonii	MW801162	1,61,116	89,380	18,900	26,418	131 (113)	86 (79)	37 (30)	8 (4)	37.1	35.0	31.4	42.8

genera in other studies (Bolmgren and Oxelman 2004; Hauenschild et al. 2016). To increase the robustness of the phylogenetic analysis results, we constructed phylogenetic trees using the maximum likelihood (ML) and Bayesian inference (BI) methods, respectively. All complete chloroplast genome sequences were aligned using MAFFT v7.308 to develop the phylogenetic trees. The most suitable evolutionary model was selected using JModelTest v2.1.10 (Darriba et al. 2012) based on the Bayesian Information Criterion (BIC). The TVM + F + I + G4 and TVM + G models were chosen as the best models for the ML and BI trees based on the results. The ML tree was constructed using IQ-TREE v2.1.3 (Trifinopoulos et al. 2016). Each branch node was given 1000 bootstrap replications, while the other settings remained at their default values. MrBayes v3.2.7 (Huelsenbeck and Ronquist 2001) was utilized to generate a BI tree, employing Markov Chain Monte Carlo (MCMC) for 2,000,000 iterations. Every 5000 generations, the tree was sampled, and only the top 25% of trees were regarded as "burn-in" before the consensus tree was constructed from the remaining trees.

Results

Characterization of the chloroplast genomes

The chloroplast genome characteristics of the 13 species (*Frangula alnus*, *Rhamnus cathartica*, *R. crenata*, *R. davurica*, *R. diamantiaca*, *R. globose*, *R. heterophylla*, *R. lamprophylla*, *R. leptophylla*, *R. subapetala*, *R. taquetii*, *R. ussuriensis*, and *R. wilsonii*) are presented in Table 1. *R. subapetala* and *R. heterophylla* had the longest and shortest chloroplast genomes (161,426 bp and 156,514 bp, respectively). Although the chloroplast genome length of these 13 species varied, the genetic composition analysis revealed some similarities. The positions of the genes were visualized in Fig. 1. As observed in most other angiosperms, all 13 species displayed a classical quadripartite structure (Fig. 1), with two IR regions separating the LSC region and the SSC region. *R. taquetii* and *R. davurica* had the longest IR regions, each reaching 26,448 bp. *R. heterophylla* had the shortest IR region length of 25,306 bp. The shortest and longest LSC regions were found in *R. heterophylla* and *R. davurica* (86,023 bp and 88,777 bp, respectively), while the shortest and longest SSC regions were found in *R. cathartica* and *R. heterophylla* (18,887 bp and 19,879 bp, respectively). The chloroplast genome's total GC content varied slightly across *Rhamnus* species, ranging from 36.9% in *R. crenata* to 37.2% in *R. davurica*. The highest average GC content was found in the IR regions (42.8%), followed by the LSC (35.0%) and SSC (31.4%) regions. There were 131 predicted functional genes in most of the 13 chloroplast genomes, with 113 unique genes, including 79 protein-coding genes, 30

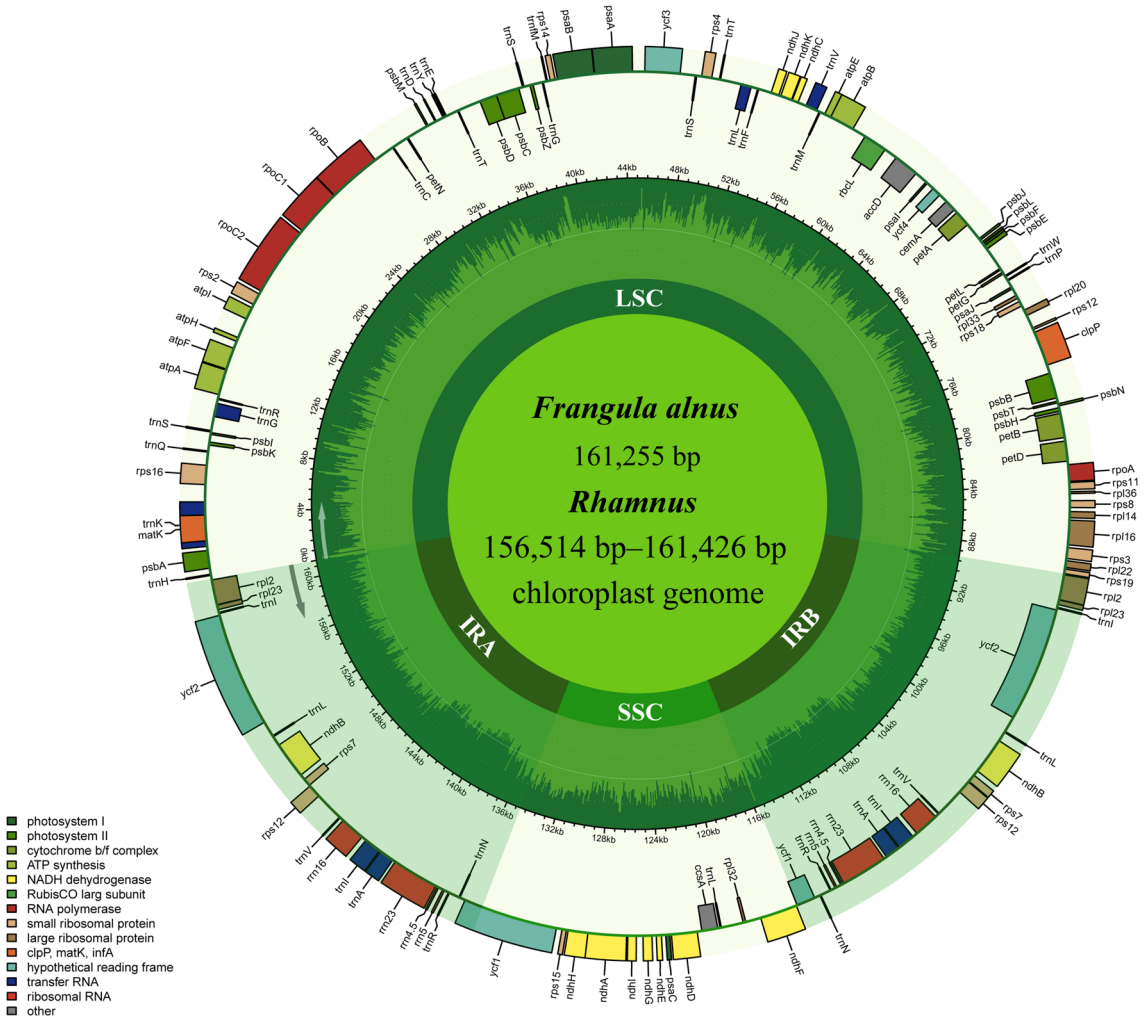

Fig. 1 The circular map of the complete chloroplast genomes of *Rhamnus* species and *Frangula alnus*. The genes are shown along the inner and outer sides of the circle. Genes on the outside of the circle are transcribed counterclockwise, and the genes inside the circle are transcribed clockwise. The dark green area and light green area of the inner circle represent the GC content and AT content of the genome, respectively. The large single copy (LSC), the small single copy (SSC), and the inverted repeat (IRA and IRB) regions are also represented. Different colors in genes represent different functions are shown in the bottom left corner

tRNA genes, and 4 rRNA genes. Notably, the *infA* gene was absent in five species (*F. alnus*, *R. cathartica*, *R. globose*, *R. heterophylla*, and *R. taquetii*). In addition, *R. cathartica* and *R. globosa* did not contain the *psbL* gene (Tables 1 and 2). The *ycf1* pseudogene of *R. heterophylla* and *R. taquetii* was lost at the junction of IRb/SSC (Figure S1). A single intron can be found in 15 genes, including those involved in photosynthesis and self-replication. The *rps12*, *clpP*, and *ycf3* genes included two introns (Table 2).

Identification of simple sequence repeats (SSRs) and long repeats

The distribution of SSRs in the chloroplast genomes of *Rhamnus* species and *Frangula alnus* was studied. MISA analysis was used in the analysis. The number of SSRs varied from 100 in *R. heterophylla* to 126 in *R. lamprophylla*, *R. subapetala*, and *R. wilsonii*, with all species having similar SSR type distributions (Table S1). Six different types of SSR were observed, which included mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeats (Fig. 2A). Mononucleotide repeats were the most common, accounting for about 70.99% of all SSRs. Dinucleotide repeats made up 12.42% of the total, while tetranucleotide repeats made up 9.94%, and trinucleotide repeats made up 3.83%. Pentanucleotide and hexanucleotide repeats were extremely rare in these chloroplast genomes, accounting for only 1.54% and 1.28%, respectively (Fig. 2B). The chloroplast genomes of *Rhamnus* species and *Frangula alnus* had a higher number of SSRs in the LSC region compared to the IR and SSC

Table 2 Lists of genomic genes for *Frangula alnus* and various *Rhamnus* species

Category	Gene group	Gene name
Photosynthesis	Subunits of photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Subunits of photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	Subunits of NADH dehydrogenase	*ndhA*, ndhB**(2), *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Subunits of cytochrome b/f complex	*petA, petB*, petD*, petG, petL, petN*
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI*
	Large subunit of rubisco	*rbcL*
Self-replication	Proteins of large ribosomal subunit	*rpl14, rpl16*, rpl2**(2), *rpl20, rpl22, rpl23*(2), *rpl32, rpl33, rpl36*
	Proteins of small ribosomal subunit	*rps11, rps12***(2), *rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7*(2), *rps8*
	Subunits of RNA polymerase	*rpoA, rpoB, rpoC1*, rpoC2*
	Ribosomal RNAs	*rrn16*(2), *rrn23*(2), *rrn4.5*(2), *rrn5*(2)
	Transfer RNAs	*trnA*-UGC*(2), *trnC*-GCA, *trnD*-GUC, *trnE*-UUC, *trnF*-GAA, *trnG*-GCC, *trnG*-UCC*, *trnH*-GUG, *trnI*-CAU(2), *trnI*-GAU*(2), *trnK*-UUU*, *trnL*-CAA(2), *trnL*-UAA*, *trnL*-UAG, *trnM*-CAU, *trnN*-GUU(2), *trnP*-UGG, *trnQ*-UUG, *trnR*-ACG(2), *trnR*-UCU, *trnS*-GCU, *trnS*-GGA, *trnS*-UGA, *trnT*-GGU, *trnT*-UGU, *trnV*-GAC(2), *trnV*-UAC*, *trnW*-CCA, *trnY*-GUA, *trnfM*-CAU
Other genes	Maturase	*matK*
	Protease	*clpP***
	Envelope membrane protein	*cemA*
	Acetyl-CoA carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
	Translation initiation factor	*infA*
Genes of unknown function	Conserved hypothetical chloroplast ORF	*ycf1*(2), *ycf2*(2), *ycf3**, ycf4*

Gene*: Gene with one introns; Gene**: Gene with two introns; #Gene: Pseudo gene; Gene(2): Number of copies of multi-copy genes

regions. Furthermore, the analysis showed that the intergenic sequences (IGS) had a larger proportion of SSRs than the coding sequences (CDS) (Fig. 2C).

Herein, we identified all repeats longer than 30 bp in the chloroplast genomes of *Rhamnus* species and *F. alnus*. In general, we identified four types of repeats: forward (F), inverted (R), complementary (C), and palindromic (P) repeats. Figure 2D showed the results of a repeat analysis of *Rhamnus* species and *F. alnus*, which revealed the presence of 16 to 83 forward repeats, 0 to 2 inverted repeats, 0 to 2 complementary repeats, and 17 to 35 palindromic repeats (Table S2). Forward and palindromic repeats were discovered to be more common than reverse and complement repeats.

Codon usage bias and selective pressure analysis

The 13 species had comparable codon preferences, with average codons ranging from 26,286 to 26,694. The lowest average codons were found in *R. taquetii* and *R. heterophylla*, while the highest were found in *R. leptophylla* and *R. diamantiaca* (Table S3). According to the RSCU values, these species were divided into two groups: *R. globose, R. heterophylla*, and *R. taquetii* in one group; *Frangula alnus, R. crenata, R. cathartica, R. subapetala, R. diamantiaca, R. lamprophylla, R. leptophylla, R. wilsonii, R. davurica*, and *R. ussuriensis* in the other (Fig. 3). The 64 codons were classified into three categories depending on their RSCU values. The first group contained only two codons (UUG and AGA) with RSCU values greater than 1.91. The second group included 29 codons with RSCU values greater than one, while the remaining codons were sorted into the third group. Tryptophan (UGG) and methionine (AUG) both had RUSC values of 1.

The selective pressure of protein-coding genes was analyzed by calculating the Ka/Ks values of protein-coding genes in *Rhamnus* species and *Frangula alnus*. *R. crenate* was used as a reference. Statistical analysis revealed that 79 of the protein-coding genes analyzed were relatively stable (Fig. 4). The majority of genes had Ka/Ks values less than one, ranging from 0 to 1.8679 (Table S4). The highest Ka/Ks value for the *ndhB* gene was 1.8679 in *R. subapetala* and *R. taquetii*. Except for *R. globosa*, all species had Ka/Ks ratios greater than 1 for the *rpl33* gene. The Ka/Ks value for the *rpl20* gene was higher than 1 in

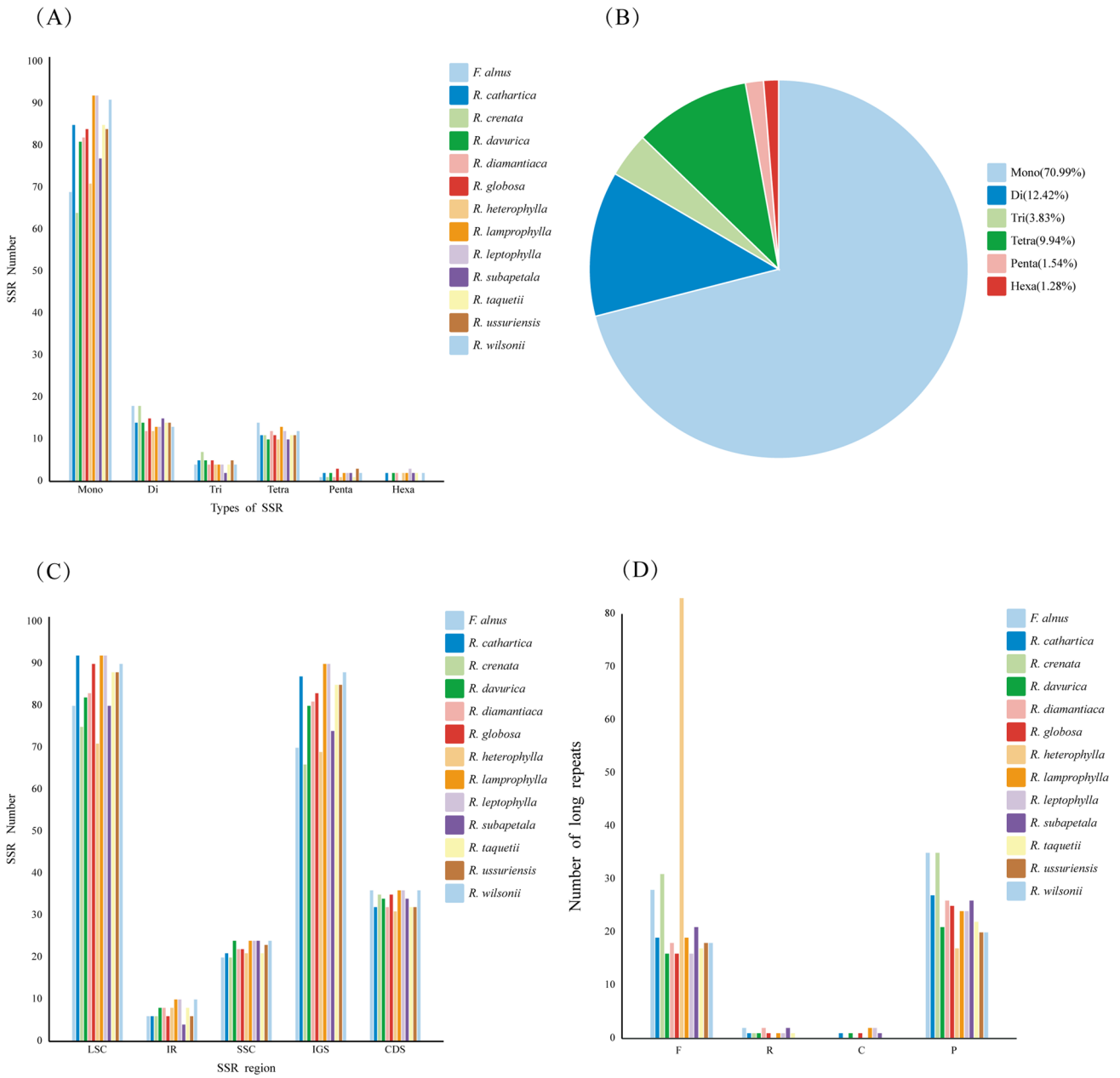

Fig. 2 Comparison of SSRs and long repeats in the chloroplast genomes. **A** The number of SSRs of different types of SSR for *Rhamnus* species and *Frangula alnus*. **B** Proportions in SSR types among *Rhamnus* species and *Frangula alnus*. **C** The number of SSR markers in the LSC /IR/SSC region along with IGS and CDS. **D** Number of four long repeat sequences in *Rhamnus* species and *Frangula*

most species except for five (*R. cathartica, R. davurica*, *R. globose*, *R. lamprophylla*, and *R. ussuriensis*). Moreover, the Ka/Ks values of four genes (*rpl33*, *rps2*, *ccsA*, and *ycf2*) in *R. lamprophylla* were greater than 1. With the exception of five species (*R. cathartica*, *R. davurica*, *R. globose*, *R. lamprophylla*, and *R. ussuriensis*), the Ka/Ks value for the *rpl20* gene was greater than 1. Four genes in *R. lamprophylla* had Ka/Ks values greater than 1, including *rpl33*, *rps2*, *ccsA*, and *ycf2*.

Divergence analysis of chloroplast genomes

By comparing the chloroplast genomes of *Rhamnus* species and *Frangula alnus* using mVISTA and *R. crenate* as the reference, the highly variable regions of the chloroplast genomes were identified (Fig. 5). In general, the IR regions of these species had less diversity than the SSC and LSC regions. Moreover, non-coding regions showed greater divergence than coding regions. The *ycf3*, *ndhA*, *rpl32*, and *ycf1* genes showed relatively significant divergence in

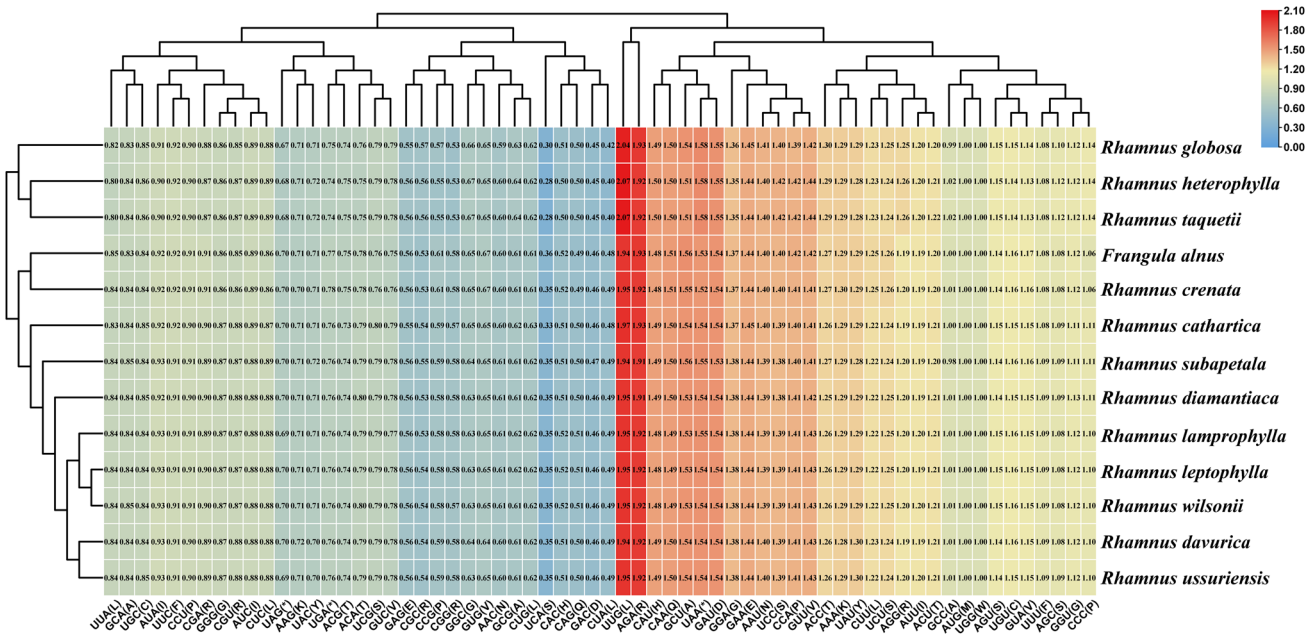

Fig. 3 Codon RSCU clustering in the chloroplast genomes of *Rhamnus* species and *Frangula alnus*. Tree clustering was based on the RSCU values of 64 codons. The RSCU value rises as the red color gets darker and falls as the blue color darkens. The various codons are represented by each column. Each different species is represented by each row

the coding regions. The intergenic spacer regions (*trnK*-UUU–*trnQ*-UUG, *atpH–atpI*, *trnY*-GUA–*trnE*-UUC, *trnG*-GCC–*trnfM*-CAU, *trnT*-UGU–*trnF*-GAA, *rpl20–rps12*, and *rpl22–rps19*) were the most diverse regions of the chloroplast genomes.

The Pi values ranged from 0 to 0.04394 (average 0.00526) (Table S5). In addition, the Pi value was found to be extremely low (0.001) in the IR regions, while it was observed to be greater in the LSC region (0.00776) compared to the SSC region (0.01883). Two genes, *rpl32* (0.02006) and *ycf1* (0.0026), and two intergenic regions, *atpH–atpI* (0.01952) and *rpl20–rps12* (0.04394), were found to display high variability in the chloroplast genomes of *Rhamnus* species and *Frangula alnus* (Fig. 6).

Phylogenetic analysis

To gain a better understanding of the phylogenetic relationships between *Rhamnus* and *Frangula* species, phylogenetic trees were constructed using complete chloroplast genome sequences. Both ML and BI trees based on the complete chloroplast genome exhibited the same tree structure (Fig. 7). The Bayesian posterior probability (PP) values and ML bootstrap values both showed high levels. Based on complete chloroplast genomes, phylogenetic trees can be divided into two major branches. *R. crenata* and *F. alnus* formed a large branch. Other *Rhamnus* species (*R. leptophylla*, *R. wilsonii*, *R. lamprophylla*, *R. davurica*, *R. heterophylla*, *R. taquetii*, *R. globosa*, *R. cathartica*, *R. ussuriensis*, and *R. subapetala*) were grouped into the second large branch.

Discussion

Rhamnaceae is a diverse plant family with significant variations in both morphology and genetics (Wanichthanarak et al. 2023). The discovery of *Rhamnaceae* fossils from the Cretaceous period highlights the significance of studying this plant family in relation to the rapid diversification of flowering plants (He and Lamont 2022; Shi et al. 2022). Herein, we selected *Rhamnus* and *Frangula* species (*Rhamnaceae*) with some medicinal value for our study (Whitfeld et al. 2014; Nigussie et al. 2021). The relationship between *Rhamnus* and *Frangula* species still needs to be further explored (Bolmgren and Oxelman 2004). Characterizing and comparing the complete chloroplast genomes has been shown to be a useful strategy for understanding the relationships between species (Yan et al. 2019). Herein, the complete chloroplast genome resources of the medicinal plants *R. cathartica* and *F. alnus* have been added. The *Rhamnus* species and *Frangula* species were analyzed at the complete chloroplast genome level for the first time.

These two chloroplast genomes were compared to the genomes of 11 other *Rhamnus* species. *R. cathartica* and *F. alnus* had chloroplast genomes that were 161,149 bp and

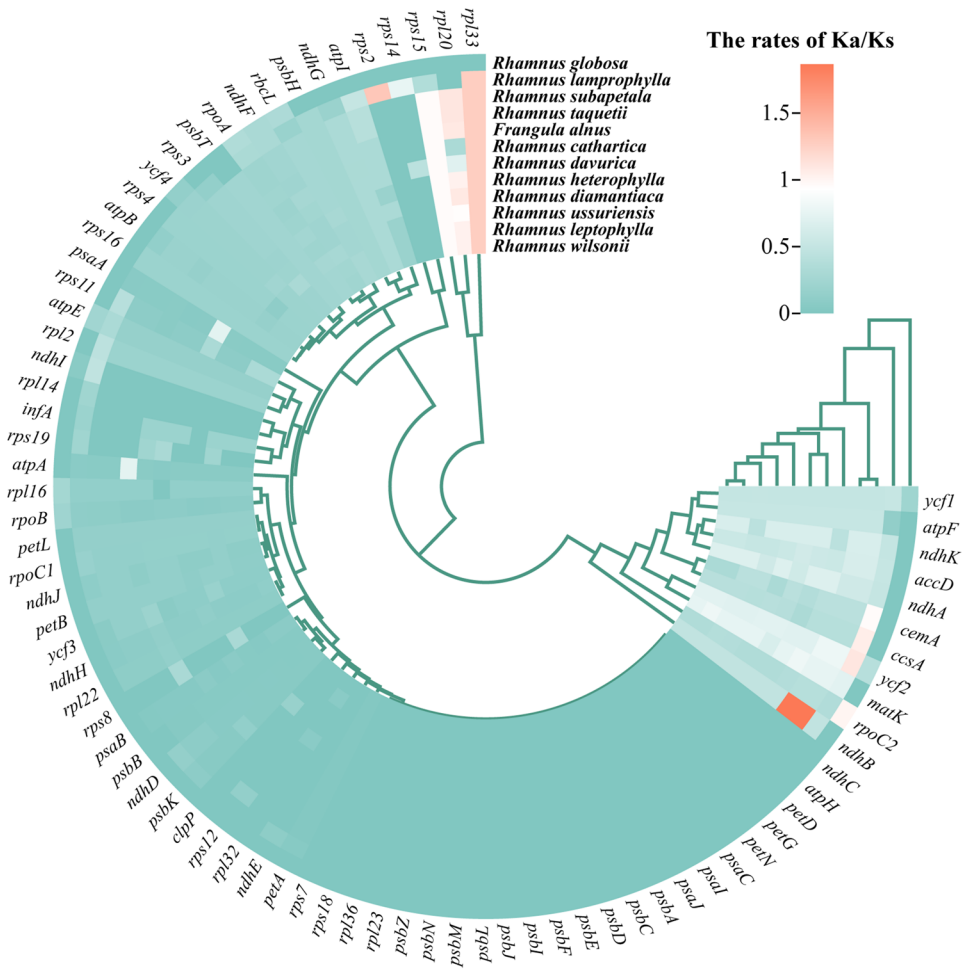

Fig. 4 The heat map for the Ka/Ks value statistics of *Rhamnus* species and *Frangula alnus*. The Ka/Ks value rises as the red color gets darker and falls as the green color darkens

161,255 bp in total, respectively. This is similar to the chloroplast genome length of *Rhamnaceae* species (Asaf et al. 2022). The quadripartite structure (LSC, SSC, and IRs) shown by the complete chloroplast genomes of *R. cathartica* and *F. alnus* is common in angiosperms (Wicke et al. 2011). The number and content of genes in *F. alnus* are similar to those of various *Rhamnus* species. These species identified a total of 129–131 chloroplast genes. The chloroplast genomes of *Ziziphus* (*Rhamnaceae*) species were previously found to have similar gene content and gene order (Huang et al. 2017; Asaf et al. 2022). These results reflect the conserved nature of the chloroplast genomes of *Rhamnaceae* species. However, gene loss and divergence in the chloroplast genomes of *Rhamnus* and *Frangula* species remain. The *infA* gene, which encodes a translation initiation factor, was absent from *F. alnus*, *R. cathartica*, *R. globose*, *R. heterophylla*, and *R. taquetii*. The *infA* gene may have moved from the chloroplast genome to the nucleus during the development of angiosperms (Millen et al. 2001). Several *Rhamnaceae* species also experience loss of the *infA* gene (Wanichthanarak et al. 2023). Two *Rhamnus* species lack the *psbL* gene. Previous studies suggest that the normal functions of chloroplasts are unaffected by the absence of the *psbL* gene (Ikeuchi et al. 1995; Suorsa et al. 2004). The absence of the *ycf1* pseudogene is common in *Rhamnaceae* species, and may be due to the contraction and expansion of the IR regions (Asaf et al. 2022).

SSRs are typically tandem repeats with 1–6 nucleotide patterns (Subramanian et al. 2003). SSRs in the chloroplast genome were thought to play a significant role in population genetics and phylogenetic analyses (Terrab et al. 2006). Herein, the distribution of SSRs in the chloroplast genomes of *Rhamnus* species and *Frangula alnus* was studied. The number of SSRs among these species ranged from 100 to 126. SSRs are primarily located in the LSC and IGS regions because these regions are more prone to developing microstructural changes (Yamane et al. 2006; Orton et al. 2017). Mono-nucleotide SSRs were the most common in the *Rhamnus* and *Frangula* species. The same finding was observed in the previous study on *Rhamnaceae* species (Huang et al. 2017; Yan et al. 2019; Asaf et al. 2022; Wanichthanarak et al. 2023), indicating that mono-nucleotide SSRs are more likely to cause chloroplast genome divergence (Ebert and Peakall 2009). Long repeat sequences have played a crucial

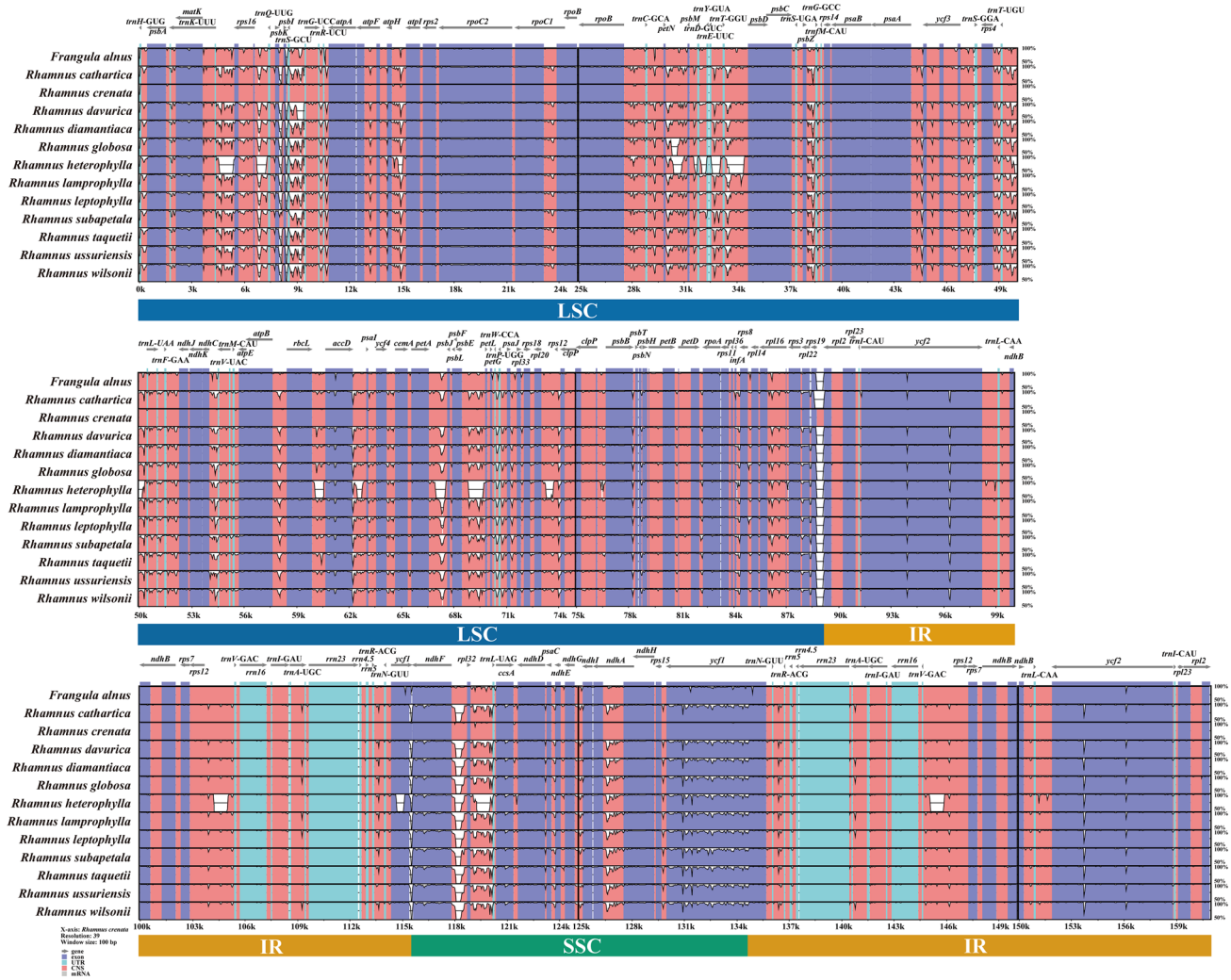

Fig. 5 The mVISTA map of the complete chloroplast genomes of *Rhamnus* species and *Frangula alnus*. *R. crenata* was used as the reference species. The dark gray arrows indicate the orientation of individual genes; purple bars indicate exons; blue bars indicate untranslated regions (UTR); pink bars indicate non-coding sequences (CNS); gray bars indicate mRNA. The x-axis indicates the coordinates within the chloroplast genome and the y-axis indicates the percentage identity, ranging from 50 to 100%

role in the analysis of chloroplast genome rearrangement, reorganization, and phylogenetics (Chumley et al. 2006). We discovered that the chloroplast genomes of *Rhamnus* and *Frangula* species had a high proportion of forward and palindromic repeats while examining long repeats. The outcome is similar to that of *Ziziphus* species and can be used to establish genetic markers for population genetics assessment (Huang et al. 2017; Asaf et al. 2022).

Codon usage bias is common in protein-coding genes in chloroplast genomes (Zhou et al. 2008). Analyzing codon preferences can provide valuable insights into species evolution and molecular development (Behura and Severson 2013). Codon usage analysis revealed that out of *Rhamnus* species and *Frangula alnus*, 29 codons had RSCU values greater than 1. Among these, 14 ended in U, 8 in A, 5 in C, and 2 in G, indicating a higher frequency of codons ending in U/A. We discovered that the codon preferences of *F. alnus* and several *Rhamnus* species are quite similar. This suggests a close relationship between *Frangula* and *Rhamnus* species. Notably, *F. alnus* and *R. crenata* exhibit highly similar codon preferences. In our phylogenetic analysis, we will concentrate on determining if these two species are closely related. The Ka/Ks ratio is frequently used to estimate the pressure of natural selection and the rates of evolution of protein-coding genes (Yang and Nielsen 2000). Herein, most of the Ka/Ks values in the protein-coding genes of the *Frangula* and *Rhamnus* species were less than 1, suggesting that these genes are experiencing purification selection and evolving relatively slowly. The same phenomenon has been observed in the study of *Ziziphus* species (Huang et al. 2017). The

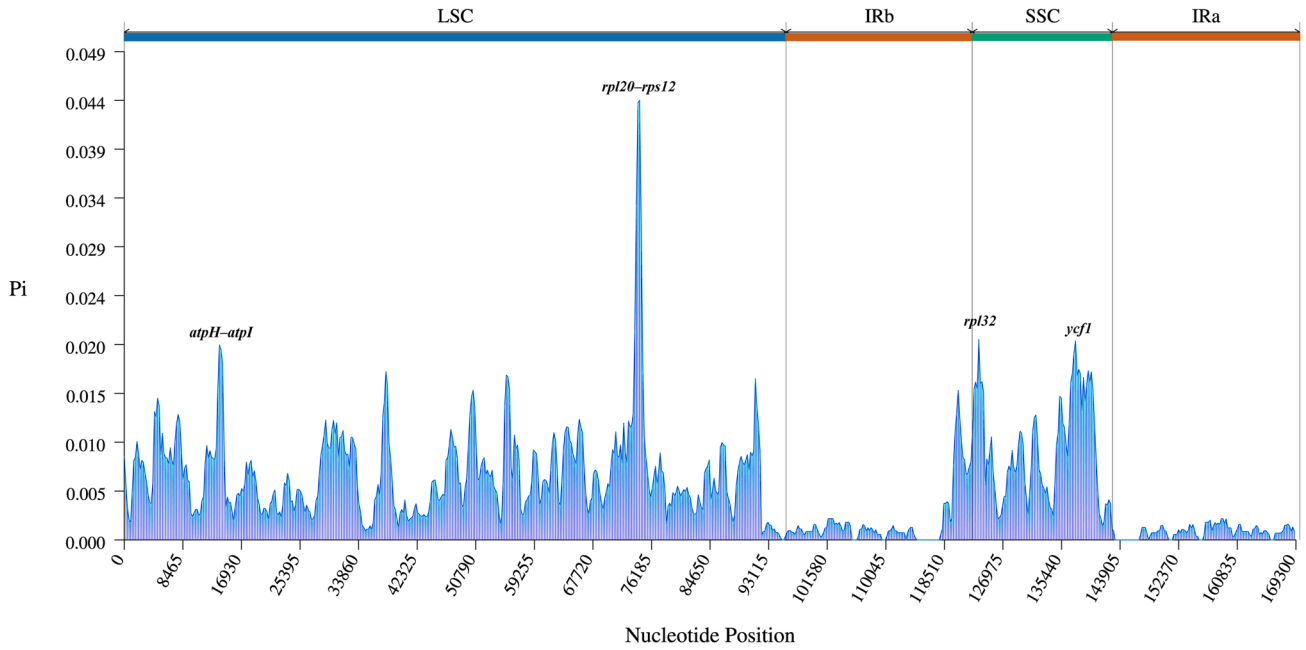

Fig. 6 Nucleotide diversity (Pi) in the complete chloroplast genomes of *Rhamnus* species and *Frangula alnus*. The x-axis and y-axis represent the base position along the sequences and Pi value, respectively

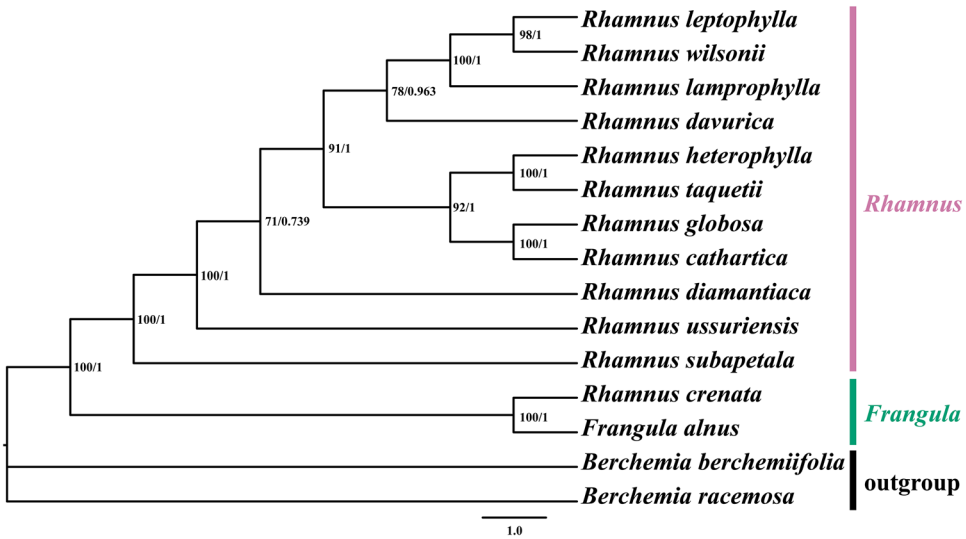

Fig. 7 Phylogenetic analysis. Phylogenetic trees were constructed using the maximum likelihood (ML) and Bayesian inference (BI) methods based on the complete chloroplast genomes of *Rhamnus* species and *Frangula alnus*. *Berchemia berchemiifolia* and *Berchemia racemose* were used as outgroups. The numbers above the nodes indicate support values

Ka/Ks values of the *rpl20* and *rpl33* genes were greater than one, suggesting rapid evolution under positive selection. This supports the hypothesis that protein-coding genes evolve faster than photosynthetic genes (Saina et al. 2018).

By comparing various chloroplast genome sequences, it is possible to find highly variable regions that may help in distinguishing closely associated species or genera (Nock et al. 2011; Li et al. 2015). According to the mVISTA comparison, the chloroplast genome sequences of 13 species displayed less variation in the IR regions but greater divergence in the LSC and SSC regions. The higher GC content in the IR regions may account for this phenomenon, as there is a demonstrated positive correlation between sequence stability and GC content (Zhang et al. 2011). Our analysis of mVISTA results and Pi values revealed several specific regions (*trnK*-UUU–*trnQ*-UUG, *atpH–atpI*, *trnY*-GUA–*trnE*-UUC, *trnG*-GCC–*trnfM*-CAU, *trnT*-UGU–*trnF*-GAA, *rpl20–rps12,* and *rpl22–rps19*) as well as genes (*ycf3*, *ndhA*, *rpl32,* and *ycf1*) that exhibit significant divergence. These variable regions may be utilized for both species discrimination and phylogenetic studies (Shi et al. 2023b). However, further study is needed to

determine which of these variable regions can be utilized for phylogenetic analysis.

Phylogenetic studies of the *Rhamnus* genus and related species based on ITS and *trnL–trnF* sequences may display some poor support values in phylogenetic trees (Bolmgren and Oxelman 2004; Hauenschild et al. 2016). Additional genetic information is required to enhance the resolution of the *Frangula* and *Rhamnus* species' phylogenies. Chloroplasts, which are inherited from a single parent, offer a clear and stable genetic lineage that is highly suitable for evolutionary and phylogenetic research (Palmer 1987; Jansen et al. 2007; Shi et al. 2023b). In contrast, the biparental inheritance of nuclear genes makes it challenging to trace evolutionary relationships over generations (Olmstead and Palmer 1994; Hare 2001; Seehausen 2004). Moreover, the nuclear genome is complex and may be difficult to access or too fragmented for analysis (Palmer 1987; Olmstead and Palmer 1994; Zhang and Hewitt 2003). Therefore, we selected the complete chloroplast genome, which contains more informative loci (Zhang et al. 2022), for the phylogenetic analysis of *Rhamnus* species and *Frangula alnus*. The support values for phylogenetic trees constructed using the two methods (ML and BI) were high. The majority of *Rhamnus* species formed a large branch, which is consistent with previous studies (Bolmgren and Oxelman 2004; Hauenschild et al. 2016; Azeem et al. 2019). Interestingly, *R. crenata* and *Frangula alnus* were closely related, forming a single cluster. Moreover, our codon preference analysis revealed similarities between these two species. However, previous studies have shown that the *Frangula* genus is monophyletic (Bolmgren and Oxelman 2004; Hauenschild et al. 2016). This suggests that *R. crenata* may be more appropriately classified under the *Frangula* genus. Sampling can impact the phylogenetic tree (Heath et al. 2008), and there is often a conflict between the phylogenies of nuclear and chloroplast DNA (Chat et al. 2004). Therefore, future studies will require broader sampling and extensive evaluation of nuclear data, or even whole genomes including nuclear genomes, to determine whether *R. crenata* should indeed be reclassified under *Frangula* genus. *R. cathartica* was found to be most closely related to *R. globose*, which also has medicinal value (Al-Judaibi and Al-Yousef 2014). This suggests that there may be a correlation between medicinal value and phylogenetic differences. We further reviewed relevant studies to determine whether the species analyzed possessed medicinal value (Table S6). *Frangula alnus* formed the base of the phylogenetic trees, whereas *R. heterophylla* and *R. davurica*, both possessing medicinal properties (Chen et al. 2016; Wang et al. 2019), were relatively distantly related to *R. cathartica*. This indicates that there is no significant correlation between the medicinal value of the species analyzed and their phylogenetic relationships. The phylogenetic analysis based on the complete chloroplast genome in this study will provide insights for future studies.

Conclusions

In this study, the complete chloroplast genomic resources of the medicinal plants *Rhamnus cathartica* and *Frangula alnus* have been added. Then, the first comparative analysis of the *Rhamnus* and *Frangula* species based on complete chloroplast genomes was provided. Our findings revealed similarities and differences in the chloroplast genomes of these species. The lack of the *infA* and *psbL* genes does not negatively impact the normal functioning of *Rhamnus* and *Frangula* species. The absence of the *ycf1* pseudogene is common in *Rhamnaceae* species, and may be due to the contraction and expansion of the IR regions. The *rpl20* and *rpl33* genes of *Rhamnus* and *Frangula* species are undergoing rapid evolution under positive selection conditions. *Rhamnus* and *Frangula* species prefer amino acids with A/U-terminal codons. SSRs, long repeats, and variable regions have been identified in the *Rhamnus* and *Frangula* chloroplast genomes, with potential applications as markers for population genetics and phylogenetic studies. The phylogenetic relationships between *Rhamnus* species and *Frangula alnus* were reconstructed based on the complete chloroplast genome.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12298-023-01331-7.

Author contributions WBS: Conceptualization, Data curation, Investigation, Methodology, Project administration, Software, Visualization, Writing—Review & Editing, Writing—Original Draft. SH: Conceptualization, Formal analysis, Writing—Original Draft. WCS: Data curation, Formal analysis, Investigation, Project administration. YH: Data curation, Formal analysis. CS: Conceptualization, Data curation, Funding acquisition, Project administration, supervision, Writing—Original Draft. SW: Data curation, Funding acquisition, Project administration. All authors contributed to the article and approved the final version.

Funding This work was supported by the State Key Laboratory of Palaeobiology and Stratigraphy (No. 223123 and No. 213119), National Natural Science Foundation of China (NO. 31801022), and Shandong Province Natural Science Foundation of China (NO. ZR2019BC094).

Data availability The data that backs up the conclusions of this study can be found in the GenBank database under the accession numbers NC_068506 and NC_068507 by visiting https://www.ncbi.nlm.nih.gov/.

Declarations

Competing interests The authors declare no conflicts of interest in their work as presented in this manuscript.

References

Al-Judaibi A, Al-Yousef F (2014) Antifungal effect of ethanol plant extract on Candida SP. Am J Agric Biol Sci 9:277–283. https://doi.org/10.3844/ajabssp.2014.277.283

Asaf S, Ahmad W, Al-Harrasi A, Khan AL (2022) Uncovering the first complete plastome genomics, comparative analyses, and phylogenetic dispositions of endemic medicinal plant *Ziziphus hajarensis* (Rhamnaceae). BMC Genomics 23:1–16. https://doi.org/10.1186/s12864-022-08320-2

Azeem F, Bilal A, Rana MA et al (2019) Drought affects aquaporins gene expression in important pulse legume chickpea (Cicer arietinum L). Pakistan J Bot 51:81–88. https://doi.org/10.30848/PJB2019

Behura SK, Severson DW (2013) Codon usage bias: causative factors, quantification methods and genome-wide patterns: With emphasis on insect genomes. Biol Rev 88:49–61. https://doi.org/10.1111/j.1469-185X.2012.00242.x

Beier S, Thiel T, Münch T et al (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolmgren K, Oxelman B (2004) Generic limits in *Rhamnus* L. s.l. (Rhamnaceae) inferred from nuclear and chloroplast DNA sequence phylogenies. Taxon 53:383–390. https://doi.org/10.2307/4135616

Brown J, Pirrung M, Mccue LA (2017) FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics 33:3137–3139. https://doi.org/10.1093/bioinformatics/btx373

Chat J, Jáuregui B, Petit RJ, Nadot S (2004) Reticulate evolution in kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. Am J Bot 91:736–747. https://doi.org/10.3732/ajb.91.5.736

Chen C, Chen H, Zhang Y et al (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13:1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Chen G, Li X, Saleri F, Guo M (2016) Analysis of flavonoids in *Rhamnus davurica* and its antiproliferative activities. Molecules 21:1275. https://doi.org/10.3390/molecules21101275

Christenhusz MJM, Byng JW (2016) The number of known plants species in the world and its annual increase. Phytotaxa 261:201–217. https://doi.org/10.11646/phytotaxa.261.3.1

Chumley TW, Palmer JD, Mower JP et al (2006) The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23:2175–2190. https://doi.org/10.1093/molbev/msl089

Darriba D, Taboada GL, Doallo R, Posada D (2012) JModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772. https://doi.org/10.1038/nmeth.2109

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: *De novo* assembly of organelle genomes from whole genome data. Nucleic Acids Res 45:1–9. https://doi.org/10.1093/nar/gkw955

Dodsworth S (2015) Genome skimming for next-generation biodiversity analysis. Trends Plant Sci 20:525–527. https://doi.org/10.1016/j.tplants.2015.06.012

Ebert D, Peakall R (2009) Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour 9:673–690. https://doi.org/10.1111/j.1755-0998.2008.02319.x

Frazer KA, Pachter L, Poliakov A et al (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res 32:273–279. https://doi.org/10.1093/nar/gkh458

Gassmann A, Tosevski I, Skinner L (2008) Use of native range surveys to determine the potential host range of arthropod herbivores for biological control of two related weed species, *Rhamnus cathartica* and *Frangula alnus*. Biol Control 45:11–20. https://doi.org/10.1016/j.biocontrol.2007.12.004

Green MR, Sambrook J (1972) Agarose gel electrophoresis. Cold Spring Harb Protoc 124:7–19. https://doi.org/10.1101/pdb.prot100404

Grubov VI (1949) Monography of *Rhamnus* L. sl. In: Schischkin BK (ed) Flora et systematica plantae vasculares. Academiae Scientiarum USSR, Leningrad, pp 14–425

Guo C, Luo Y, Gao LM et al (2023) Phylogenomics and the flowering plant tree of life. J Integr Plant Biol 65:299–323. https://doi.org/10.1111/jipb.13415

Hare MP (2001) Prospects for nuclear gene phylogeography. Trends Ecol Evol 16:700–706. https://doi.org/10.1016/S0169-5347(01)02326-6

Hauenschild F, Favre A, Salazar GA, Muellner-Riehl AN (2016) Analysis of the cosmopolitan buckthorn genera *Frangula* and *Rhamnus* s.l. supports the description of a new genus, Ventia. Taxon 65:65–78. https://doi.org/10.12705/651.5

He T, Lamont BB (2022) Ancient Rhamnaceae flowers impute an origin for (fire-prone) flowering plants exceeding 250-million-years ago. iScience 25:104642. https://doi.org/10.1016/j.isci.2022.104642

Heath TA, Zwickl DJ, Kim J, Hillis DM (2008) Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Syst Biol 57:160–166. https://doi.org/10.1080/10635150701884640

Henriquez CL, Abdullah AI et al (2020) Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). Genomics 112:2349–2360. https://doi.org/10.1016/j.ygeno.2020.01.006

Huang J, Chen R, Li X (2017) Comparative analysis of the complete chloroplast genome of four known *Ziziphus* species. Genes (basel) 8:340. https://doi.org/10.3390/genes8120340

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Ikeuchi M, Shukla VK, Pakrasi HB, Noue Y (1995) Directed inactivation of the *psbl* gene does not affect Photosystem II in the cyanobacterium *Synechocystis* sp. PCC 6803. MGG Mol Gen Genet 249:622–628. https://doi.org/10.1007/BF00418031

Jansen RK, Cai Z, Raubeson LA et al (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci U S A 104:19369–19374. https://doi.org/10.1073/pnas.0709121104

Jheng C, Chen T, Lin J et al (2012) The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. Plant Sci 190:62–73. https://doi.org/10.1016/j.plantsci.2012.04.001

Kang KB, Ernst M, van der Hooft JJJ et al (2019) Comprehensive mass spectrometry-guided phenotyping of plant specialized metabolites reveals metabolic diversity in the cosmopolitan plant family Rhamnaceae. Plant J 98:1134–1144. https://doi.org/10.1111/tpj.14292

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Kearse M, Moir R, Wilson A et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and

analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199
Kirchhoff H (2019) Chloroplast ultrastructure in plants. New Phytol 223:565–574. https://doi.org/10.1111/nph.15730
Kremer D, Kosalec I, Locatelli M et al (2012) Anthraquinone profiles, antioxidant and antimicrobial properties of *Frangula rupestris* (Scop.) Schur and *Frangula alnus* Mill. bark. Food Chem 131:1174–1180. https://doi.org/10.1016/j.foodchem.2011.09.094
Kurtz S, Choudhuri JV, Ohlebusch E et al (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29:4633–4642. https://doi.org/10.1093/nar/29.22.4633
Lehwark P, Greiner S (2018) GB2sequin—a file converter preparing custom GenBank files for database submission. Genomics 111:759–761. https://doi.org/10.1016/j.ygeno.2018.05.003
Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191/4994778
Li H, Guo Q, Xu L et al (2023) CPJSdraw : analysis and visualization of junction sites of chloroplast genomes. PeerJ 11:e15326. https://doi.org/10.7717/peerj.15326
Li X, Yang Y, Henry RJ et al (2015) Plant DNA barcoding: from gene to genome. Biol Rev Camb Philos Soc 90:157–166. https://doi.org/10.1111/brv.12104
Millen RS, Olmstead RG, Adams KL et al (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13:645–658. https://doi.org/10.1105/tpc.13.3.645
Nielsen AZ, Mellor SB, Vavitsas K et al (2016) Extending the biosynthetic repertoires of cyanobacteria and chloroplasts. Plant J 87:87–102. https://doi.org/10.1111/tpj.13173
Nigussie G, Alemu M, Ibrahim F et al (2021) Phytochemicals, traditional uses and pharmacological activity of *Rhamnus* prinoides: a review. Int J Second Metab 8:136–151. https://doi.org/10.21448/ijsm.833554
Nock CJ, Waters DLE, Edwards MA et al (2011) Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol J 9:328–333. https://doi.org/10.1111/j.1467-7652.2010.00558.x
Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. Am J Bot 81:1205–1224. https://doi.org/10.1002/j.1537-2197.1994.tb15615.x
Orton LM, Burke SV, Wysocki WP, Duvall MR (2017) Plastid phylogenomic study of species within the genus *Zea*: rates and patterns of three classes of microstructural changes. Curr Genet 63:311–323. https://doi.org/10.1007/s00294-016-0637-8
Palmer JD (1987) Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am Nat 130:S6–S29. https://doi.org/10.1086/284689
Ranwez V, Douzery EJP, Cambon C et al (2018) MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol Biol Evol 35:2582–2584. https://doi.org/10.1093/molbev/msy159
Richardson JE, Chatrou LW, Mols JB et al (2004) Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. Philos Trans R Soc B Biol Sci 359:1495–1508. https://doi.org/10.1098/rstb.2004.1537
Richardson JE, Fay MF, Cronk QCB, Chase MW (2000) A revision of the tribal classification of Rhamnaceae. Kew Bull 55:311–340. https://doi.org/10.2307/4115645
Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248
Saina JK, Gichira AW, Li ZZ et al (2018) The complete chloroplast genome sequence of *Dodonaea viscosa* : comparative and phylogenetic analyses. Genetica 146:101–113. https://doi.org/10.1007/s10709-017-0003-x
Seehausen O (2004) Hybridization and adaptive radiation. Trends Ecol Evol 19:198–207. https://doi.org/10.1016/j.tree.2004.01.003
Shi C, Hu N, Huang H et al (2012) An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. PLoS ONE 7:e31468. https://doi.org/10.1371/journal.pone.0031468
Shi C, Wang S, Cai HH et al (2022) Fire-prone Rhamnaceae with South African affinities in Cretaceous Myanmar amber. Nat Plants 8:125–135. https://doi.org/10.1038/s41477-021-01091-w
Shi L, Chen H, Jiang M et al (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res 47:W65–W73. https://doi.org/10.1093/nar/gkz345
Shi W, Song W, Chen Z et al (2023a) Comparative chloroplast genome analyses of diverse *Phoebe* (Lauraceae) species endemic to China provide insight into their phylogeographical origin. PeerJ 11:e14573. https://doi.org/10.7717/peerj.14573
Shi W, Song W, Liu J et al (2023b) Comparative chloroplast genome analysis of *Citrus* (Rutaceae) species : insights into genomic characterization, phylogenetic relationships, and discrimination of subgenera. Sci Hortic 313:111909. https://doi.org/10.1016/j.scienta.2023.111909
Song W, Chen Z, Shi W et al (2022) Comparative analysis of complete chloroplast genomes of nine species of Litsea (Lauraceae): hypervariable regions, positive selection, and phylogenetic relationships. Genes (basel) 13:1550. https://doi.org/10.3390/genes13091550
Subramanian S, Mishra RK, Singh L (2003) Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biol 4:1–10. https://doi.org/10.1186/gb-2003-4-2-r13
Suorsa M, Regel RE, Paakkarinen V et al (2004) Protein assembly of photosystem II and accumulation of subcomplexes in the absence of low molecular mass subunits *PsbL* and *PsbJ*. Eur J Biochem 271:96–107. https://doi.org/10.1046/j.1432-1033.2003.03906.x
Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38:3022–3027. https://doi.org/10.1093/molbev/msab120
Terrab A, Paun O, Talavera S et al (2006) Genetic diversity and population structure in natural populations of Moroccan Atlas cedar (*Cedrus atlantica*; Pinaceae) determined with cpSSR markers. Am J Bot 93:1274–1280. https://doi.org/10.3732/ajb.93.9.1274
Tillich M, Lehwark P, Pellizzer T et al (2017) GeSeq—versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11. https://doi.org/10.1093/nar/gkx391
Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44:W232–W235. https://doi.org/10.1093/NAR/GKW256
Wang D, Zhang Y, Zhang Z et al (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinform 8:77–80. https://doi.org/10.1016/S1672-0229(10)60008-3
Wang L, Fan S, Wang X et al (2019) Physicochemical aspects and sensory profiles as various potential factors for comprehensive quality assessment of Nü-Er-Cha produced from *Rhamnus heterophylla* Oliv. Molecules 24:3211. https://doi.org/10.3390/molecules24183211
Wanichthanarak K, Nookaew I, Pasookhush P et al (2023) Revisiting chloroplast genomic landscape and annotation towards comparative chloroplast genomes of Rhamnaceae. BMC Plant Biol 23:1–22. https://doi.org/10.1186/s12870-023-04074-5
Whitfeld TJS, Lodge AG, Roth AM, Reich PB (2014) Community phylogenetic diversity and abiotic site characteristics influence abundance of the invasive plant *Rhamnus cathartica* L. J Plant Ecol 7:202–209. https://doi.org/10.1093/jpe/rtt020
Wicke S, Schneeweiss GM, DePamphilis CW et al (2011) The evolution of the plastid chromosome in land plants: gene content, gene

order, gene function. Plant Mol Biol 76:273–297. https://doi.org/10.1007/s11103-011-9762-4

Yamane K, Yano K, Kawahara T (2006) Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. DNA Res 13:197–204. https://doi.org/10.1093/dnares/dsl012

Yan M, Zhao X, Zhou J et al (2019) The complete chloroplast genomes of *Punica granatum* and a comparison with other species in Lythraceae. Int J Mol Sci 20:2886. https://doi.org/10.3390/ijms20122886

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43. https://doi.org/10.1093/oxfordjournals.molbev.a026236

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Mol Ecol 12:563–584. https://doi.org/10.1046/j.1365-294X.2003.01773.x

Zhang J, Kuo CCJ, Chen L (2011) GC content around splice sites affects splicing through pre-mRNA secondary structures. BMC Genomics 12:1–11. https://doi.org/10.1186/1471-2164-12-90

Zhang YM, Han LJ, Yang CW et al (2022) Comparative chloroplast genome analysis of medicinally important *Veratrum* (Melanthiaceae) in China: insights into genomic characterization and phylogenetic relationships. Plant Divers 44:70–82. https://doi.org/10.1016/j.pld.2021.05.004

Zheng S, Poczai P, Hyvönen J et al (2020) Chloroplot: an online program for the versatile plotting of organelle genomes. Front Genet 11:576124. https://doi.org/10.3389/fgene.2020.576124

Zhou M, Long W, Li X (2008) Patterns of synonymous codon usage bias in chloroplast genomes of seed plants. Stud China 10:235–242. https://doi.org/10.1007/s11632-008-0047-1

Zurawski G, Clegg MT (1987) Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Annu Rev Plant Physiol 38:391–418. https://doi.org/10.1146/annurev.pp.38.060187.002135