

Comparative chloroplast genome analysis of four Hippophae rhamnoides subspecies and its phylogenetic analysis

Ruoqiong Wang (✉ wangr_q@163.com)

Shanxi Agricultural University <https://orcid.org/0009-0006-0003-8836>

Xiangjie Wang

Shanxi Agricultural University

Zhihong Liu

Shanxi Agricultural University

Youzhi Han

Shanxi Agricultural University

Qing Xie

Shanxi Agricultural University <https://orcid.org/0000-0003-2889-5670>

Research Article

Keywords: Hippophae rhamnoides, Chloroplast genome, Comparative analysis, Phylogeny analysis

Posted Date: June 26th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3077329/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on November 16th, 2023. See the published version at <https://doi.org/10.1007/s10722-023-01788-1>.

Abstract

Sea buckthorn(*Hippophae rhamnoides*), a hardy deciduous shrub of the family Elaeagnaceae, grows wild at high altitudes in Asia and Europe and is known for its great nutritional, medicinal and ecological value. In this study, the chloroplast genomes of four *H. rhamnoides* subspecies, *H. rhamnoides* subsp. *mongolica* 'wulanshalin', *H. rhamnoides* subsp. *caucasica*, *H. rhamnoides* subsp. *sinensis* 'wucixiong', and *H. rhamnoides* subsp. *yunnanensis*, were characterized. The results showed that the genome length of these four subspecies ranged from 157,436 bp to 157,822 bp, with a typical quadripartite structure. Comparing the genome structure of these four subspecies, it was found that the chloroplast genomes were relatively conserved, retaining the same gene order. The annotation contained a total of 132 genes in each chloroplast genome, with 86 protein-coding genes, 38 tRNA genes, and 8 rRNA genes. This study identified 14 highly differentiated regions and 423 simple sequence repeats loci, which can be used as potential molecular markers for *H. rhamnoides*. Meanwhile, Phylogenetic analysis showed that all *Hippophae* taxa were clustered in the same group and formed a sister clade with *Elaeagnus* taxa supported by Bayesian posterior probabilities. Among *Hippophae* taxa, *H. gyantsensis*, *H. neurcar* and *H. salicifolia* were grouped together, but *H. tibetana* was clustered with *H. rhamnoides* and the other six *H. rhamnoides* subspecies. The findings of this research will be useful for further studies on resource protection and the taxonomic classification of sea buckthorn.

Introduction

Chloroplast is the main place for photosynthesis in plants. It has the functions of synthesizing starch, aliphatic acid, phytohormones, pigments and amino acids, and usually exists in terrestrial plants, algae and some protists (Neuhaus and Emes 2000). Chloroplasts contain relatively independent genetic material, double-chained chloroplast DNA. Usually, the chloroplast genome of higher plants is a highly conserved tetrad structure, including a large single-copy region (LSC), a small single-copy region (SSC) and two reverse repeat regions (IR) with the same sequence and opposite directions (Sugiura 1992). Different from the nuclear genome and mitochondrial genome, the chloroplast genome is characterized by monophyletic inheritance, structural stability, and slow evolution rate and its size is moderate and easier to sequence than complex nuclear genomes (Jansen et al. 2007). With the development of next-generation sequencing technology and the reduction of cost, the research on chloroplast genome has attracted the attention of many scholars, and this technique plays an important role in plant phylogenetic research, species identification, and genetic relationship analysis (Lee et al., 2019; Gong et al. 2022).

Hippophae rhamnoides, popularly known as sea buckthorn or vitamin tree, is a bush in the Elaeagnaceae family. These plants have strong ecological adaptability, cold resistance, and high-temperature resistance, and can resist wind, drought, and saline-alkali water (Huang 2007). In addition, *H. rhamnoides* is a medicinal and edible plant, and its various organs are rich in nutrients and physiologically active substances, such as vitamins, carotenoids, organic acids, unsaturated fatty acids and some essential amino acids, etc. (Gâtlan and Gutt, 2021). At present, the research of *H. rhamnoides* mainly focuses on

systematic taxonomy, chemical composition and pharmacology (Guo et al. 2017; Li et al. 2019; Ma et al. 2022).

H. rhamnoides has the most abundant subspecies in the genus *Hippophae*. There are nine *H. rhamnoides* subspecies, which are *H. rhamnoides* subsp. *sinensis*, *H. rhamnoides* subsp. *yunnanensis*, *H. rhamnoides* subsp. *wolongensis*, *H. rhamnoides* subsp. *turkestanica*, *H. rhamnoides* subsp. *mongolica*, *H. rhamnoides* subsp. *caucasica*, *H. rhamnoides* subsp. *carpatica*, *H. rhamnoides* subsp. *fluviatilis* and *H. rhamnoides* subsp. *rhamnoides* (Lian and Chen 1996). In China, *H. rhamnoides* subsp. *sinensis* is the most widely distributed plant of *H. rhamnoides*, and it is also the largest group for development, utilization and planting. *H. rhamnoides* subsp. *sinensis* 'wucixiong' is a single clone variety formed by systematic selection based on the germplasm resources of *H. rhamnoides* subsp. *sinensis*. (Huang 2007). *H. rhamnoides* subsp. *mongolica* 'wulanshalin' an excellent clone variety selected from the natural hybrid offspring of *H. rhamnoides* subsp. *mongolica*, which is mainly cultivated in Mongolia. It is famous for its large fruit, thick skin, cold resistance, and strong adaptability to the environment (Huang 2007).

H. rhamnoides subsp. *yunnanensis* is a unique group in China. It is mainly distributed in the northwest of Yunnan Province, the east of Tibet, and the vast area of Sichuan Province. *H. rhamnoides* subsp. *yunnanensis* is rich in vitamin C, which is 0.2-1 times higher than that of *H. rhamnoides* subsp. *sinensis*. At the same time, it is also rich in vitamins B1, B2, E, P and other nutrients. So, it has the potential to be widely planted as an economic crop in the local area (Yuan 1988). *H. rhamnoides* subsp. *caucasica* is distributed in Bulgaria, Iran, North Caucasus, Turkey and other places. It often grows along the river and floodplain at an altitude of 1000–3000 m (Lian and Chen 1996). Studies have shown that changes in the microstructure of chloroplasts in plants may be due to the alpine environment. Therefore, studying the chloroplast genome of species in the plateau region helps to understand the evolution of chloroplasts. (Lütz and Engel 2007).

In this research, the chloroplast genomes of four *H. rhamnoides* subspecies, *H. rhamnoides* subsp. *mongolica* 'wulanshalin', *H. rhamnoides* subsp. *caucasica*, *H. rhamnoides* subsp. *sinensis* 'wucixiong', and *H. rhamnoides* subsp. *yunnanensis* were sequenced, assembled, and annotated. The structure, nucleotide diversity, simple sequence repeat site and codon bias of these 4 subspecies chloroplast genomes were compared. Moreover, a phylogenetic tree was constructed depending on the available chloroplast genome of Elaeagnaceae. The findings of this research will be useful for further studies on resource protection and the taxonomic classification of *Hippophae*. The research results will provide a basis for the effective protection and utilization of seabuckthorn resources.

Materials and methods

Chloroplast DNA extraction and sequencing

The young leaves of each subspecies were taken back to the laboratory in envelopes containing discolored silicone. The samples of *H. rhamnoides* subsp. *yunnanensis* were collected from Kangding County, Sichuan Province (30.04512° N, 101.56306° E), and the other three taxa were collected from the

Desert Forestry Experimental Center of Chinese Academy of Forestry (40.45172° N, 106.76247° E). The CTAB method was used to extract the whole genome DNA of four *H. rhamnoides* subspecies. The purity and concentration were detected by agarose gel electrophoresis and ultra-micro ultraviolet spectrophotometer. Complete random interrupt sequencing of chloroplast genomes of 4 seabuckthorn subspecies was performed using Illumina X-10, Breeding Biotechnologies Company (Shaanxi, China). Finally, at least 5 Gb of clean data were obtained for each sample.

Chloroplast genome assembly and annotation

The chloroplast genome sequence of *H. rhamnoides* (NC_035548) was selected as the reference, and all clean data were aligned to this sequence using Burrow Wheel Aligner (BWA) tool (Li and Durbin 2009). Sequences that could be successfully aligned were extracted from Picard software and assembled by Fast-Plast to obtain contigs. ClustalW (Thompson et al. 2002) was used to align these contigs with the sequences of *H. rhamnoides*, and the end of the invalid sequence was intercepted to obtain the sequence with the best connectivity, which was the final assembly result for later annotation.

CPGAVAS2 online software (Shi et al. 2019) was used to annotate the chloroplast genome. Making the whole genome circular map by using Organellar GenomeDRAW V.1.3.1 (Greiner et al. 2019).

Chloroplast genome Comparison

The chloroplast genome sequences of four subspecies were compared and mapped using the mVISTA program with the LAGAN mode (Frazer et al. 2004). IRscope was used to visually display and compared the boundaries of small single copy (SSC), large single copy (LSC), and inverted repeat (IR) regions in four subspecies of *H. rhamnoides*. (Amiryousefi A 2018).

To detect sequence polymorphism, 113 coding gene sequences and 157 intergenic spacers (IGS) sequences of four taxa are extracted by PhyloSuite ver.1.2.3 (Zhang et al. 2020), and batch comparisons were performed by MAFFT software (Katoh and Standley 2013) using the-auto=strategy and codon alignment model. After this step, nucleotide diversity (Pi) values were calculated for 270 sites using DnaSPver.6.12.03 (Rozas et al. 2017).

SSR locus analysis

Simple sequence repeats (SSRs) were detected by MISA-web online tool (Beier 2017) with the following parameters: The minimum repetition times of single nucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide were 10, 5, 4, 3, 3 and 3, respectively. Set the shortest distance between any two SSR sites to 0 bp, and the c* type was manually removed.

Codon usage analyses

The coding sequence (CDS) of the chloroplast genome was extracted and screened, and the screening conditions were as follows: 1. The total number of bases of each CDS sequence should be an integer

multiple of three; 2. The sequence length should be ≥ 300 bp and the base types in the sequence only include A, T, C, and G; 3. Per sequence includes a start codon and a stop codon; 4. Termination codons were not included in the sequence, and the repetitive sequence is eliminated (Zhou et al. 2008). Finally, 52 CDS sequences were retained for subsequent analysis.

Calculation of RSCU values using the platform JSHYCloud (<http://cloud.genepioneer.com:9929/#/>). Using EMBOSS online software, the GC content of each codon at positions 1, 2 and 3 (namely GC1, GC2 and GC3) was calculated for each gene, and the total GC content (GCall) and effective number of codon (ENC) values for each codon of each gene were determined.

Phylogenetics analysis

To obtain more comprehensive clustering results. We downloaded 16 previously reported chloroplast genomes of Elaeagnaceae (NC _ 058887, LC _ 523635, MN _ 599114, LC _ 522137, LC _ 522136, NC _ 040992, NC _ 036932, NC _ 062127, NC _ 056188, MT _ 512451, MT _ 512450, MW _ 791512, MT _ 512453, NC _ 062882, NC _ 044478 and ON _ 584766) from GenBank and constructed the phylogenetic tree together with these four *H. rhamnoides* subspecies chloroplast genome sequences in this study, and using *Daphne odora* (MT_627479) as outgroup. We use PhyloSuite v1.2.3 software (Zhang et al. 2020) to extract (the CDS sequence), compare, optimize, prune and link sequences, select the MrBayes model in the Partition Finder plug-in, and finally use ITOL software (Letunic and Bork 2021) to view and beautify the constructed tree.

Results and analysis

Characteristics of the four *H. rhamnoides* subspecies chloroplast genome

The chloroplast genomes of four subspecies of *H. rhamnoides* were typical double-stranded tetrads, with the ranged from 157, 437 bp to 157, 822 bp (Fig. 1). The total GC contents were from 36.42 % to 36.53 %, which were consistent with other *Hippophae* chloroplast genomes (Zhou et al. 2020; Wang et al. 2021).

The LSC region length of the four subspecies was 83559-85727 bp, and the GC content was 34.11-34.33%. There were 18,828-18,867bp bases in SSC region, and the GC content ranges from 29.88% to 29.11%. The inverted region size was 26559-26664 bp, and the GC content was 42.42-42.46% (Table 1). The GC content in the IR region was usually higher than the other two regions, probably, due to the presence of four rRNA. (He et al. 2016).

The chloroplast genomes of four subspecies were annotated with 132 genes, including 86 protein-coding genes (PCG), 8 rRNAs, and 38 tRNAs. A total of 15 genes had introns, including 5 tRNA genes (trnA-UGC, trnI-GAU, trnK-UUU, trnL-UAA, and trnV-UAC) and 10 PCGs (rpl2, rpl16, rpoC1, ndhA, ndhB, petB, petD, atpF, clpP, and ycf3), of which ycf3 and clpP have two introns. There were 19 genes duplicated in the IR region, including 7 PCG, 8 tRNA genes, and 4 rRNA genes (Table 2).

The outer gene is transcribed counterclockwise, and the inner gene is transcribed clockwise. Different gene functions have different colors. In the inner circle, the dark part represents GC content, light part represents AT content.

Table 1 Characteristics of chloroplast genomes of four *H. rhamnusus* subspecies

Species	ALL		LSC		SSC		IR	
	Length(bp)	GC%	Length(bp)	GC%	Length(bp)	GC%	Length(bp)	GC%
HM	157822	36.51	85728	34.29	18846	29.91	26624	42.43
HC	157437	36.5	85491	34.25	18828	29.89	26559	42.46
HS	157515	36.42	85359	34.11	18840	29.91	26658	42.42
HY	157736	36.53	85541	34.33	18867	29.88	26664	42.42

Notes: HM: *H. rhamnoides* subsp. *mongolica* 'wulanshalin'; HC: *H. rhamnoides* subsp. *caucasica*; HS: *H. rhamnoides* subsp. *sinensis* 'wucixiong'; HY: *H. rhamnoides* subsp. *Yunnanensis*. This acronym also applies in the following tables and pictures.

Table 2 The chloroplast genome genes of the four *H. rhamnusus* subspecies

Gene function	Gene category	Gene name
Self-replication	The large subunit of the ribosome	rpl2(×2)*, rpl14, rpl16*, rpl20, rpl22, rpl23(×2), rpl32, rpl33, rpl36
	The small subunit of the ribosome	rps2, rps3, rps4, rps7(×2), rps8, rps11, rps12(×2), rps14, rps15, rps16, rps18, rps19
	RNA polymerase	rpoA, rpoB, rpoC1*, rpoC2
	Ribosomal RNAs	rrn4.5(×2), rrn5(×2), rrn16(×2), rrn23(×2)
	Transfer RNAs	trnA-UGC(×2)*, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnFM-CAU, trnG-GCC*, trnG-UCC, trnH-GUG(×2), 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
Photosynthesis	Photosystem I	psaA, psaB, psaC, psal, psaJ
	Photosystem II	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ
	NADPH dehydrogenase	ndhA*, ndhB(×2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK
	Cytochrome b/f complex	petA, petB*, petD*, petG, petL, petN
	ATP synthase	atpA, atpB, atpE, atpF*, atpH, atpI
	RubisCO large subunit	rbcL
Other genes	Protease	clpP**
	Maturase	matK
	Hypothetical	ycf1(×2), ycf2(×2), ycf3**, ycf4
	Other genes	ccsA, cemA, accD, infA

Note: *: gene has an intron; * *: gene has two introns; x2: repeated genes in the IR region.

Comparison analysis of chloroplast genomes

To compare the differences in chloroplast genome sequences of 4 taxa of sea buckthorn, the mVISTA software with Shuffle-LAGAN mode was used to compare the annotation sequence of the four *H. rhamnoides* subspecies, *H. rhamnoides* (KY_794808) as the reference sequence (Fig. 2). The results displayed that these taxa had relatively conservative chloroplast genome sequences, but some variations

were detected. The IR region was more conservative than the LSC and SSC regions, and the CDS was more conservative than the non-coding sequence (CNS), while the differences were mainly located in the intergenic regions of LSC and SSC regions, including trnSGCU-trnGGCC, trnCGCA-petN, psbZ-trnGUCC, rps4-ndhJ, rpl33-rps18 and ndhF-trnLUAG, protein-coding genes ycf1. These different sequences can be used as a possible basis for the identification of plant taxa.

To determine the chloroplast genome polymorphic loci of four subspecies. We figured up the nucleotide diversity (P_i) values of genes and IGS. In the gene coding region, the average P_i of the common genes was 0.002634 (range from 0 to 0.02083, Fig. 3 a). A total of 5 mutation sites ($P_i > 0.01$) were found, including petN, petL, psaJ, infA, and rpl32. In the IGS, the average P_i value was 0.010998 (ranging from 0 to 0.36391, Fig. 3 b), and there were 9 mutation sites in these spacers ($P_i > 0.025$), including ccsA-ndhD, ndhB-rps7, ndhD-psaC, ndhI-ndhA2, psbT-psbN, rrn4.5-rrn5, trnLUAG-ccsA, and trnSGCU-trnGGCC1.

Black arrows represent gene direction; the purple stripe represents the exon; the blue stripe represents rRNA; the pink stripe represents CNS; the gray bar represents mRNA; the white peak represents the difference of gene sequence; the y-axis represents the percentage identity (range display: 50-100 %).

IR contraction and expansion

As mentioned above, the IR region was the most conserved. In the process of species evolution, the chloroplast genome will also undergo partial variation, and the corresponding boundary will expand or shrink. Therefore, we compared the SSC/IR and LSC/IR boundaries among these chloroplast genomes (Fig. 4). Founded, six genes, ycf1, ndhF, trnH-GUG, rpl2, rpl22 and rps19, were located at the SSC/IR and LSC/IR borders. In the four subspecies, the ycf1 gene crosses the SSC/IRa boundary, resulting in incomplete replication of the gene at the SSC/IRb boundary, while the partially replicated ycf1 gene crosses the SSC/IRb boundary and intertwines with the ndhF gene extending in the SSC region. The rpl22 gene of the four taxa was located in the LSC region of the LSC/IRb boundary and the trnH-GUG gene and rpl2 gene were replicated in the IR region near the LSC/IR boundary. Among the four subspecies, the rps19 gene in the chloroplast genome of *H. rhamnoides* subsp. *sinensis* 'wucixiong' was 43 bp away from the JLB boundary, while the remaining three taxa extended from the IRb region to the LSC region by 219-221 bp. In general, the IR boundary regions of the four taxa had similar features, with only slight differences in the flanks of the boundary or the length across.

Codon usage bias analyses

Different codon frequencies among synonymous codons were referred to as codon use bias (CUB). Understanding molecular evolution, environmental adaptation, and genetic characteristics can be aided by CUB analysis. Firstly, for 52 PCG, we determined the relative synonymous codon usage (RSCU) value. The RSCU (Fig. 5) results showed that the four subspecies had similar codon preferences. In addition to the stop codon, the codon UUA encoding leucine was the highest, followed by the codon AGA encoding

arginine, the AGC codon, which encodes serine acid, had the lowest usage frequency methionine and tryptophan has no codon preference. Most codons tended to end with A/U, which was the same as found in other chloroplast genomes (Chi et al. 2020; Zeng et al. 2023). Then we studied the base composition of codons at each position in four subspecies, showing GC1 > GC2 > GC3, indicating that the third base of the codon was mainly A/T ending (Table 3). The ENC of these subspecies' chloroplast genes ranged from 48.735% to 48.783%, which implied that the CUB of genes was relatively weak and the endogenous gene expression was low (Table 3).

Table 3 Analysis of effective codon number and GC content in different positions of codons

	GCall %	GC1 %	GC2 %	GC3 %	ENC
HM	37.28	45.63	37.58	28.62	48.783
HC	37.25	45.62	37.59	28.54	48.735
HS	37.28	45.64	37.58	28.62	48.735
HY	37.28	45.67	37.56	28.59	48.778

SSR analysis

Six types of SSR (mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide) were detected by MISA analysis of four *H. rhamnoides* subspecies chloroplast genome sequences (Fig. 6A). The total numbers of the SSRs were 100, 103, 110, and 110 in *H. rhamnoides* subsp. *mongolica* 'wulanshalin', *H. rhamnoides* subsp. *caucasica*, *H. rhamnoides* subsp. *sinensis* 'wucixiong', and *H. rhamnoides* subsp. *yunnanensis*, respectively (Fig. 6 a). The main types of SSRs were mononucleotide repeats, accounting for 76.3%, and the majority of them tended to use A/T, followed by the dinucleotide repeat unit (AT/TA), accounting for 13.9%. Meanwhile, three hexanucleotide repeats, GAAGGG, TAAAAA and AAAAAT, were detected only in *H. rhamnoides* subsp. *yunnanensis*. In the meantime, AAATA and AAAT sequences were detected in the chloroplast genomes of *H. rhamnoides* subsp. *mongolica* 'wulanshalin' and *H. rhamnoides* subsp. *caucasica*, respectively. We also found that these SSRs were mainly located in the LSC region or CDS. (Fig. 6 b, Fig. 6 c). The SSRs will provide molecular markers for studying the genetic diversity and population structure of sea buckthorn.

Phylogenetic analysis

In this study, we downloaded 16 previously reported chloroplast genomes of Elaeagnaceae from NCBI and constructed the phylogenetic tree with these four *H. rhamnoide* subspecies chloroplast genome sequences, using *D. odora* (MT_627479) as an outgroup. The results showed that *Elaeagnus* and *Hippophae* formed two monophyletic sister groups with high support rates. *Hippophae* was composed of two evolutionary clades. In one of these, *H. gyantsensis* and *H. neurcar* were first gathered together, and

then they were clustered with *H. salicifolia*. In another, *H. tibetana*, *H. rhamnoides* and six *H. rhamnoides* subspecies get gathered. The *H. rhamnoides*, could be divided into three small evolutionary clades. Among them, *H. rhamnoides* subsp. *yunnanensis* clustered into a single clade. *H. rhamnoides* subsp. *mongolica* and *H. rhamnoides* subsp. *sinensis* were clustered together to form a small clade. The other three subspecies, *H. rhamnoides* subsp. *caucasica*, *H. rhamnoides* subsp. *rhamnoides* and *H. rhamnoides* subsp. *turkestanica*, formed another evolutionary clade. This study clarified the positional relationship of Elaeagnaceae at the chloroplast genome level and provided a reference for the subsequent classification of *Hippophae*.

Discussion

In this study, we comprehensively analyzed the chloroplast genomes of four *H. rhamnoides* subspecies. The gene order and structure of the four taxa were similar. The results revealed that the chloroplast genomes were consistent with most angiosperms and were a typical tetrad structure, including a large single-copy region (LSC), a small single-copy region (SSC) and two reverse repeat regions (IR) with the same sequence and opposite directions, which consisted of 86 PCG, 38 tRNA and 8 rRNA, with a genome size of 157,437 – 157,822 bp and GC% content of 36.42–36.53%.

The four taxa's genomes were shown to be relatively conserved by comparative genome analysis using mVISTA. However, LSC and SSC areas were more divergent than the IR regions, and the non-coding portions were more divergent than the coding regions. Similar outcomes have been previously reported for numerous chloroplast genomes (Huo et al. 2019). There were still several issues with the identification of related species, particularly with the identification of species with unclear morphology. Therefore, finding appropriate DNA markers for these species was crucial. Fortunately, these divergences between chloroplast segments can be converted into corresponding molecular markers to distinguish different *Hippophae* taxa. Due to their slower evolution than nuclear genomes, the chloroplast genomes were frequently used for phylogenetic investigations and species identification. (Song et al. 2017). In this study, the IGS (ccsA-ndhD, ndhB-rps7, ndhD-psaC, ndhI-ndhA2, psbT-psbN, rrn4.5-rrn5, trnLUAG-ccsA and trnSGCU-trnGGCC1) and genes (petN, petL, psaJ, infA, and rpl32) with high variation were recognized. These high variability regions, it has potential values as an effective DNA barcode marker for the authentication of sea buckthorn in the future.

The expansion and contraction of IR boundaries were general in the chloroplast genome, which made the whole chloroplast genome of the plant populations different (Guo et al. 2022). In this research, the IR boundaries of the four *H. rhamnoides* subspecies were compared. Among them, *H. rhamnoides* subsp. *caucasica* was more obvious, and its rps19 gene didn't expand to IRb (42bp from the JLB boundary), which indicated that the IR region of *H. rhamnoides* subsp. *caucasica* has a more obvious contraction than the other three taxa. This contraction probably be the primary cause why the length of the chloroplast genome of *H. rhamnoides* subsp. *caucasica* was shorter than that of the other three subspecies. The ycf1 gene crosses the IRA/SSC boundary, resulting in a 5' deletion in the IRb region, generating the ycf1 pseudogene. Therefore, the ycf1 gene segment can be used to identify sea buckthorn.

Codon usage bias was a crucial evolutionary feature in the genome that contributes to our understanding of how genetic information is transmitted among species and is useful for studying the evolution, gene function, and foreign gene expression of organisms (Quax et al. 2015). The findings showed that four subspecies' chloroplast genomes had generally comparable patterns of codon use. Consistent use implied great conservation, according to the mode. They had a bias for A/U-ending synonymous codons in particular. Other angiosperm species, including Fabaceae, Euphorbiaceae, and Solanaceae, were also found to exhibit this codon usage pattern (Nie et al. 2014; wang et al. 2020; Mehmood et al. 2020). Moreover, among 64 synonymous codons, UUA, encoding leucine was the highest, followed by the codon AGA encoding arginine, and AGC encoding serine had the lowest usage frequency. At the same time, the ENC of the chloroplast genomes of these taxa was all less than 35, showing that their codon preference was weak and the expression of endogenous genes was low. In the long-term evolution of species, codon preference was influenced by natural selection, base mutation, genetic drift and selective translation, which was a result of the long-term evolution of species' response to environmental selection (Bhattacharyya et al. 2019).

SSR was often as a molecular marker to investigate the degree of evolution among species. In the SSR analysis, a total of 423 SSR loci were found, most of which were located in IGS and LSC. The SSRs identified in the chloroplast genomes of the four subspecies can be used for the construction of genetic maps, genetic diversity analysis, and identification of sea buckthorn plants.

Chloroplast genomes were often used to infer developmental and evolutionary relationships among species due to their stability and conservation. In this study, *D. odora* was the outgroup, and the phylogenetic tree was constructed by the MrBayes method. The results showed that *H. tibetan* and *H. rhamnoides* were clustered into one branch, and *H. gyantsensis*, *H. neurocarpa* and *H. salicifolia* were clustered into another, which was consistent with the results of Wang (2020). According to the classical taxonomy, *H. salicifolia* and *H. rhamnoides* were classified as Sect. *Hippophae*, *H. tibetica*, *H. gyantse*, and *H. neurocarpa* were classified as Sect. *Gyantsenses* (Lian and Chen 1996). The study on the origin of *Hippophae* showed that the plants of this genus might originate from the middle and high-altitude areas from the eastern Himalayas to the Hengduan Mountains. With the uplift of the Himalayas and the ridge of the Qinghai-Tibet Plateau, the alpine climate was cold. It may be due to the effect of cold that the fruit-related traits of sea buckthorn have changed. The phylogenetic tree constructed by the chloroplast genome might overlap with its geographical distribution, resulting in frequent gene exchange. The study found that the six subspecies of *H. rhamnoides* were clustered into one clade, which agreed with the classical taxonomy, indicating the feasibility of chloroplast genome analysis in the study. Among the six subspecies, *H. rhamnoides* subsp. *yunnanensis* was clustered into a single clade. *H. rhamnoides* subsp. *sinensis* and *H. rhamnoides* subsp. *mongolica* were clustered into a clade. The other three subspecies, *H. rhamnoides* subsp. *caucasica*, *H. rhamnoides* subsp. *rhamnoides* and *H. rhamnoides* subsp. *turkestanica*, formed another evolutionary clade. This result is consistent with the phylogenetic tree constructed based on the ITS sequence (Ma et al. 2014). We will further study the phylogenetic relationships of *Elaeagnus* in the future.

Conclusion

In the present study, the chloroplast genomes of four *H. rhamnoides* subspecies were characterized. Their genome length of them ranged from 157,436 bp to 157,822 bp, with a typical quadripartite structure. The annotation contained a total of 132 genes in each chloroplast genomes genome, with 86 protein-coding genes, 38 tRNA genes, and 8 rRNA genes. Fourteen highly differentiated regions and 423 SSR loci were identified, which can be used as potential molecular markers for *H. rhamnoides*. Meanwhile, the Phylogenetic tree of 21 taxa of the family Elaeagnaceae was constructed by MrBayes based on CDS sequence. The findings of this research will be useful for further studies on resource protection and the taxonomic classification of sea buckthorn.

Declarations

Author's contributions

XQ and HYZ conceived and designed the project. WRQ and WXJ analyzed the data. WRQ drafted the original manuscript. LZH and XQ provided financial support and reviewed and revised the paper.

Funding

This research was supported by the Starting Fund for Doctoral Research of Shanxi Agricultural University (2015YJ15 and 2015YJ16), the Shanxi Province Applied Basic Research Program (201901D211360), and the Shanxi Province Basic Research Program (202103021224141).

Conflict of interest

The authors declare no competing interests.

References

1. Amiryoussefi A, Hyövänen J, Pocai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34 (17): 3030–3031. <https://doi.org/10.1093/bioinformatics/bty220>
2. Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33 (16): 2583–2585. <https://doi.org/10.1093/bioinformatics/btx198>
3. Bhattacharyya D, Uddin A, Das S, Chakraborty S (2019) Mutation pressure and natural selection on codon usage in chloroplast genes of two species in *Pisum* L. (Fabaceae: Faboideae). *Mitochondrial DNA Part A* 30 (4): 664–673. <https://doi.org/10.1080/24701394.2019.1616701>
4. Chi XF, Zhang FQ, Dong Q, Chen SL (2020) Insights into comparative genomics, codon usage bias, and phylogenetic relationship of species from Biebersteiniaceae and Nitrariaceae based on complete chloroplast genomes. *Plants* 9 (11): 1605. <https://doi.org/10.3390/plants9111605>

5. Dong W, Xu C, Li C, Sun J, Zuo Y et al. (2015) *ycf1*, the most promising plastid DNA barcode of land plants. *Sci Rep* 5:8348. <https://doi.org/10.1038/srep08348>
6. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32(W): 273-279. <https://doi.org/10.1093/nar/gkh458>
7. Gâtlan AM, Gutt G (2021) Sea Buckthorn in Plant Based Diets. An Analytical Approach of Sea Buckthorn Fruits Composition: Nutritional Value, Applications, and Health Benefits. *Int J Environ Res Public Health* 18 (17): 8986. <https://doi.org/10.3390/ijerph18178986>
8. Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 47 (W1): W59-64. <https://doi.org/10.1093/nar/gkz238>
9. Gong L, Ding XX, Guan W, Zhang DC, Zhang J et al. (2022) Comparative chloroplast genome analyses of *Amomum*: insights into evolutionary history and species identification. *BMC Plant Biol* 22 (1): 520. <https://doi.org/10.1186/s12870-022-03898-x>
10. Guo RX, Guo XB, Li Tong, Fu X, Liu RH (2017) Comparative assessment of phytochemical profiles, antioxidant and antiproliferative activities of Sea buckthorn (*Hippophae rhamnoides* L.) berries. *Food Chemistry* 221: 997-1003. <https://doi.org/10.1016/j.foodchem.2016.11.063>
11. Guo S, Liao XJ, Chen SY, Liao BS, Guo YM (2022) A Comparative analysis of the chloroplast genomes of four *Polygonum* medicinal plants. *Front Genet* 13:764534. <https://doi.org/10.3389/fgene.2022.764534>
12. He Y, Xiao HT, Deng C, Xiong L, Yang J (2016) The complete chloroplast genome sequences of the medicinal plant *Pogostemon cablin*. In. *J Mol Sci* 17 (6): 820. <https://doi.org/10.3390/ijms17060820>
13. Huang Q (2007) *Hippophae* breeding and cultivation. Science Press, Beijing
14. Huo YM, Gao LM, Liu BJ, Yang YY, Kong SP et al. (2019) Complete chloroplast genome sequences of four *Allium* species: comparative and phylogenetic analyses. *Sci Rep* 9 (1): 12250. <https://doi.org/10.1038/s41598-019-48708-x>
15. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW 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 USA* 104 (49): 19369–19374.
16. Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30 (4): 772-780. <https://doi.org/10.1093/molbev/mst010>
17. Lee KJ, Lee GA, Lee JR, Sebastin R, Shin MJ et al. (2019) Genetic Diversity of Sweet Potato (*Ipomoea batatas* L. Lam) Germplasms Collected Worldwide Using Chloroplast SSR Markers. *Agronomy-Basel* 9 (11): 752. <https://doi.org/10.3390/agronomy9110752>
18. Letunic I, Bork P (2021) Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49 (W1): 293-296. <https://doi.org/10.1093/nar/gkab301>
19. Lian YS, Chen XL (1996) Systematic classification of *Hippophae*. *Hippophae* 9 (1): 15-24.

20. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14): 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>
21. Li H, Ruan CJ, Teixeira da Silva JA (2009) Identification and genetic relationship based on ISSR analysis in a germplasm collection of sea buckthorn (*Hippophae* L.) from China and other countries. *Sci Hortic* 123 (2): 263-271. <https://doi.org/10.1016/j.scienta.2009.08.007>
22. Lütz C, Engel L (2007) Changes in chloroplast ultrastructure in some high-alpine plants: adaptation to metabolic demands and climate? *Protoplasma* 231 (3-4): 183–192. <https://doi.org/10.1007/s00709-007-0249-8>
23. Ma QG, Guan Y, Song ZP, Dong JH, Wei RR (2022) Isolation and characterization of auronlignan derivatives with hepatoprotective and hypolipidemic activities from the fruits of *Hippophae rhamnoides* L. *Food func* 13 (14): 7750-7761. <https://doi.org/10.1039/D2FO01079H>
24. Ma YH, Ye GS, Xiang QS, Gao Y, Yang CJ et al. (2014) Phylogenetic relationships of sea buckthorn based on ITS sequences. *Chinese Journal of Applied Ecology* 25 (10): 2985-2990. <https://doi.org/10.13287/j.1001-9332.20140731.009>
25. Mehmood F, Abdullah I, Shahzadi I, Ahmed I, Waheed MT et al. (2020) Characterization of *Withania Somnifera* chloroplast genome and its comparison with other selected species of Solanaceae. *Genomics* 112 (2): 1522–1530. <https://doi.org/10.1016/j.ygeno.2019.08.024>
26. Neuhaus HE, Emes MJ (2000) Nonphotosynthetic metabolism in plastids. *Annu Rev Plant Physiol Plant Mol Biol* 51 (1): 111–140. <https://doi.org/10.1146/annurev.arplant.51.1.111>
27. Nie X, Deng P, Feng K, Liu P, Du X et al. (2014) Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. *Plant Mol Biol Rep* 32 (4): 828–840. <https://doi.org/10.1007/s11105-013-0691-z>
28. Quax TEF, Claassens NJ, Soll D, van der Oost J (2015) Codon bias as a means to fine-tune gene expression. *Molecular Cell* 59 (2): 149-161. <https://doi.org/10.1016/j.molcel.2015.05.035>
29. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P et al. (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution* 34 (12): 3299-3302. <https://doi.org/10.1093/molbev/msx248>
30. Shi LC, Chen HM, Jiang M, Wang LQ, Wu X et al. (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res* 47 (W1): W65-W73. <https://doi.org/10.1093/nar/gkz345>
31. Song Y, Chen Y, Lv J, Xu J, Zhu S et al. (2017) Development of chloroplast genomic resources for *Oryza* species discrimination. *Frontiers in Plant Science* 8: 1854. <https://doi.org/10.3389/fpls.2017.01854>
32. Sugiura M (1992) The chloroplast genome. *Plant Molecular Biology* 19 (1): 149-168. <https://doi.org/10.1007/BF00015612>
33. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequences alignment using ClustalW and ClustalX. *Curr protoc Bioinform*. <https://doi.org/10.1002/0471250953.bi0203s00>
34. Wang L, Wang J, He C, Zhang J, Zeng Y (2021) Characterization and comparison of chloroplast genomes from two sympatric *Hippophae* species (Elaeagnaceae). *J For Res* 32 (01): 307-318.

<https://doi.org/10.1007/s11676-019-01079-5>

35. Wang J (2020) Study on the evolution of chloroplast genome of *Hippophae*. Dissertation, Chinese Academy of Forestry
36. Wang ZJ, Xu BB, Li B, Zhou QQ, Wang GY (2020) Comparative analysis of codon usage patterns in chloroplast genomes of six *Euphorbiaceae* species PeerJ 8: e8251.
<https://doi.org/10.7717/peerj.8251>
37. Yuan W (1988) *Hippophae rhamnoides* subsp. *yunnanensis* resources and their ecological distribution. Journal of Yunnan Agricultural University 3 (02): 218-219.
[https://doi.org/10.16211/j.issn.1004-390x\(n\).1988.02.022](https://doi.org/10.16211/j.issn.1004-390x(n).1988.02.022)
38. Zeng YJ, Shen LW, Chen SQ, Qu S, Hou N (2023) Codon usage profiling of chloroplast genome in Juglandaceae. Forests 14 (2): 378. <https://doi.org/10.3390/f14020378>
39. Zhang D, Gao FL, Jakovlić I, Zou H, Zhang J et al. (2020) PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour.20 (1): 348-355. <https://doi.org/10.1111/1755-0998.13096>
40. Zhou M, Long W, Li X (2008) Analysis of synonymous Codon usage in chloroplast genome of *Populus alba*. J For Res 19 (4): 293-297. <https://doi.org/10.1007/s11676-008-0052-1>
41. Zhou W, Dong Q, Wang HL, Hu N (2020) Characterization of the complete chloroplast genome of *Hippophae tibetana*. Mitochondrial DNA Part B-Resources 5 (1): 593-594.
<https://doi.org/10.1080/23802359.2019.1710595>.

Figures

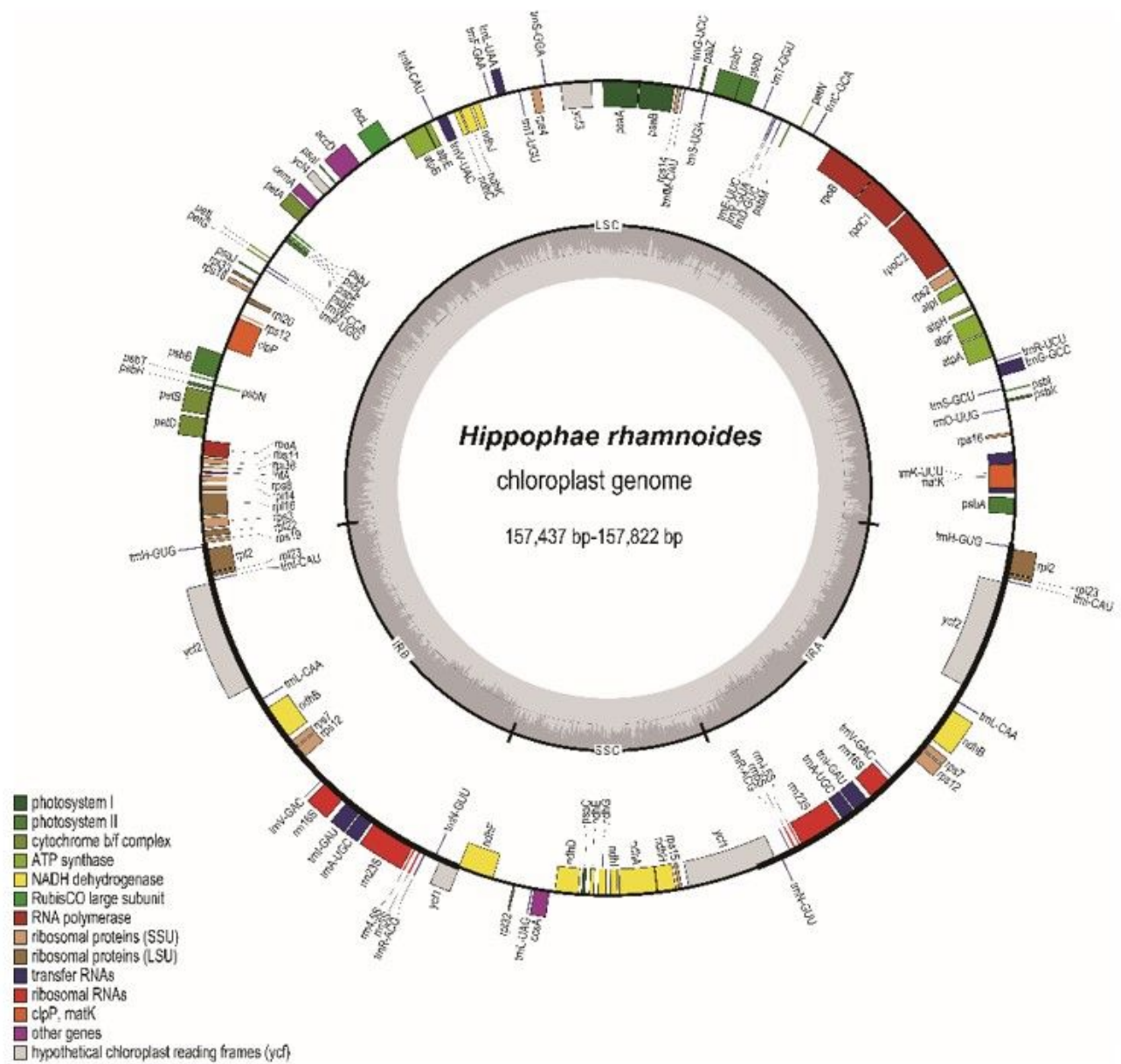


Figure 1

A circular map of chloroplast genomes of four *H. rhamnoides* subspecies

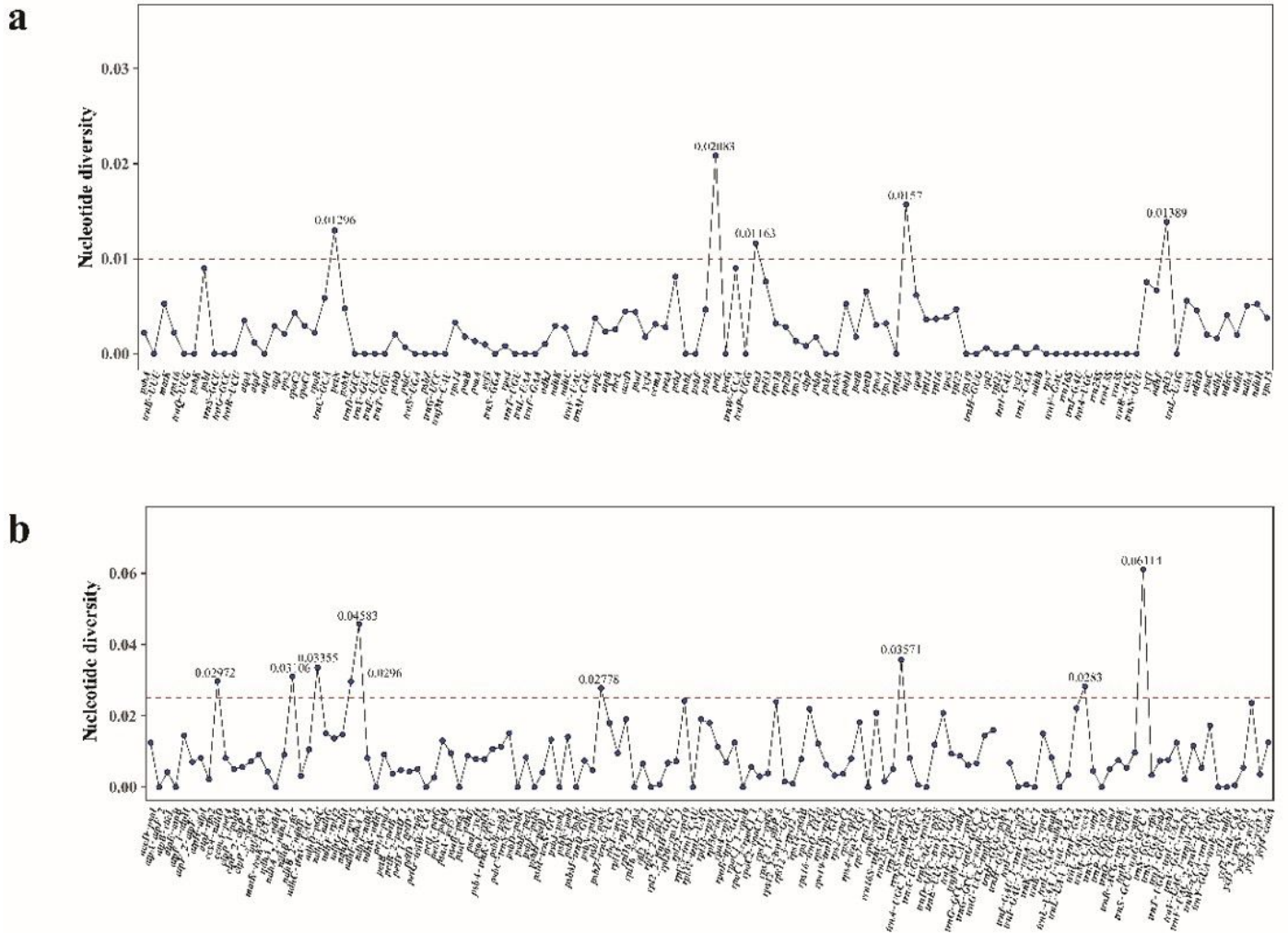
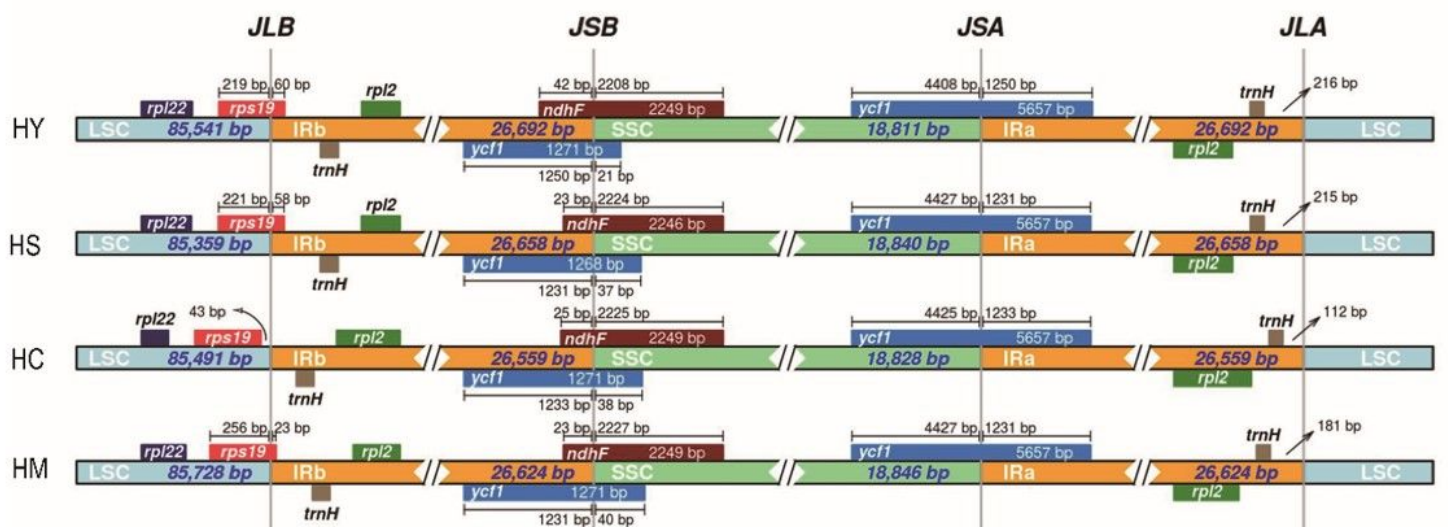


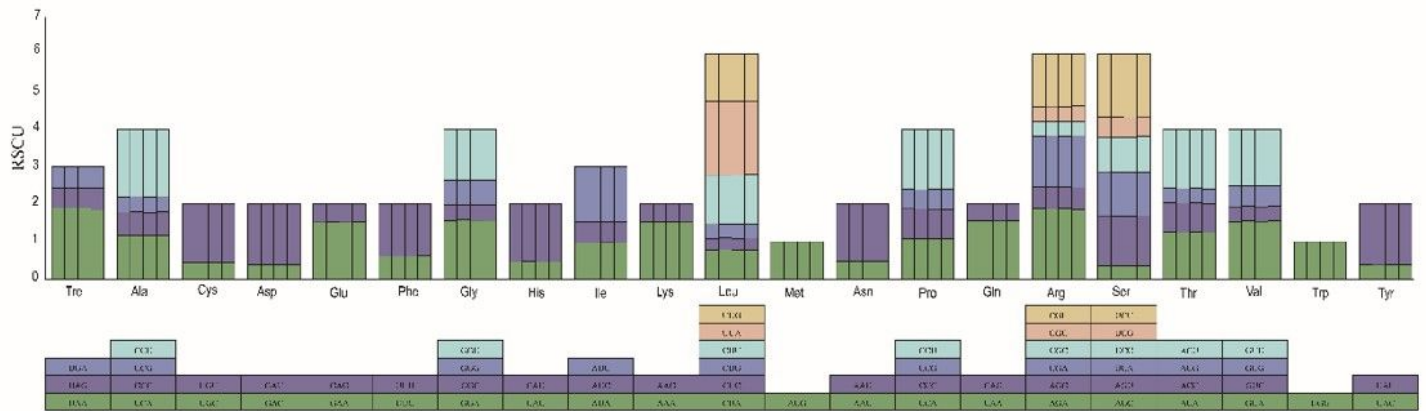
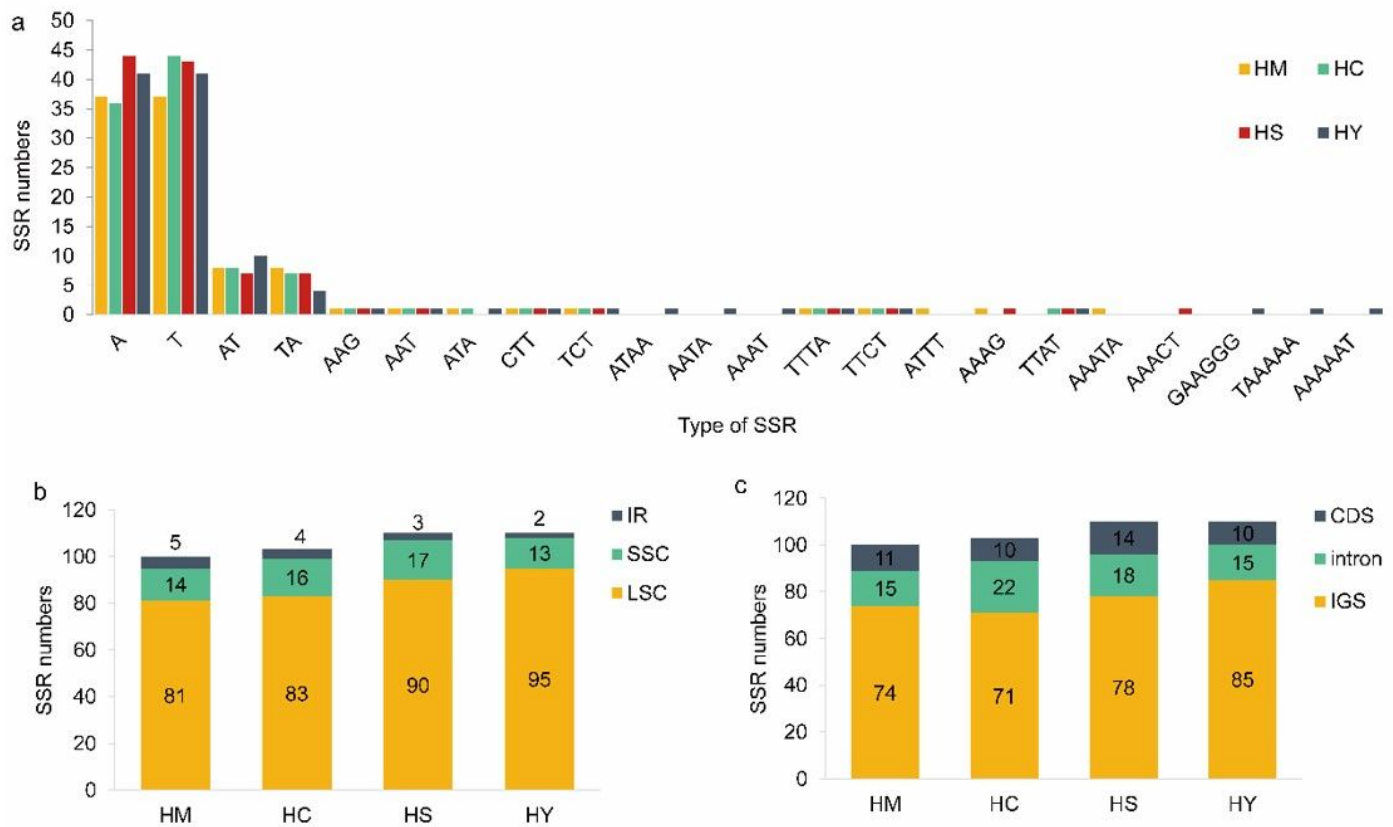
Figure 3

The nucleotide variability (Pi) values among four *H. rhamnoides* subspecies

a: The P-distance value of genes. b: The P-distance value of IGS.



Comparative analysis of IR boundary regions of four *H. rhamnosus* subspecies

Chloroplast genome RSCU of four *H. rhamnoides* subspecies

The distribution of SSRs detected in the chloroplast genomes of four *H. rhamnoides* subspecies

a: the number of SSR types; b: the number of SSRs in LSC, SSC, and IR regions; c: The number of SSRs in the CDS, intron region, and IGS.

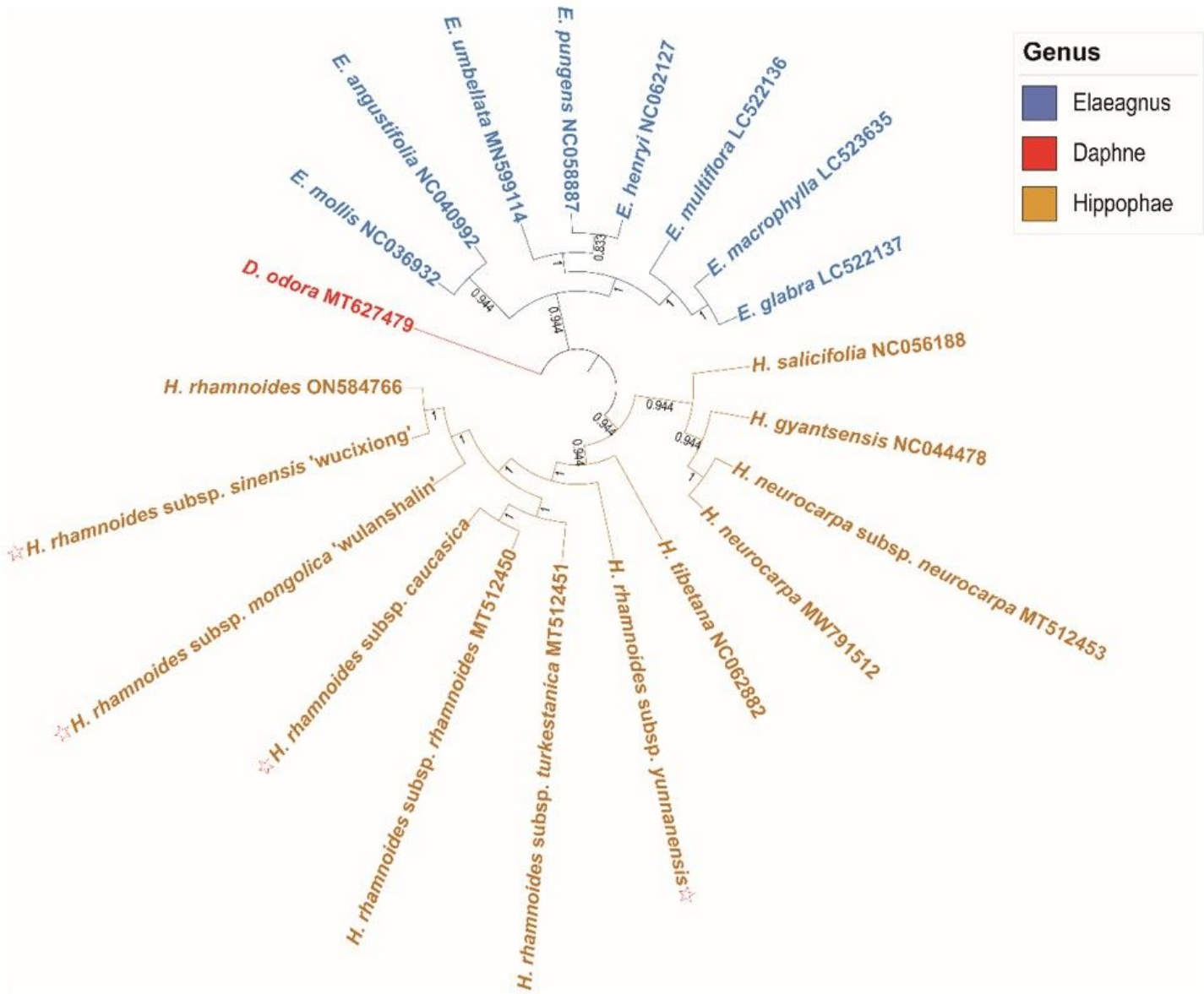


Figure 7

Phylogenetic tree of 21 taxa of the family Elaeagnaceae using MrBayes inference based on CDS sequence