Journal of Animal & Plant Sciences, 33(2): 2023, Page: 462-473
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2023.2.0638

COMPARISON AND PHYLOGENETIC ANALYSIS OF THE CHLOROPLAST GENOMES OF *SABINA PRZEWALSKII* F. *PENDULA* AND *JUNIPERUS PRZEWALSKII*

S. P. Si[1], J. Z. Wan[2], L. H. Yang[3], Z. Z. Wu[4], B. Xiong[5], C. H. Zhang[2], L. Wang[1,2*] and C. J. Wang[1*]

[1]College of Agriculture and Animal Husbandry, Qinghai University, Xining 810016, China
[2]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, China
[3]College of Continuing Education, Qinghai University, Xining 810016, China
[4]Beishan Forest Farm, Haidong 810599, China
[5]College of Tea Science, Guizhou University, Guiyang 550000, China
*Corresponding Author's E-mail: *Chun-Jing Wang: wangchunjing00@163.com
*Le Wang: wangleqhu@163.com

ABSTRACT

Most species of the family Cupressaceae are widely distributed throughout China, but *Sabina przewalskii* f. *pendula* and *Juniperus przewalskii* are endemic to the Qinghai-Tibet Plateau region. Therefore, they are of great significance to the plateau ecosystem. *Sabina przewalskii* and *J. przewalskii* have similar characteristics, but they be distinguished by the presence or absence of drooping branchlets. However, this feature can be difficult to distinguish with the naked eye. In this study, the complete chloroplast (CP) genomes of *S. przewalskii* and *J. przewalskii* were determined by high-throughput sequencing, aiming to unveil their differences at the molecular level. The sizes of the CP genomes of *J. przewalskii* and *S. przewalskii* are 128,748 bp and 127,315 bp, respectively, sharing the typical structure of conifer species. The GC content of the *S. przewalskii* plastome is 35.08%, encoding 84 proteins, 32 tRNAs and 4 rRNAs. The GC content of the *J. przewalskii* CP genome is 34.98%, and one additional *rpl32* gene was detected. Phylogenetic analysis proved that *J. przewalskii* showed a close relationship to *Juniperus tibetica* and *S. przewalskii*, both of which belong to the family Cupressaceae. Our results of the plastome of *J. przewalskii* and *S. przewalskii* further clarify the evolution of Cupressaceae species. The divergent regions and the repetitive sequences identified could be developed as molecular markers that would benefit breeding and species discrimination.

Keywords Chloroplast genomes, Cupressaceae, *Juniperus przewalskii*, *Sabina przewalskii* f. *pendula*, Phylogenetic analysis

Published first online March 02, 2023 Published final March 24, 2023

INTRODUCTION

Sabina przewalskii f. *pendula* W.C.Cheng & L.K.Fu.and *Juniperus przewalskii* Kom. are evergreen trees in the family Cupressaceae and are similar in phenotype (http://www.theplantlist.org/; Editorial Committee of Woody Flora of Qinghai Province, 1987). These two tree species are increasingly endangered, due to logging and deforestation (Yang and Liao, 2013). *Juniperus przewalskii* is endemic to the Qinghai-Tibet Plateau (Zhang *et al*., 2005), and is mainly distributed in the northeast region from 2,900 to 3,760 m in altitude (Editorial Committee of Woody Flora of Qinghai Province, 1987). Previously, *S. przewalskii* f. *pendula* were classified as *S. przewalskii* and then was renamed as *J. przewalskii*, but there was no basis for this change. Morphologically, the main difference between *S. przewalskii* f. *pendula* and *J. przewalskii* is that *J. przewalskii* has no drooping branchlets. This difference in morphology could be caused by complicated factors, such as wood formation (Hollender and Dardick, 2015), alleles (Hollender *et al*., 2018) or plant hormones (Mao *et al*., 2020).

Photosynthesis occurs in the chloroplasts (CPs) of green plants. CP genomes are non-meiotic with mostly uniparental inheritance (Drouin *et al*., 2008; Redwan *et al*., 2015). In most angiosperms, the CP gene is a circular structure composed of four regions, including two inverted repeat (IR) regions, a large single-copy (LSC) region and a small single-copy (SSC) region (Kerstin *et al*., 2009). Generally, CPs are approximately 120–160 kb and contain 120–130 genes (Lei *et al*., 2016). Therefore, CP genomes have been used as essential resources for phylogenic studies (Song *et al*., 2019). The emergence and advancement of high-throughput sequencing technology has promoted development in plant genetics and genomics (Daniell *et al*., 2016). Since the first CP genome from tobacco (*Nicotiana tabacum*) was sequenced in 1986, a number of CP genomes have been sequenced (Shinozaki *et al*., 1986), which is beneficial to

the study of taxonomy and evolution. Plastome-based phylogenies have been used for the exploration of evolutionary relationships among various species (Daniell *et al*., 2016; Xu *et al*., 2020).

In the present study, we sequence and assemble the complete CP genomes of *J. przewalskii* and *S. przewalskii* f. *pendula*, and construct a plastome-based phylogeny to infer their relationship. Our results provide plastome references needed for species discrimination and phylogenic inference of *Juniperus* species.

MATERIALS AND METHODS

Material sampling and DNA extraction: In this study, genomic DNA was extracted from fresh leaves of *J. przewalskii* and *S. przewalskii* f. *pendula*, both of which were collected from Hush Beishan National Forest Park in Qinghai Province, China (36° 5′ 26.48″ N, 102° 20′ 49.31″ E) in December 2020. The fresh leaves were washed and kept in liquid nitrogen and were brought back to the lab in College of Agriculture and Animal Husbandry, Qinghai University, China. Extractions of genomic DNA were performed by using the Plant Genomic DNA Prep Kit (Takara, Dalian, China).

Chloroplast genome sequencing, assembly and annotation: Genomic DNA was subjected to library construction following the manufacture's protocol (Illumina Inc. San Diego, CA, USA), and the paired-end reads (150 bp) were sequenced using the Illumina NovaSeq platform (Borgstrom *et al*., 2011). Filtering of the sequencing data was performed with the SOAPnuke v1.3.0 (Chen *et al*., 2018). Clean reads were then mapped to the reference [*Juniperus recurva* (accession no. NC_042763.1)] to extract CP-like reads, using the BLASTN software (v. BLAST 2.2; Parameter: E-value $1e^{-5}$). The CP-like reads were then subjected to de novo assembling using SPAdes v3.6.1 (Bankevich *et al*., 2012). Drafts of the chloroplast genomes of *J. przewalskii* and *S. przewalskii* f. *pendula* were constructed with Unicycler v0.4.8 (Wick *et al*., 2017). Gene annotation was performed using GeSeq (Tillich *et al*., 2017). All annotations were checked manually by BLAST. Chloroplot software (Zheng *et al*., 2020) was used to draw circular maps of the *S. przewalskii* f. *pendula* and *J. przewalskii* chloroplast genomes.

Long repeats, SSRs and genome comparison: MISA software (v. 1.0; default parameters; the minimum repetition of each repeating unit was 1–8, 2–4, 3–4, 4–3, 5–3 and 6–3; 1–8 represents a mononucleotide repeating unit in which at least eight repeats were detected) was used to detect SSRs in the CP genomes of *S. przewalskii* f. *pendula* and *J. przewalskii* (Thiel *et al*., 2003； Beier *et al*., 2017). Long-repeat fragments in the CP genome were distinguished by using Vmatch (http://www.vmatch.de/; parameter: minimal repeat size 30 bp). To visualize the structural variation, CP genomes of *J. przewalskii*, *S. przewalskii* f. *pendula*, *J. communis* (MH121052.1), *J. squamata* (NC 044076.1), *J. tibetica* (NC 041523.1) and *Pinus tabuliformis* (NC 028531.1) were compared using the mVISTA (Mayor *et al*., 2000) program in the Shuffle-LAGAN mode (Frazer *et al*., 2004).

Phylogenetic analysis: The chloroplast genome sequences of *J. przewalskii* and *S. przewalskii* f. *pendula*, together with sequences from 16 other species (the accession numbers of the sequences were shown on the figure in Fig 4) the, were used for phylogenic inference. Sequence alignments were performed using the MAFFT software (Katoh *et al*., 2002) with the FFT-NS-1 strategy. The aligned sequences were then subjected to MEGA X for phylogeny construction (Kumar *et al*., 2018).

RESULTS

Sequencing and plastome assembly: A total of 4.65 and 4.24 GB of raw data were generated for the sequencing libraries of QLYB (*J. przewalskii*) and CZYB (*S. przewalskii* f. *pendula*), respectively. After filtering, 4.62 Gb and 4.21 Gb clean data were used for CP-like read extraction. The complete assemblies of QLYB and CZYB were 128,748 bp and 127,315 bp in length, with GC contents of 34.98% and 35.08%, respectively (Fig. 1; Table 1). Similar to most conifer species, no canonical inverted repeats (IRs) were detected in the CP genome of QLYB and CZYB, but two short repeats of 224 bp were identified in the plastomes of both species. The large single-copy (LSC) regions were close in size, having lengths of 92,257 bp and 90,986 bp for QLYB and CZYB, respectively. The sizes of the small single-copy (SSC) regions were also slightly different (36,043 and 35,881 for QLYB and CZYB, respectively), showing the typical structure of *Juniperus* species (Table 1). The complete CP genome sequences of QLYB and CZYB have been submitted to GenBank and deposited under the accession numbers of OL470654.1 and OL470655.1, respectively.

Plastome features of QLYB (*J. przewalskii*) and CZYB (*S. przewalskii* f. *pendula*): A total of 121 genes were identified in the QLYB plastome, including 85 protein-coding genes, 32 tRNAs and 4 rRNAs (Table 2). A *rpl32* gene was found missing in the CZYB cp genome, resulting in a total of 120 annotated genes. In sum, introns were detected in 11 genes (*rps12, rpoC1, trnI-GAU, trnK-UUU, trnA-UGC, trnG-UCC, trnV-UAC, ndhA, ndhB, atpF* and *ycf3*) of the QLYB plastome. In the CP genome of CZYB, 14 genes (*rps12, rpl2, rpl16, rpoC1, trnI-GAU, trnK-UUU, trnA-UGC, trnV-UAC, ndhA, ndhB, petD, petB, atpF* and *ycf3*) were identified as intron-containing genes.

Codon usage in QLYB (*J. przewalskii*) and CZYB (*S.*

***przewalskii* f. *pendula*) CP genomes:** According to the results of the codon usage analysis, there were no substantial differences in codon usage between the QLYB and CZYB CP genomes, but the codon usage of different amino acids varied significantly. Sixty-one codons that encode 20 amino acids were identified in both species (Table 3). The leucine codons showed the highest usage frequency, while cysteine showed the lowest. In sum, 30 codons have a higher usage rate (RSCU > 1; Table 3) and 29 codons are used less frequently (RSCU < 1; Table 3). The usage of the AUG codon was unbiased (RSCU = 1; Table 3).

Repeat sequences and SSR analysis: Repeats of 30–50 bp were found to be the most abundant in the CZYB and QLYB CP genomes (Fig. 2A). A total of 38 repeat sequences were detected in the CZYB plastome, which were classified into three categories: forward (28), palindromic (4) and tandem repeats (6) (Fig. 2B). A larger number of repeats were identified in the CP genome of QLYB (85 in sum), including 72 forward, 5 palindromic and 8 tandem repeats (Fig. 2B).

Simple sequence repeats (SSRs) are 1–6 nucleotides repeated multiple times. In this study, the composition of SSRs was similar in the CP genomes of QLYB and CZYB. As shown in Fig. 2C, D and E, mononucleotide SSRs were the most frequently identified type, among which A/T SSRs were the most abundant. Statistics of the SSRs revealed that QLYB and CZYB shared similar composition features.

Genome sequence divergence among QLYB, CZYB and *Juniperus chinensis*: Overall, the structure *Juniperus* species plastomes are relatively conserved. However, by comparing the structural features of six species, divergent regions were also identified, especially in noncoding regions such as *rpl*23 (Fig. 3). Strong variation was also detected in regions of some coding genes, such as *accD* (Fig. 3).

Phylogenetic analysis: A large proportion of the gymnosperms belong to the Cupressaceae and the Pinaceae family, and *Juniperus* is one of the most important genera of the Cupressoideae subfamily (Eckenwalder, 1976). Studying the phylogenetic relationship of *J. Przewalskii, S. przewalskii* f. *pendula* and other Cupressaceae species can offer better guidance for studying related gymnosperms. Hence, plastome sequences of 18 gymnosperm plants, including species from *Juniperus* and *Thuja*, were used for phylogenetic exploration. As shown in Fig. 4, *J. przewalskii* (QLYB) is more closely related to *J. Tibetica,* whereas and *S. przewalskii* f. *pendula* (CZYB) is more closely related to *J. recurve, J. communis* and *J. squamata*. *Juniperus* plants form a mainly monophyletic group except *J. formosana*, which has a closer relationship to *Thuja* species (Fig. 4).

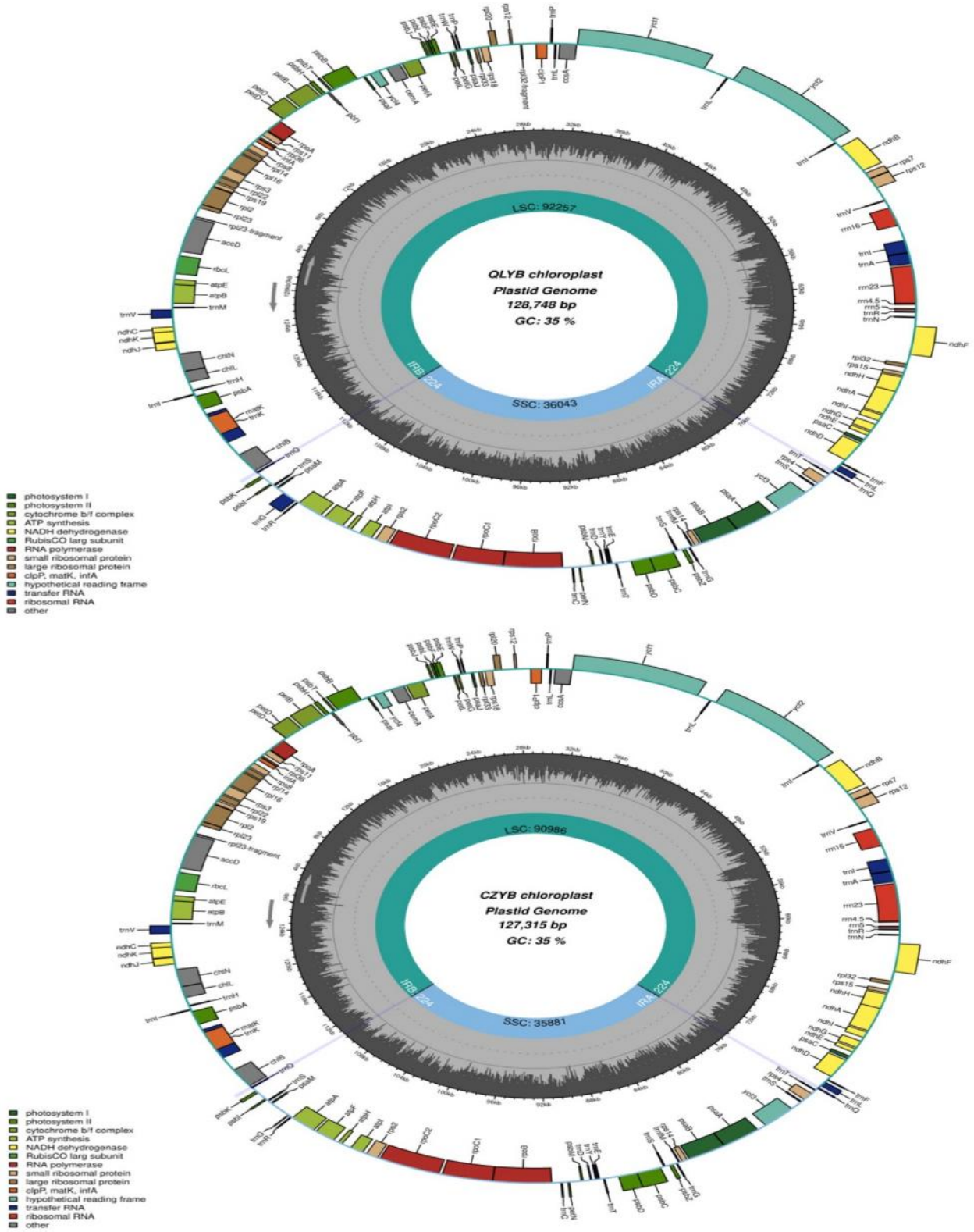

Fig. 1. Chloroplast map of QLYB (*Juniperus przewalskii*) and CZYB (*Sabina przewalskii* f. *pendula*). Genes drawn inside the circle are transcribed clockwise; genes outside the circle are transcribed counterclockwise. The different color boxes represent different conservative genes. The darker gray of the inner circle revels the GC content, whereas the blue lines show two gene copies in the IR region.

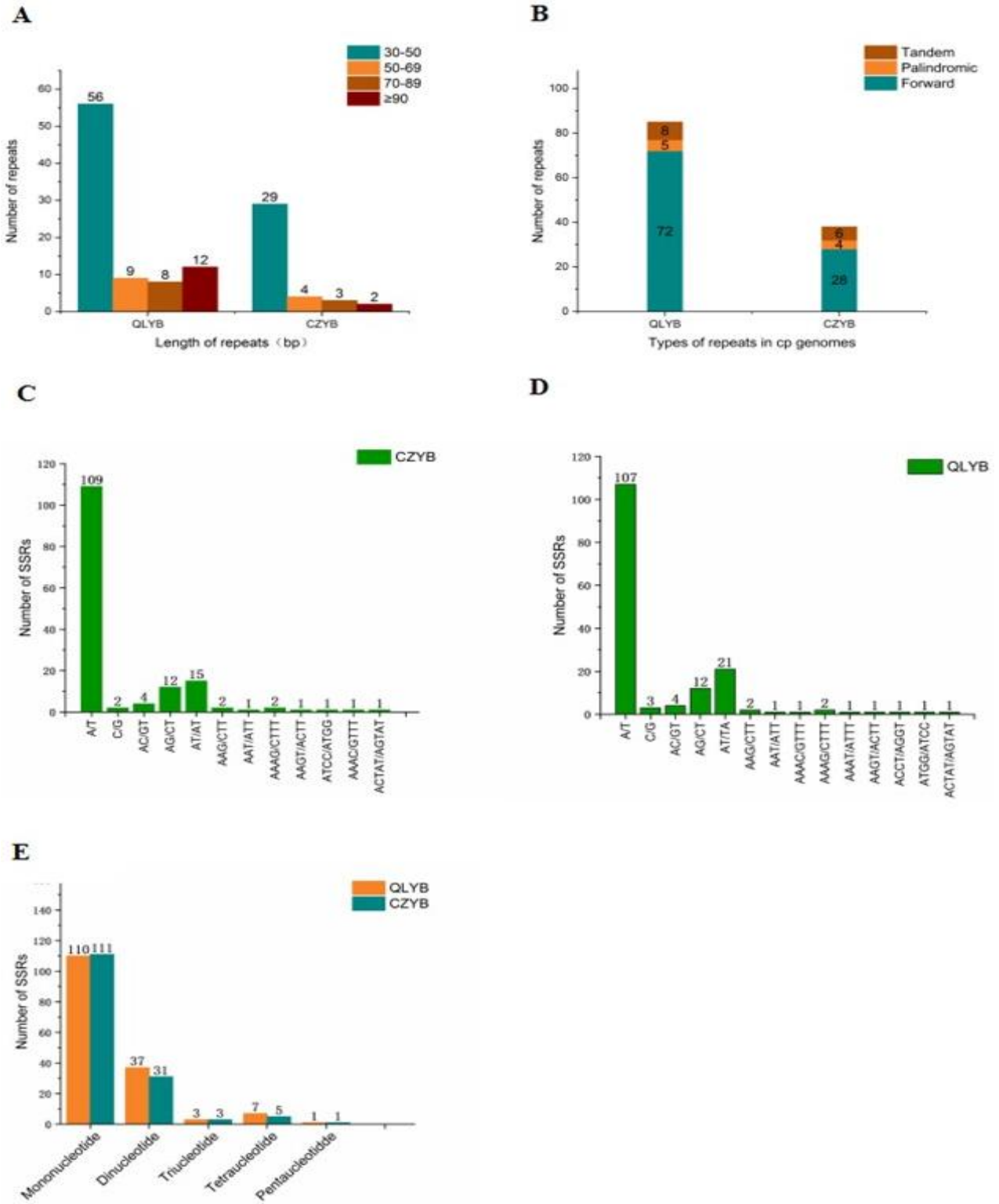

Fig. 2. Repeated sequences in the chloroplast genomes of QLYB (*Juniperus przewalskii*) and CZYB (*Sabina przewalskii* f. *pendula*). A: Lengths of repeats in QLYB and CZYB; B: Types of repeat genomes in QLYB and CZYB; C: Frequency of diversified SSR motifs in CZYB; D: Frequency of diversified identified SSR motifs in QLYB; E: Total number of the six repeat types in the chloroplast genomes of CZYB and QLYB.

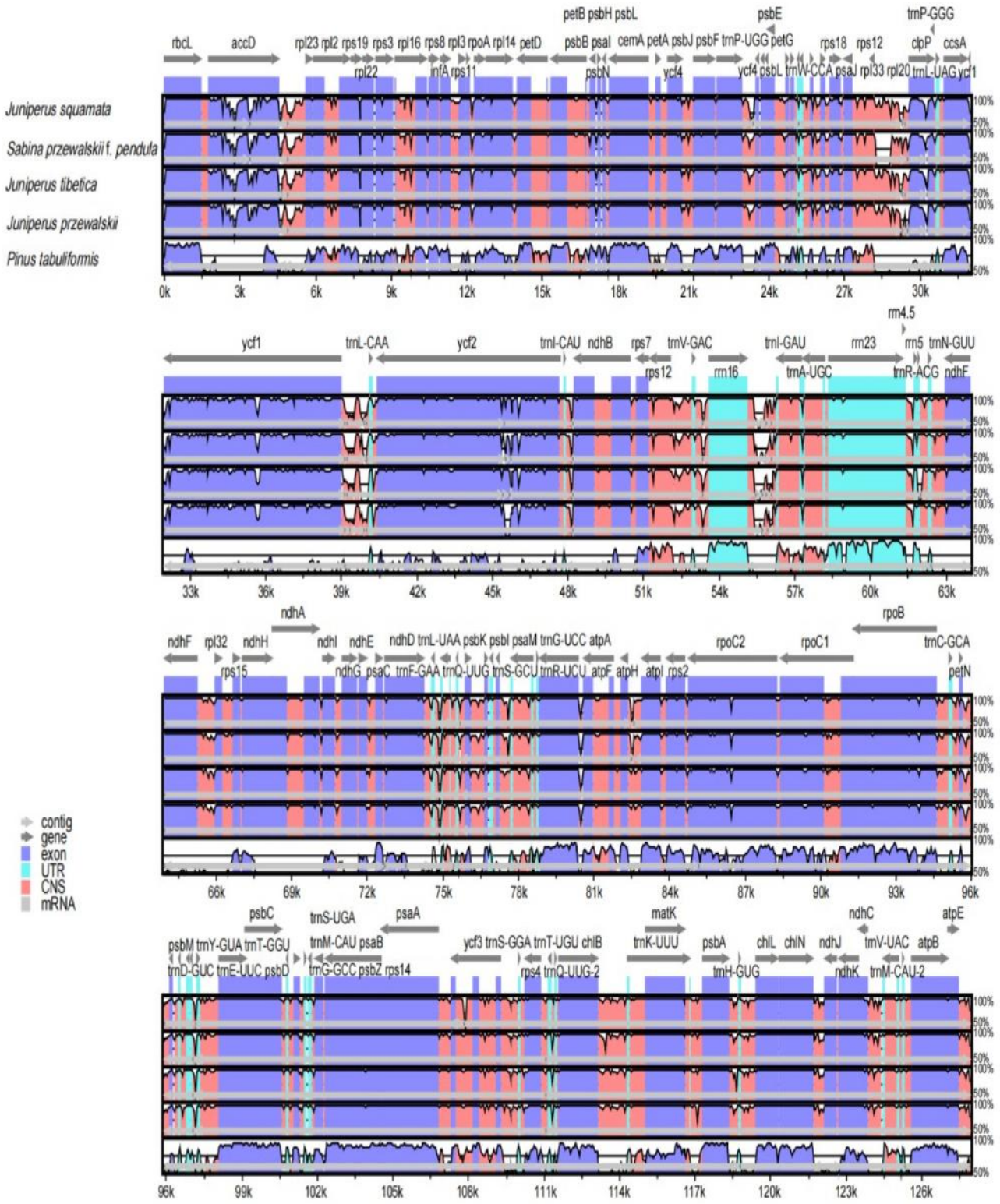

Fig. 3. Comparative plots based on sequence identity of the chloroplast genome of three species using *Juniperus communis* as the reference genome. Plots were constructed with the mVISTA program. Chloroplast coding regions are indicated in blue and noncoding regions in red; notice the reduction in sequence identity indicated by the reduction of blue/red shadowing (white spaces).

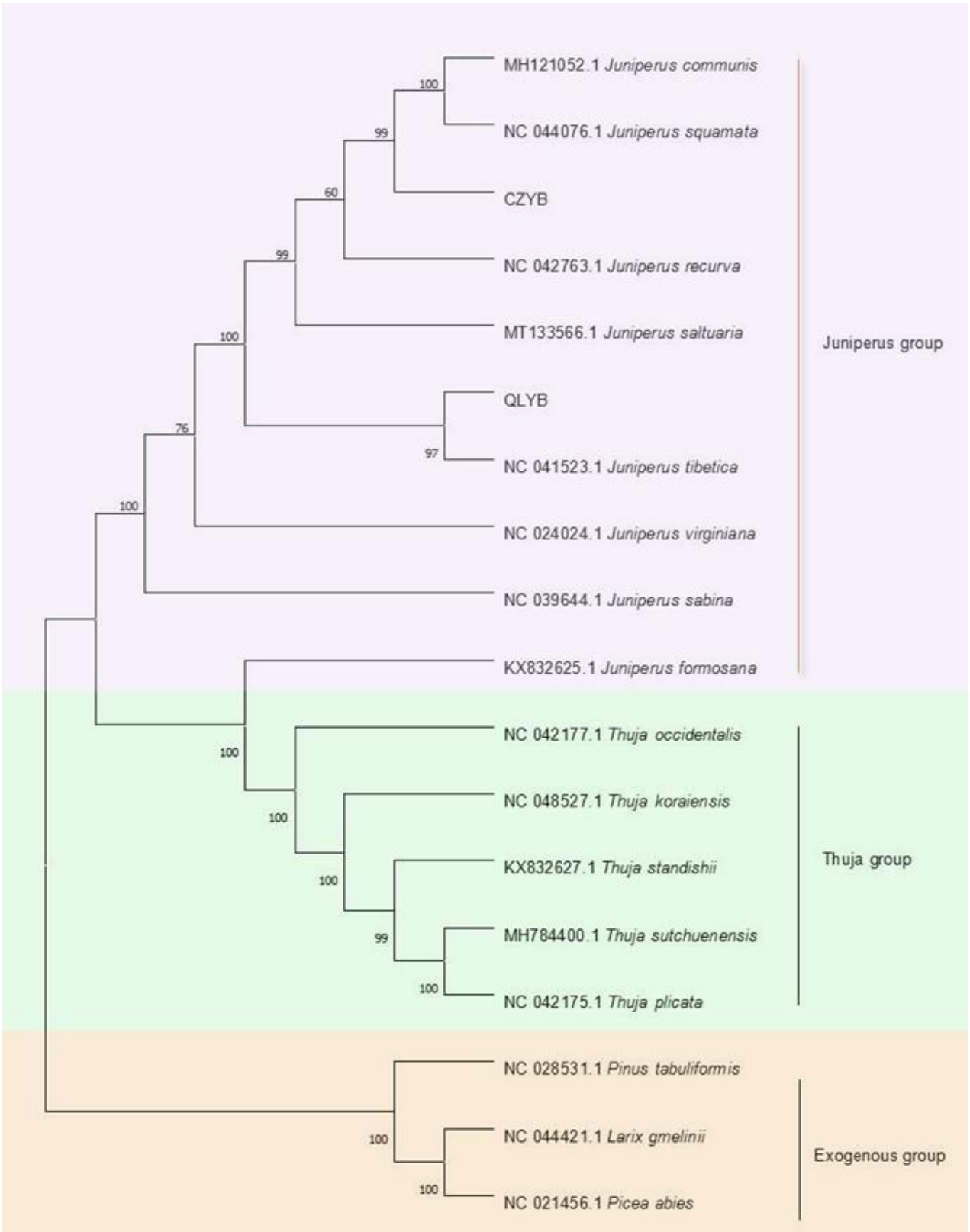

Fig. 4. The chloroplast-genome-based phylogenetic tree of CZYB (*Sabina przewalskii* f. *pendula),* QLYB (*Juniperus przewalskii)* and 16 related species. The bootstrap values were based on 1,000 resamplings, which are shown next to the branches.

Table 1. Summary of QLYB and CZYB chloroplast genome features

Category	QLYB	CZYB
Genome size (bp)	128748	127315
LSC size (bp)	92257	90986
SSC size (bp)	36043	35881
IRa/IRb size (bp)	224/224	224/224
Total GC content (%)	34.98	35.08
Gene (total /different)	121/113	120/114
Protein-coding gene	85	84
tRNAs	32	32
rRNAs	4	4
Genes duplicated in IR	2	2

CZYB: *Sabina przewalskii* f. *pendula*; IR, inverted repeat; LSC, large single-copy region; SSC, small single-copy region; QLYB: *Juniperus przewalskii*

Table 2. Genes in QLYB and CZYB chloroplast genomes

Groups of genes	Names of genes
Ribosomal proteins (SSU)	rps14, rps3, rps7, rps18, rps8, rps11, rps12[ab], rps15, rps19, rps4, rps2
Ribosomal proteins (LSU)	rpl36, rpl33, rpl2[b], rpl32 (×2)①, rpl16[b], rpl20, rpl14, rpl22, rpl23 (×2)
RNA polymerase	rpoC2, rpoA, rpoB, rpoC1[ab]
Ribosomal RNAs	rrn4.5, rrn16, rrn5, rrn23
Transfer RNAs	trnL-UAG, trnI-GAU[ab], trnT-GGU, trnK-UUU[ab], trnY-GUA, trnP-GGG, trnR-UCU, trnV-GAC, trnA-UGC[ab], trnI-CAU (×2), trnL-CAA, trnfM-CAU, trnM-CAU, trnG-GCC, trnD-GUC, trnE-UUC, trnC-GCA, trnQ-UUG (×2), trnL-UAA, trnW-CCA, trnN-GUU, trnG-UCC[a], trnH-GUG, trnS-GCU, trnS-UGA, trnP-UGG, trnR-ACG, trnS-GGA, trnT-UGU, trnV-UAC[ab], trnF-GAA
Photosystem Ⅰ	psaM, psaC, psaB, psaA, psaI, psaJ
Photosystem Ⅱ	psbD, psbB, psbK, psbM, psbZ, psbH, psbL, psbF, psbJ, psbI, psbA, psbE, psbT, psbC
NADH dehydrogenase	ndhA[ab], ndhG, ndhD, ndhJ, ndhC, ndhH, ndhF, ndhK, ndhE, ndhI, ndhB[ab]
Cytochrome b/f complex	petL, petG, petA, petD[b], petN, petB[b]
ATP synthase	atpA, atpH, atpE, atpB, atpF[ab], atpI
Rubisco large subunit	rbcL
Maturase	matK
ATP-dependent protease	clpP1
Hypothetical chloroplast reading frames	ycf2, ycf4, ycf1, ycf3[ab]
Translational initiation factor	infA
Other genes	pbf1, chlN, chlL, cemA, chlB, accD, ccsA

CZYB: *Sabina przewalskii* f. *pendula*; QLYB: *Juniperus przewalskii*
Note: (×2): two gene copies; [a]: intron containing genes of QLYB; [b]: intron containing genes of CZYB; ①: two gene copies existing in the QLYB plastome only

Table 3. Codon usage in QLYB and CZYB chloroplast genomes.

Amino Acids	Condon	Number (QLYB/CZYB)	RSCU (QLYB/CZYB)	Amino Acids	Condon	Number (QLYB/CZYB)	RSCU (QLYB/CZYB)
Ala	GCU	626/626	1.93/1.93	Pro	CCU	406/404	1.77/1.76
	GCC	155/154	0.48/0.47		CCC	155/154	0.68/0.67
	GCA	404/403	1.24/1.24		CCA	242/242	1.06/1.05
	GCG	114/114	0.35/0.35		CCG	114/118	0.5/0.51
Cys	UGU	186/189	1.43/1.44	Gln	CAA	644/651	1.57/1.57
	UGC	75/74	0.57/0.56		CAG	179/180	0.43/0.43
Asp	GAU	811/809	1.62/1.61	Arg	CGU	316/313	1.47/1.46

	GAC	193/193	0.38/0.39		CGC	98/98	0.45/0.46
Glu	GAA	1120/1129	1.6/1.6		CGA	268/265	1.24/1.23
	GAG	280/279	0.4/0.4		CGG	68/69	0.32/0.32
Phe	UUU	934/937	1.43/1.44		AGA	432/431	2/2
	UUC	374/368	0.57/0.56		AGG	112/114	0.52/0.53
Gly	GGU	487/487	1.41/1.41	Ser	UCU	506/506	1.94/1.95
	GGC	131/131	0.38/0.38		UCC	208/204	0.8/0.78
	GGA	594/594	1.71/1.72		UCA	319/321	1.23/1.23
	GGG	174/173	0.5/0.49		UCG	121/119	0.47/0.46
His	CAU	386/388	1.59/1.6		AGU	308/311	1.18/1.2
	CAC	99/98	0.41/0.4		AGC	99/99	0.38/0.38
Ile	AUU	971/974	1.48/1.48	Thr	ACU	461/462	1.77/1.77
	AUC	321/323	0.49/0.49		ACC	160/161	0.61/0.62
	AUA	675/672	1.03/1.02		ACA	305/307	1.17/1.18
Lys	AAA	1143/1133	1.52/1.52		ACG	116/113	0.45/0.43
	AAG	358/362	0.48/0.48	Val	GUU	420/415	1.44/1.43
Leu	UUA	822/827	1.99/1.99		GUC	133/136	0.46/0.47
	UUG	517/516	1.25/1.24		GUA	410/406	1.41/1.4
	CUU	496/499	1.2/1.2		GUG	203/203	0.7/0.7
	CUC	159/166	0.38/0.4	Trp	UGG	409/408	1/1
	CUA	350/347	0.85/0.84	Tyr	UAU	675/675	1.58/1.57
	CUG	136/138	0.33/0.33		UAC	182/185	0.42/0.43
Met	AUG	515/514	1/1	Ter	UAG	10/10	0.58/0.59
Asn	AAU	834/834	1.56/1.55		UGA	8/8	0.46/0.47
	AAC	237/240	0.44/0.45				

CZYB: *Sabina przewalskii* f. *pendula*; QLYB: *Juniperus przewalskii*; Ter: stop codon

DISCUSSION

In this study, the chloroplast genomes of two *Juniperus* species were first sequenced using the Illumina NovaSeq system. Although the size of the CP genome of *J. przewalskii* (128,748 bp) differed slightly from that of *S. przewalskii* f. *pendula* (127,315 bp), plastomes of both species had two small IRs of 224 bp. No canonical IRs were identified. The GC content of the *S. przewalskii* plastome is 35.08%, encoding 84 proteins, 32 tRNAs and 4 rRNAs. And the GC content of the *J. przewalskii* CP genome is 34.98%, and one additional *rpl32* gene was detected. The results revealed that the plastomes of these two species had similar structure characteristics to *Juniperus* species (Lu *et al*., 2018). However, the IR sequences are absent in the CP genome of *Juniperus microsperma* (Tso *et al*., 2019), as well as some other pinaceae and cypsaceae species (Hao *et al*., 2016). Gymnosperm land plants have CP genomes that consist of a relatively conserved quadripartite structure that contains a pair of IR (IRA and IRB) regions, an LSC region and an SSC region (Yang *et al*., 2021). The length of the CP genome of *J. przewalskii* (128,748 bp) differed slightly from that of *S. przewalskii* f. *pendula* (127,315 bp), but two *Juniperus* species had the same IR lengths of 224 bp.

The IR regions are the most conserved regions in the CP genome (Daniell *et al*., 2016), and contraction or expansion of the IR regions is considered the main cause of size changes of CP genomes (Kode *et al*., 2005; Khakhlova and Bock, 2006; Yao *et al*., 2015). Comparison of angiosperm plastomes reveals that the length of IRs varies substantially among taxa, and plastome rearrangements of gymnosperm species are also found to be tightly related to IRs (Guisinger *et al*., 2011; Jansen and Ruhlman, 2012). An additional *rpl32* gene was detected in the CP genome of *J. przewalskii*, which could be gained by transferring from the nuclear genome. The transfer of *rpl32* has been observed in Populus (Steane, 2005; Okumura *et al*., 2006), but whether this additional *rpl32* is originated from the nuclear genome still needs confirmation.

Owing to their high variability, SSRs and other repetitive sequences have been utilized as markers for species discrimination (Ebert and Peakall, 2009; Dong *et al*., 2013; Dong *et al*., 2016). In this study, the CP genomes of *J. przewalskii* and *S. przewalskii* f. *pendula* were rich in SSRs (287 SSRs in *S. przewalskii* f. *pendula* and 305 SSRs in *J. przewalskii*). These SSRs will provide a useful reference in future research on the identification, genetic diversity and evolutionary process of *Juniperus* species.

We also compared variations among the CP genomes of six species using mVISTA. This facilitated the identification of highly variable regions, especially non-coding regions and the coding region of the accD gene. These sequences, together with the identified SSRs and long repeats, could be used as potential markers for

species discrimination.

The phylogenetic tree revealed that *J. przewalskii* is more closely related to *J. tibetica*, whereas *S. przewalskii* f. *pendula* is more closely related to *J. recurva, J. communis* and *J. squamata*. Using sequenced plastomes, we further confirmed the phylogenetic relationship of *Juniperus* species, including *J. przewalskii* and *S. przewalskii* f. *pendula*. Discriminating between *J. przewalskii* and *S. przewalskii* f. *pendula* can be challenging using only morphological traits. However, using the plastome sequences generated in the present study, a molecular approach could be developed and adopted in further studies.

Conclusions: Herein, the complete CP genomes of *S. przewalskii* f. *pendula* and *J. przewalskii* were generated by Illumina sequencing. The structures of the plastomes of these two species are highly similar to those of other *Juniperus* species. The results of this study provided detailed information on the CP genomes of *S. przewalskii* f. *pendula* and *J. przewalskii*. The divergent regions and the repetitive sequences identified could be developed as molecular markers that would benefit breeding and species discrimination.

Acknowledgments: We thank Prof. Qiang-Feng Li and Guang-Jing Yin for the supporting of our work.

Funding: This research was supported by the Basic Research Project of Qinghai Province, China (No. 2019-ZJ-962Q).

Data availability: The complete CP genome sequences that support the findings of this study are openly available in NCBI at GenBank database with accession numbers, OL470654.1 and OL470655.1. (https://www.ncbi.nlm.nih.gov).

Compliance with ethical standards: All field sampling work was allowed by the local government and all the experiments did not involve humans and animals. There was no issue on ethics in this study.

Conflict of interest: The authors declare no conflicts of interest.

Author's contributions: Conceptualization and experimental design: Chun-Jing Wang (60%), Le Wang (40%); Formal analysis and investigation: Chun-Jing Wang (30%), Shuai-Peng Si (40%), Ji-Zhong Wan (10%), Ling-Hua Yang (5%), Ze-Zhu Wu (5%), Biao Xiong (5%), Chun-Hui Zhang (5%); Project administration: Chun-Jing Wang (70%), Le Wang (30%); Writing—original draft: Shuai-Peng Si (100%); Writing—review and editing: Chun-Jing Wang (50%), Le Wang (50%).

REFERENCES

Bankevich A., S. Nurk, D. Antipov,A.A. Gurevich , M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev and P.A. Pevzner (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput Biol. 19(5): 455-77. doi.org/10.1089/cmb.2012.0021

Beier S., T. Thiel, T. Münch, U. Scholz and M. Mascher (2017). MISA-web: a web server for microsatellite prediction. Bioinformatics. 33: 2583–2585. doi.org/10.1093/bioinformatics/btx198

Borgstrom E., S. Lundin and J. Lundeberg (2011). Large scale library generation for high througthput sequencing. PLoS. One. 6(4): e19119. doi.org/10.1371/journal.pone.0019119

Chen Y., Y. Chen, C. Shi, Z. Huang, Y. Zhang, S. Li, Y. Li, J. Ye, C. Yu, Z. Li, X. Zhang, J. Wang, H. Yang, L. Fang and Q. Chen (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 7(1): 1–6. doi/10.1093/gigascience/gix120

Daniell H., C.S. Lin, M. Yu and W.J. Chang (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17: 134. doi.org/10.1186/s13059-016-1004-2

Dong W., C. Xu, T. Cheng, K. Lin and S. Zhou (2013). Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of saxifragales. Genome Biol. Evol. 5(5): 989-997. doi.org/10.1093/gbe/evt063

Dong W.P., C. Xu, D.L. Li, X.B. Jin, R.L. Li, Q. Lu and Z.L. Suo (2016). Comparative analysis of the complete chloroplast genome sequences in psammophytic Haloxylon species (Amaranthaceae). Peerj. 4(2): e2699. doi.org/10.7717/PEERJ.2699

Drouin G., H. Daoud and J. Xia (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol. Phylogenet Evol. 49(3): 827–831. doi.org/10.1016/j.ympev.2008.09.009

Ebert D. and R. Peakall (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(3): 673-690. doi.org/10.1111/j.1755-0998.2008.02319.x

Eckenwalder J.E. (1976). Re-evaluation of Cupressaceae and Taxodiaceae: a proposed merger. Madrono 23(5): 237–256.

Editorial Committee of Woody Flora of Qinghai Province (1987). Woody Flora of Qinghai Province. 1st Ed, Qinghai People's Publishing House; Xining (China). 48p-49p

Frazer K.A., L. Pachter, A. Poliakov, E.M. Rubin and I. Dubchak (2004). VISTA: computational tools for comparative genomics. Nucleic Acids. Res. 32: W273-279. doi.org/10.1093/nar/gkh458

Guisinger M.M., J.V. Kuehl, J.L. Boore and R.K. Jansen (2011). Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol. Biol. Evol. 28(1): 583-600. doi.org/10.1093/molbev/msq229

Hao Z.D., T.L. Cheng, R.H. Zhang, H.B. Xu, Y.W. Zhou, M.P. Li, F.J. Lu, Y.N. Dong, X. Liu, J.H. Chen and J.S. Shi (2016). The Complete Chloroplast Genome Sequence of a Relict Conifer Glyptostrobus pensilis : Comparative Analysis and Insights into Dynamics of Chloroplast Genome Rearrangement in Cupressophytes and Pinaceae. PLoS. One. 11(8): e0161809. doi.org/10.1371/journal.pone.0161809

Hollender C.A. and C. Dardick (2015). Molecular basis of angiosperm tree architecture. New Phytol. 206: 541-556. doi.org/10.1111/nph.13204

Hollender C.A., T. Pascal, A. Tabb, T. Hadiarto, C. Srinivasan, W. Wang, Z. Liu, R. Scorza and C. Dardick (2018). Loss of a highly conserved sterile alpha motif domain gene (WEEP) results in pendulous branch growth in peach trees. P. Natl. Acad. Sci. USA. 115(20): E4690-E4699. doi.org/10.1073/pnas.1704515115

Jansen R.K. and T.A. Ruhlman (2012). Plastid genomes of seed plants. Cardiovasc Eng. 35: 103–126. doi.org/10.1007/978-94-007-2920-9_5

Katoh K., K. Misawa1, K. Kuma and T. Miyata (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids. Res. 30(14): 3059–3066. doi.org/10.1093/nar/gkf436

Kerstin D., T.R. Hodkinson, K.H. Wolfe, R. Van, D. Bekerom, P.J. Dix and S. Barth (2009). Complete Chloroplast Genome Sequence of a Major Allogamous Forage Species, Perennial Ryegrass (*Lolium perenne* L.). DNA Res. 16(3): 165–176. doi.org/10.1093/dnares/dsp008

Khakhlova O. and R. Bock (2006). Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 46(1): 85-94. doi.org/10.1111/j.1365-313X.2006.02673.x

Kode V., E.A. Mudd, S. Iamtham and A. Day (2005). The tobacco plastid accD gene is essential and is required for leaf development. Plant J. 44(2): 237-244. doi.org/10.1111/J.1365-313X.2005.02533.X

Kumar S., G. Stecher, M. Li, K. Christina and T. Koichiro (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol. Biol. Evol. 18(6): 6. doi.org/10.1093/molbev/msy096

Lei W.J., D.P. Ni, Y.J. Wang, J.J. Shao, X.C. Wang, D. Yang, J.S. Wang, H.M. Chen and C. Liu (2016). Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of Astragalus membranaceus. Sci. Rep. 6: 21669. doi.org/10.1038/srep21669

Lu D.Y., L. Zhang, L. Hao, W.Y. Guo, Y.X. Li, Y.K. Zhang and G.S. Zhang (2018). Analysis of Chloroplast Genome Structure and Phylogenetic Evolution of *Juniperus sabina*. Acta. Bot. Bor-Occid. Sin. 38(08): 1464-1475. doi.org/10.7606/j.issn.1000-4025.2018.08.1464

Mao T.Y., H.H. Zhu, Y.Y. Liu, M.Z. Bao, J.W. Zhang, Q. Fu, C.F. Xiong and J. Zhang (2020). Weeping candidate genes screened using comparative transcriptomic analysis of weeping and upright progeny in an F1 population of Prunus mume. Physiol. Plantarum. 170(3): 328-334. doi.org/10.1111/ppl.13179

Mayor C., M. Brudno, J.R. Schwartz, A. Poliakov, E.M. Rubin, K.A. Frazer, L.S. Pachter and I. Dubchak (2000). VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 16(11): 1046–1047. doi.org/10.1093/bioinformatics/16.11.1046

Okumura S., M. Sawada, Y.W. Park, T. Hayashi, M. Shimamura, H. Takase and K.I. Tomizawa (2006). Transformation of poplar (*Populus alba*) plastids and expression of foreign proteins in tree chloroplasts. Transgenic Res. 15(5): 637-646. doi.org/10.1007/s11248-006-9009-3

Redwan R.M., A. Saidin and S.V. Kumar (2015). Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. BMC Plant Biol. 15: 196. doi.org/10.1186/s12870-015-0587-1

Shinozaki K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B.Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada and M. Sugiura (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. Embo. J. 5(9): 2043–2049. doi.org/10.1002/j.1460-

2075.1986.tb04464.x

Song E., S. Park and S. Kim (2019). Primers for complete chloroplast genome sequencing in Magnolia. Appl. Plant Sci. 7(9): e11286. doi.org/10.1002/aps3.11286

Steane A.D. (2005). Complete Nucleotide Sequence of the Chloroplast Genome from the Tasmanian Blue Gum, *Eucalyptus globulus* (Myrtaceae). DNA Res. 12(3): 215-220. doi.org/10.1093/dnares/dsi006

Thiel T., W. Michalek, R. Varshney and A. Graner (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 106(3): 411–422. doi.org/10.1007/s00122-002-1031-0

Tillich M., P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock and S. Greiner (2017). GeSeq-versatile and accurate annotation of organelle genomes. Nucleic Acids. Res. 45(W1): W6–W11. doi.org/10.1093/nar/gkx391

Tso S., J.L. Li, S.Y. Xie, J.B. Miao, Q.J. Hu and K.S. Mao (2019). Characterization of the complete chloroplast genome of *Juniperus microsperma* (Cupressaceae), a rare endemic from the Qinghai-Tibet Plateau. Conservation Genet. Resour. 11: 325-328. doi.org/10.1007/s12686-018-1027-y

Wick R.R., L.M. Judd, C.L. Gorrie and K.E. Holt (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. Plos. Computational Biology 13(6): e1005595. doi.org/10.1371/journal.pcbi.1005595.t001

Xu J., X. Shen, B. Liao, J. Xu and D. Hou (2020). Comparing and phylogenetic analysis chloroplast genome of three Achyranthes species. Sci. Rep. 10(1): 10818. doi.org/10.1093/bioinformatics/btz305

Yang X., T. Zhou, X. Su, G.B. Wang, X.H. Zhang, Q.R. Guo and F.L. Cao (2021). Structural characterization and comparative analysis of the chloroplast genome of *Ginkgo biloba* and other gymnosperms. J. For. Res. 32: 765-778. doi.org/10.1007/s11676-019-01088-4

Yang, Y. and W. Liao (2013). *Juniperus przewalskii*. The IUCN Red List of Threatened Species 2013: e.T34141A2847665.

Yao X., P. Tang, Z. Li, D. Li, Y. Liu and H. Huang (2015). The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis. PLoS. One. 10(6): e0129347. doi.org/ 10.1371/journal.pone.0129347

Zhang Q., T.Y. Chiang, M. George, J.Q. Liu and R.J. Abbott (2005). Phylogeography of the Qinghai-Tibetan Plateau endemic przewalskii (Cupressaceae) inferred from chloroplast DNA sequence variation. Mol. Ecol. Resour. 14(11): 3513-3524. doi.org/10.1111/j.1365-294X.2005.02677.x

Zheng S., P. Poczai, J. Hyvönen, J. Tang and A. Amiryousefi (2020). Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. Front. Genet. 11: 576124. doi.org/10.3389/fgene.2020.576124.