PLOS ONE

RESEARCH ARTICLE

The chloroplasts genomic analyses of four specific *Caragana* species

Maohua Yuan[1,2☯], **Xianmei Yin**[1☯], **Bixing Gao**[1,3], **Rui Gu**[1], **Guihua Jiang**[1,4,5*]

1 College of Pharmacy, Chendu University of Traditional Chinese Medicine, Chendu, China, **2** Traditional Chinese Medicine Detection, Leshan Food and Drug Inspection and Testing Center, Leshan, China, **3** Sichuan Institute for Drug Control, Key Laboratory of Quality Evaluation of Chinese Patent Medicines, Chengdu, China, **4** Key Laboratory of Standardization of Chinese Medicines of Ministry of Education, State Key Laboratory of Characteristic Chinese Medicine Resources in Southwest China, School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, China, **5** School of Ethnic Medicine, Chengdu University of TCM, Chengdu, China

☯ These authors contributed equally to this work.
* 11469413@qq.com

Citation: Yuan M, Yin X, Gao B, Gu R, Jiang G (2022) The chloroplasts genomic analyses of four specific *Caragana* species. PLoS ONE 17(9): e0272990. https://doi.org/10.1371/journal.pone.0272990

Editor: Vikas Sharma, Sant Baba Bhag Singh University, INDIA

Received: December 24, 2021

Accepted: July 30, 2022

Published: September 1, 2022

Data Availability Statement: The datasets generated during the current study are available in The Genome Warehouse in National Genomics Data Center (NGDC, https://ngdc.cncb.ac.cn/)[1][2] (accession numbers: GWHBJYO00000000, GWHBJYN00000000, GWHBJYM00000000, and GWHBJYL00000000).

Funding: This work was supported by the Major Project of "National key research and development plan project projects", under Grant #2019YFC1712302, 2019YFC1712305. The funders had no role in study design, data collection

Abstract

Background

Many species of the genus *Caragana* have been used as wind prevention and sand fixation plants. They are also important traditional Chinese medicine, and ethnic medicine resource plant. Thus, chloroplast genomes (cp-genome) of some of these important species must be studied.

Methods

In this study, we analyzed the chloroplast genomes of *C. jubata*, *C. erinacea*, *C. opulens*, and *C. bicolor*, including their structure, repeat sequences, mutation sites, and phylogeny.

Results

The size of the chloroplast genomes was between 127,862 and 132,780 bp, and such genomes contained 112 genes (30 tRNA, 4 rRNA, and 78 protein-coding genes), 43 of which were photosynthesis-related genes. The total guanine + cytosine (G+C) content of four *Caragana* species was between 34.49% and 35.15%. The four *Caragana* species all lacked inverted repeats and can be classified as inverted repeat-lacking clade (IRLC). Of the anticipated genes of the four chloroplast genomes, introns were discovered in 17 genes, most of which were inserted by one intron. A total of 50 interspersed repeated sequences (IRSs) were found among them, 58, 29, 61, and 74 simple sequences repeats were found in *C. jubata*, *C. bicolor*, *C. opulens*, and *C. erinacea*, respectively. Analyses of sequence divergence showed that some intergenic regions (between *trnK-UUU* and *rbcl*; *trnF-GAA* and *ndhJ*; *trnL-CAA* and *trnT-UGU*; *rpoB* and *trnC-GCA*; *petA* and *psbL*; *psbE* and *pebL*; and sequences of *rpoC*, *ycf1*, and *ycf2*) exhibited a high degree of variations. A phylogenetic tree of eight *Caragana* species and another 10 legume species was reconstructed using full sequences of the chloroplast genome.

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Conclusions

(1) Chloroplast genomes can be used for the identification and classification of *Caragana* species. (2) The four *Caragana* species have highly similar cpDNA G+C content. (3) IRS analysis of the chloroplast genomes showed that these four species, similar to the chloroplast genome of most legumes, lost IRLC regions. (4) Comparative cp-genomic analysis suggested that the cp genome structure of the *Caragana* genus was well conserved in highly variable regions, which can be used to exploit markers for the identification of *Caragana* species and further phylogenetic study. (5) Results of phylogenetic analyses were in accordance with the current taxonomic status of *Caragana*. The phylogenetic relationship of *Caragana* species was partially consistent with elevation and geographical distribution.

1. Introduction

About 100 *Caragana* species exist worldwide. Among the species in the arid and semi-arid regions of Asia and Europe, 66 species (32 endemic) can be found in China [1]. The genus *Caragana* is a deciduous shrub with a wide range of adaptability and strong stress tolerance. Most of the *Caragana* species are distributed in higher elevations and relatively harsh environments (barren, drought, heat, and cold); they are known to prevent wind and fixate sand [2–5]. Moreover, previous studies have shown that many plants have pharmacological, antibacterial, and antioxidant activities and anti-tumor, anti-HIV, and other effects [2, 5, 6]. All four species in this study have been documented in traditional Chinese ethnic medicine. Among them, *C. jubata* is an important Tibetan drug that can be used to treat alpine erythrocytosis and hypertension; it possesses hepatoprotective and antiviral activities[7–11].

Phylogenetic relationships in the genus *Caragana* remain obscure, and also have some problems in the identification of medicinal species. Only 10 cp-genome of the *Caragana* genus have been reported and low amount of data are available for analysis [12–14].

Chloroplasts are the posterity of ancient microbacillary endosymbionts. They are the usual organelles of green plants, which play an indispensable role in photosynthesis [15]. Ordinarily, the descendibility of the chloroplast genome is maternal in angiosperms [16]. The chloroplast genome is relatively stable in structure, and it contains a large single-copy region, small single-copy region, and two inverse repeat (IR) regions. Inverted repeat-lacking clade (IRLC) has been reported in legumes [17–20]: four *Caragana* species have been reported with IRLC [11–13]. Therefore, the *Caragana* genus can represent a lineage with extensive IRLC. However, knowledge of the pattern, origin, and evolution of plastomic IRLC within *Caragana* is presently limited by the scarcity of plastomic sequences. In addition, the chloroplast genomic model can be used to study molecular identification, phylogeny, species conservation, and evolution [21, 22].

In the present study, the four species of the genus *Caragana* from Ganzi Tibetan Autonomous Prefecture of Sichuan Province and Qinghai Province, China, were identified on the basis of the chloroplast genome. The structural characteristics, population genetics, phylogenetic relationships, and phylogenetic trees were documented.

2. Materials and methods

2.1. DNA sequencing, assembly, and validation of the chloroplast genome

The leaves of four plants were collected from Ganzi Tibetan Autonomous Prefecture of Sichuan Province and Qinghai Province, China: *C. jubata* location: E 97˚11′18″, N 32˚37′19″,

altitude: 4372 m; *C. erinacea* location: E 98˚18′13″, N 33˚3′34″, altitude: 3974 m; *C. opulens* location: E 101˚5′42″, N 31˚0′8″, altitude: 2932 m; and *C. bicolor* location: E 100˚40′26″, N 31˚24′25″, altitude: 3199 m. The whole genomic DNA of *Caragana* species was extracted by using E.Z.N.A® Plant DNA kit [23]. The library was started by reagent (TruSeq™ Nano DNA Sample Prep Kit, Illumina) at 1 μg of DNA, and the DNA was interrupted to 300–500 bp by Covaris M220 ultrasound. Libraries were enriched, and eight cycles were amplified by polymerase chain reaction (PCR). Quantification was performed using TBS380 (Picogreen). Bridge PCR amplification was performed on cBot Truseq PE Cluster Kit v3-cBot-HS to generate clusters. Sequencing was conducted with 150 bp pair-end reads on the Illumina NovaSeq platform (Illumina, San Diego, CA, USA).

The cp reads were used to assemble sequences by spades, abyss, and soapdenovo. All of the contigs were aligned to the reference cp genome of *C. korshinskii* with MUMmer. Finally, the assembly results were inhole repaired with GapCloser-1.12 (OMEGA) [24].

2.2. Gene annotation and sequence analyses

Sequences were annotated by Plann [25] using the chloroplast genome of *C. korshinskii* from NCBI and some manual corrections. BLAST and Apollo [26] were used to check the start and stop codons and intron/exon boundaries with the cp genome of *C. korshinskii* as the reference sequence. The complete chloroplast genome sequence data reported in this paper have been deposited in the Genome Warehouse in National Genomics Data Center (NGDC https://ngdc.cncb.ac.cn/, accession numbers: GWHBJYO00000000, GWHBJYN00000000, GWHBJYM00000000, and GWHBJYL00000000). The structural features of the chloroplast genome were drawn by Organellar Genome DRAW [27] (http://ogdraw.mpimp-golm.mpg.de/). Protein-coding gene sequences were extracted by Geneious.

2.3. Comparison of chloroplast genomes

The chloroplast genomes of *Caragana* species were completed by mVISTA [28] (Shuffle-LAGAN mode) using the genome of *C. korshinskii* as the reference. The detecting and testing of forward, palindromic, and tandem repeats were performed using Tandem Repeats Finder [26] and REPuter [27]. In addition, the detection of simple sequence repeats (SSRs) was executed using Misa.pl [29]. The search parameters of mononucleotides, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotide were set to $\geq$10, $\geq$8, $\geq$4, and $\geq$3 repeat units.

2.4. Phylogenetic analyses

Phylogenetic trees were constructed using plastid genomes of 18 species, in which *Ranitomeya imitator* was an outgroup. The sequences were aligned by Mafft. An unrooted phylogenetic tree with 1000 bootstrap replicates was inferred using the neighbor-joining (NJ) approach with MEGA X [30].

3. Results

3.1. DNA features of the chloroplast genome of four *Caragana* species

The size of chloroplast genome was between 127,862 and 132,780 bp, which is small because of the loss of the IR region. *C. opulens* (132780 bp) had the largest chloroplast genome, whereas *C. jubata* had the smallest (127862 bp). The mean value of the total guanine + cytosine (G+C) content of four *Caragana* species was 34.72%. Four *Caragana* species had a chloroplast genome with a similar structure, and all of them loss the IR region. After annotation, the

Table 1. Summary of complete chloroplast genomes for four *Caragana* species.

Scample name	Total length	GC Content (%)	Depth	accession number
C. jubata	127862	34.49	749	GWHBJYN00000000
C. bicolor	131318	34.54	979	GWHBJYL00000000
C. erinacea	130968	35.15	451	GWHBJYM00000000
C. opulens	132780	34.70	992	GWHBJYO00000000

https://doi.org/10.1371/journal.pone.0272990.t001

whole chloroplast genome sequence of the four *Caragana species* was submitted to NGDC: the accession numbers are listed in Table 1.

The gene maps of *C. jubata*, *C. bicolor*, *C. erinacea*, and *C. opulens* were drawn by OGDraw [24] (Fig 1) based on annotation results. A total of 112 genes were found in the chloroplast genome of *C. jubata*, *C. bicolor*, *C. erinacea*, and *C. opulens*: 30 tRNA, 4rRNA, and 78 protein-coding genes. Among the 112 genes, 43 were photosynthesis-related genes (Table 2). Most genes could be divided crudely into three groups: "photosynthesis-related," "self-replication-related,", and "other" groups (Table 2) [31].

Of the anticipated genes of the chloroplast genomes of *C. jubata*, *C. bicolor*, *C. erinacea*, and *C. opulens*, introns were discovered in 17 genes: six tRNA(*trnV-UAC*, *trnL-CAA*, *trnG-UCC*, *trnA-UGC*, *trnK-UUU*, and *trnI-AUC*) genes and 11 protein-encoding genes (*rpl2*, *rps12*, *rpoC1*, *rpl16*, *ndhA*, *ndhB*, aptF, *clpP*, *petB*, *pafI*, and *petD*; Table 3). Most of the 17 intron-containing genes were inserted by one intron except for *pafI*, which was inserted by two introns. In the four cp-genomes, the longest intron all were *trnH-UUU*, which in length (2481, 2494, 2494, and 2485 bp), and contained the whole *matK*.

3.2 Analyses of long repetitive sequences and SSRs

For *C. jubata*, *C. bicolor*, *C. erinacea*, *and C. opulens*, interspersed repeated sequences (IRSs) were evaluated in the chloroplast genome with a repeat-unit length of $\geq$ 20 bp. These sequences comprised only forward reverse and palindromic repeats, yet they lacked complementary repeats that are common in other species. Among them, a total of 50 IRSs were found. Among all types of IRS, the sequence lengths in the range of 20–39 bp occurred most frequently in *C. jubata* and 40–59 bp occurred most frequently in *C. erinacea*. Those in the range of 60–79 bp and $\geq$ 100 bp occurred most frequently in *C. opulens*. IRS analyses of chloroplast genomes are shown in Fig 2.

The key mutational mechanism generating SSR polymorphism is as follows: SSRS tended to undergo slipped-strand mispairing [32]. However, SSRs in chloroplast genomes are often used as genetic markers in evolutionary and population genetic studies because of their variability at the intra-specific level [33, 34]. We found 58 SSRs in *C. jubata*, 29 SSRs in *C. bicolor*, 61 SSRs in *C. opulens*, and 74 SSRs in *C. erinacea* (Fig 3).

3.3. Comparative genomic analysis

In elucidating the differences in genomic sequences of *C. jubata*, *C. erinacea*, *C. opulens*, and *C. bicolor*, we used mVISTA to detect sequence variations using the sequence in *C. bicolor* as a reference (Fig 4). The four genomic sequences are highly similar. However, in some intergenic spacer (IGS) regions and partial sequences, significant differences are found, such as the IGS between *trnK-UUU* and *rbcl*; *trnF-GAA* and *ndhJ*; *trnL-CAA* and *trnT-UGU*; *rpoB* and *trnC-GCA*; *petA* and *psbL*; *psbE* and *pebL*; and sequences of the *rpoC*, *ycf1*, and *ycf2*. The non-coding regions have different degrees of divergence, whereas the protein coding regions are highly conserved. This finding indicated that the IGS of the *Caragana* genus evolved rapidly.

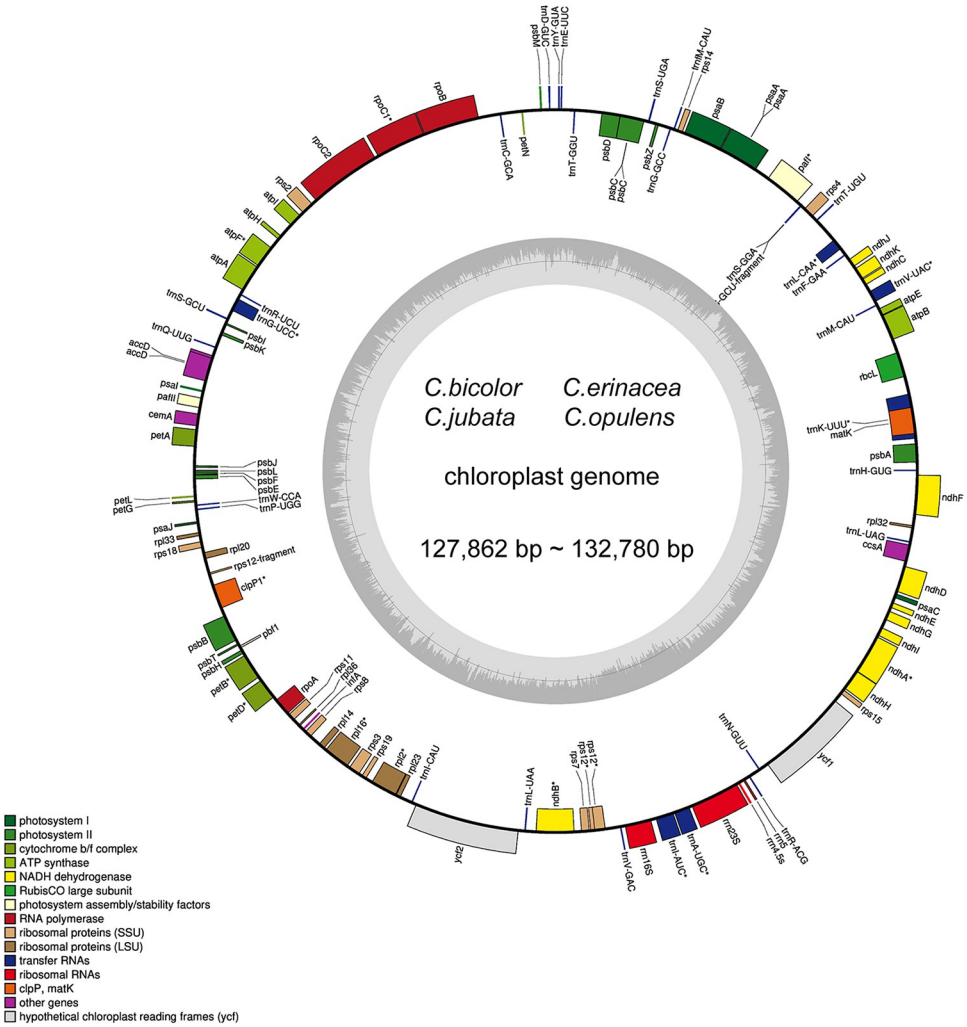

Fig 1. Gene map of the chloroplast genome of four *Caragana* species. Genes within the circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes belonging to different functional groups are color coded. The dark gray in the inner circle corresponds to the DNA G+C content, whereas the light-gray corresponds to the A+T content.

https://doi.org/10.1371/journal.pone.0272990.g001

Highly variable regions can be used to exploit markers for identification and further phylogenetic study.

3.4 Phylogenetic analyses

In determining the phylogenetic position of *Caragana* species, 18 complete chloroplast genome sequences of the Fabaceae family were constructed using the NJ tree (Fig 5). The other 14 species belong to *Lens*, *Medicago*, *Caragana*, *Astragalus*, *Glycyrrhiza*, *Lotus*, *Millettia*, *Vigna*, *Phaseolus*, *Lupinus*, and *Mimosa*.

The results showed that eight species from *Caragana* were relatives and categorized together. The following pairs showed a closer relationship: *C. kozlowii* and *C. erinacea*, *C. microphylla* and *C. korshinskii*, and *C. opulens* and *C. bicolor*. The genera *Astragalus* and *Caragana* were classified into the *Subtrib*. Astragalinae: *C. jubata* belongs to *Ser*. *Jubatae*; *C. bicolor* belongs to *Ser*. *Occidentales*; *C. erinacea* belongs to *Ser*. *Spinosae*, and *C. opulens* belongs to *Ser*.

Table 2. Genes in the chloroplast genome of four *Caragana* species.

Category	Group of genes	Name of genes
Self-replication	Large subunit of ribosomal proteins	*Rp114*, *16*, *2*, *20*, *22*, *23*, *32*, *33*, *36*
	Small subunit of ribosomal proteins	*rps2*, *3*, *4*, *7*, *8*, *11*, *12*, *14*, *15*, *18*, *19*
	DNA-dependent RNA polymerase	*rpoA*, *B*, *C1*, *C2*
	rRNA genes	*rrn16S*[a], *rrn23S*[a], *rrn4.5S*[a], *rrn5S*[a]
	tRNA genes	*trnA-UGC*, *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC*, *trnH-GUG*, *trnI-AUC*, *trnI-CAU*, *trnK-UUU*, *trnL-CAA*, *trnL-UAA*, *trnL-UAG*, *trnM-CAU*, *trnN-GUU*, *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*, *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*, *trnV-UAC*, *trnW-CCA*, *trnY-GUA*
Photosynthesis	Photosystem I	*psaA*, *B*, *C*, *I*, *J*
	Photosystem II	*psbA*, *B*, *C*, *D*, *E*, *F*, H, *I*, *J*, *K*, *L*, *M*, *T*, *Z*,
	NADH oxidoreductase	*ndhA*[*], *B*[*,a], *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*
	Cytochrome b6/f complex	*petA*, *B*[*], *D*[*], *G*, *L*, *N*
	ATP synthase	*atpA*, *B*, *E*, *F*[*], *H*, *I*
	Rubisco	*rbcL*
Other genes	Maturase	*matK*
	Protease	*clpP*[*]
	Envelope membrane protein	*cemA*
	Subunit acetyl-CoA-carboxylase	*accD*
	c-Type cytochrome synthesis gene	*ccsA*
	Conserved open-reading frames	*Ycf1*, *2*
	protein synthesis initiation factor	*infA*

https://doi.org/10.1371/journal.pone.0272990.t002

Table 3. The genes having introns in the chloroplast genome of four *Caragana* species and the length of the exons and introns.

Species	Gene	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
C. jubata	*trnL-CAA*	37	350	50		
	trnV-UAC	39	574	37		
	trnA-UGC	38	812	35		
	trnG-UCC	21	682	51		
	trnI-AUC	37	953	35		
	trnK-UUU	37	2481	29		
	rps12[*]	26	232	591		
	rpl16	399	1105	9		
	rpl2	434	689	394		
	rpoC1	430	790	1625		
	ndhA	553	1163	539		
	ndhB	762	685	777		
	aptF	145	702	410		
	petB	6	819	642		
	clpP	231	786	292		
	petD	8	717	475		
	pafI	124	711	230	886	153

(*Continued*)

Table 3. (Continued)

Species	Gene	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
C. erinacea	*trnL-CAA*	37	555	50		
	trnV-UAC	39	572	37		
	trnA-UGC	38	808	35		
	trnG-UCC	21	679	51		
	trnI-AUC	37	956	35		
	trnK-UUU	37	2494	29		
	*rps12**	26	232	588		
	rpl16	399	945	9		
	rpl2	434	689	388		
	rpoC1	430	785	1625		
	ndhA	551	1194	541		
	ndhB	723	680	762		
	petB	6	822	642		
	clpP	363	843	231		
	petD	8	729	475		
	aptF	145	697	410		
	pafI	124	714	230	891	153
C. bicolor	*trnV-UAC*	39	572	37		
	trnL-CAA	37	555	50		
	trnK-UUU	37	2494	29		
	trnI-AUC	37	956	35		
	trnA-UGC	38	808	35		
	rps12	26	232	588		
	rpoC1	430	785	1625		
	rpl2	390	686	435		
	rpl16	399	945	9		
	petD	8	729	475		
	petB	6	822	642		
	pafI	126	714	226	981	155
	ndhB	723	680	762		
	ndhA	551	1194	541		
	clpP1	363	843	231		
	atpF	167	673	412		
C. opulens	*trnV-UAC*	39	574	37		
	trnL-CAA	37	534	50		
	trnK-UUU	37	2485	29		
	trnI-AUC	37	953	35		
	trnG-UCC	21	682	51		
	trnA-UGC	38	810	35		
	rps12	26	232	600		
	rpoC1	430	789	1625		
	rpl2	396	692	435		
	rpl16	399	1101	9		
	petD	8	720	475		
	petB	6	826	642		
	pafI	126	702	226	873	155
	ndhB	723	685	762		
	ndhA	551	1169	541		
	clpP1	363	1628	223		
	atpF	167	679	412		

https://doi.org/10.1371/journal.pone.0272990.t003

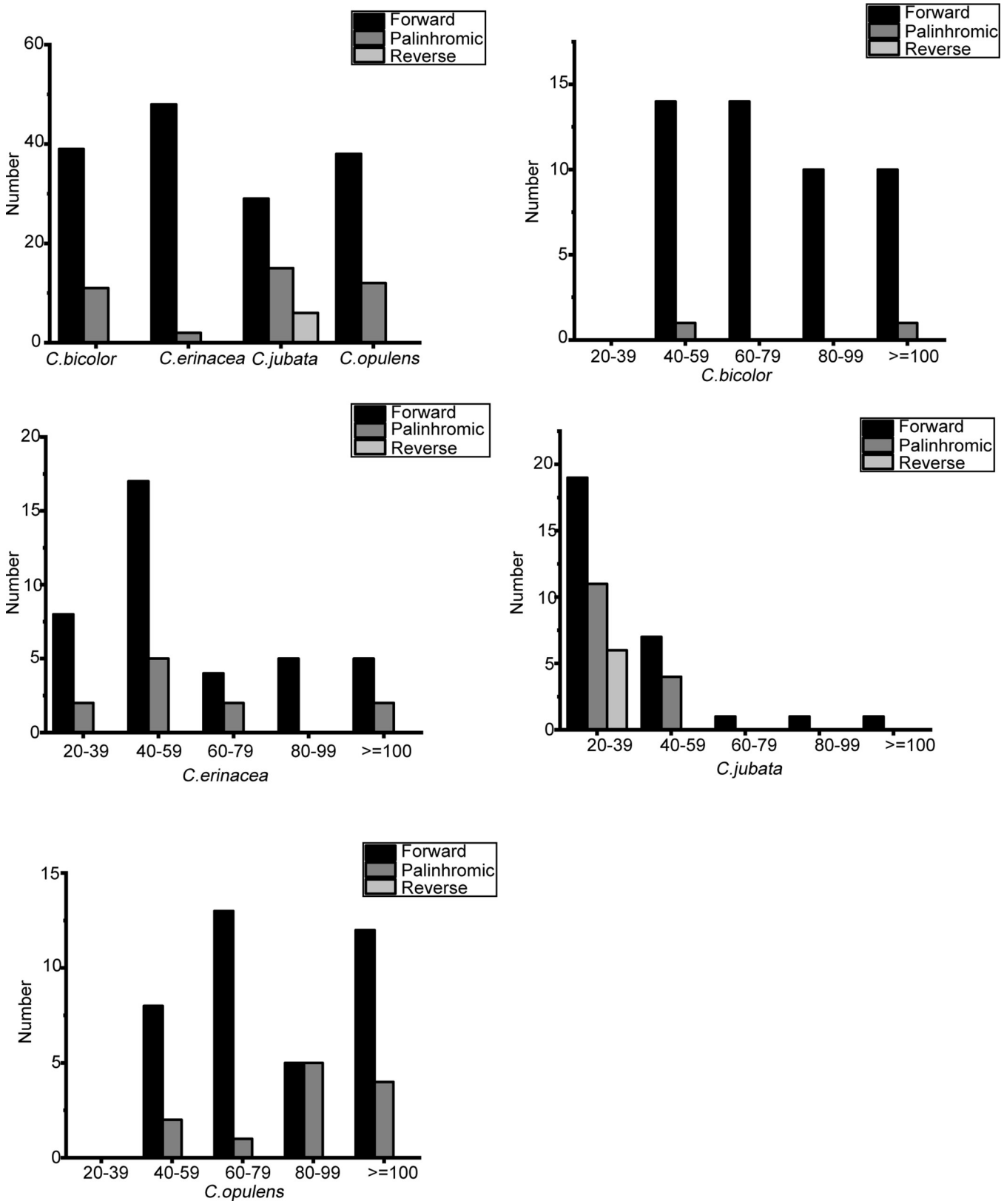

Fig 2. Long repetitive sequences in chloroplast genomes of four *Caragana* species.

https://doi.org/10.1371/journal.pone.0272990.g002

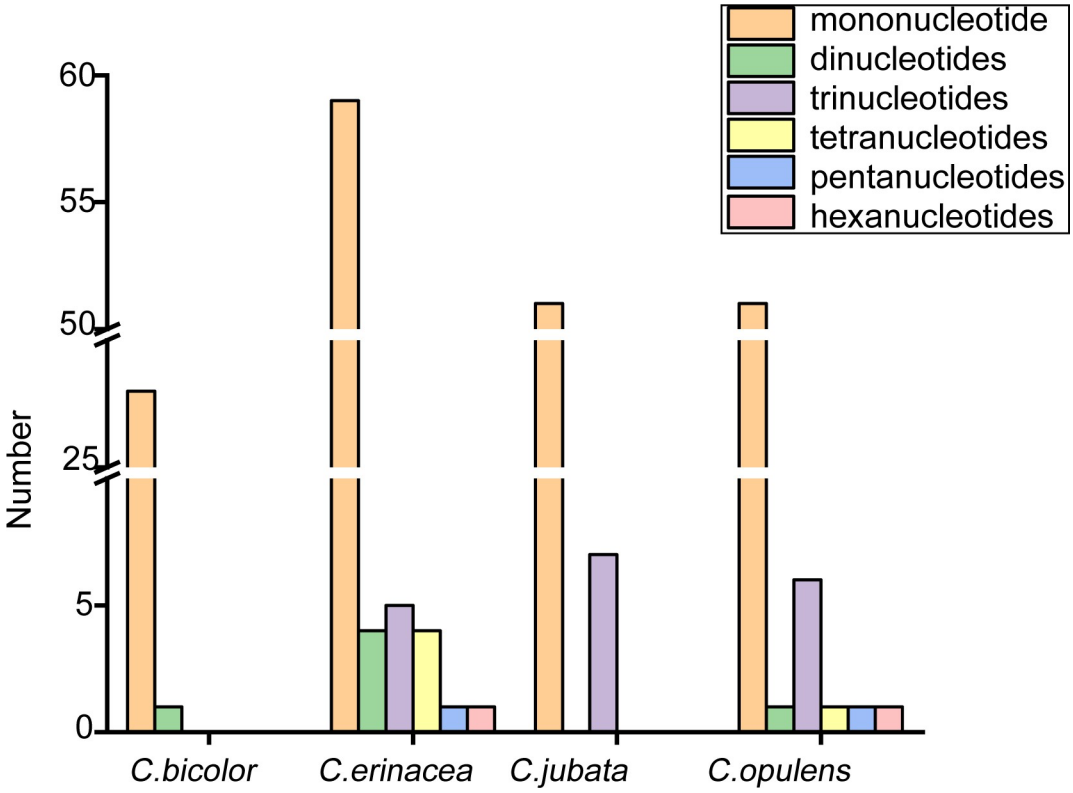

Fig 3. SSR distribution in chloroplast genomes of four Caragana species.

https://doi.org/10.1371/journal.pone.0272990.g003

Grandiflorae Pojark˙ [1]. This result also fitted the clustering results based on ITS2 sequences [35].

4. Conclusions and discussion

The chloroplast genomes of 10 species of the genus *Caragana* have been published in the National Center for Biotechnology Information; chloroplast genomes were between 127,103 and 133,122 bp in size, and they contained 110–111 genes (30–31 tRNA, 4 rRNA, and 76 protein-coding genes). Multiple species of *Caragana* have been reported to loss the IR region, such as *C. rosea*, *C. microphylla*, and *C. intermedia*[12–14]. The chloroplast genomes of *C. jubata*, *C. erinacea*, *C. opulens*, and *C. bicolor* showed high similarity with regard to gene deletion, genome size, gene sequences, gene classes, and distribution of repeat sequences, and the lacked IRLC. An important indicator of species affinity is the content of DNA G + C [36], and the four *Caragana* species in this study have highly similar cpDNA G+C content.

IRS analyses of chloroplast genomes show that the four species lacked complementary repeats. Comparative cp-genomic analysis suggested that the cp genome structure of *Caragana* was well conserved. Highly variable regions are primarily distributed in non-coding and partial coding regions, which can be used to exploit markers for identification and further phylogenetic study.

Intron and/or gene losses in chloroplast genomes have been reported in considerable literature [37–39]. Introns can play an important role in the regulation of gene expression in a temporal and tissue-specific manner [39–41]. Regulatory mechanisms of introns in some plants and animals have been reported [42–44]. However, the relationship between intron deletion

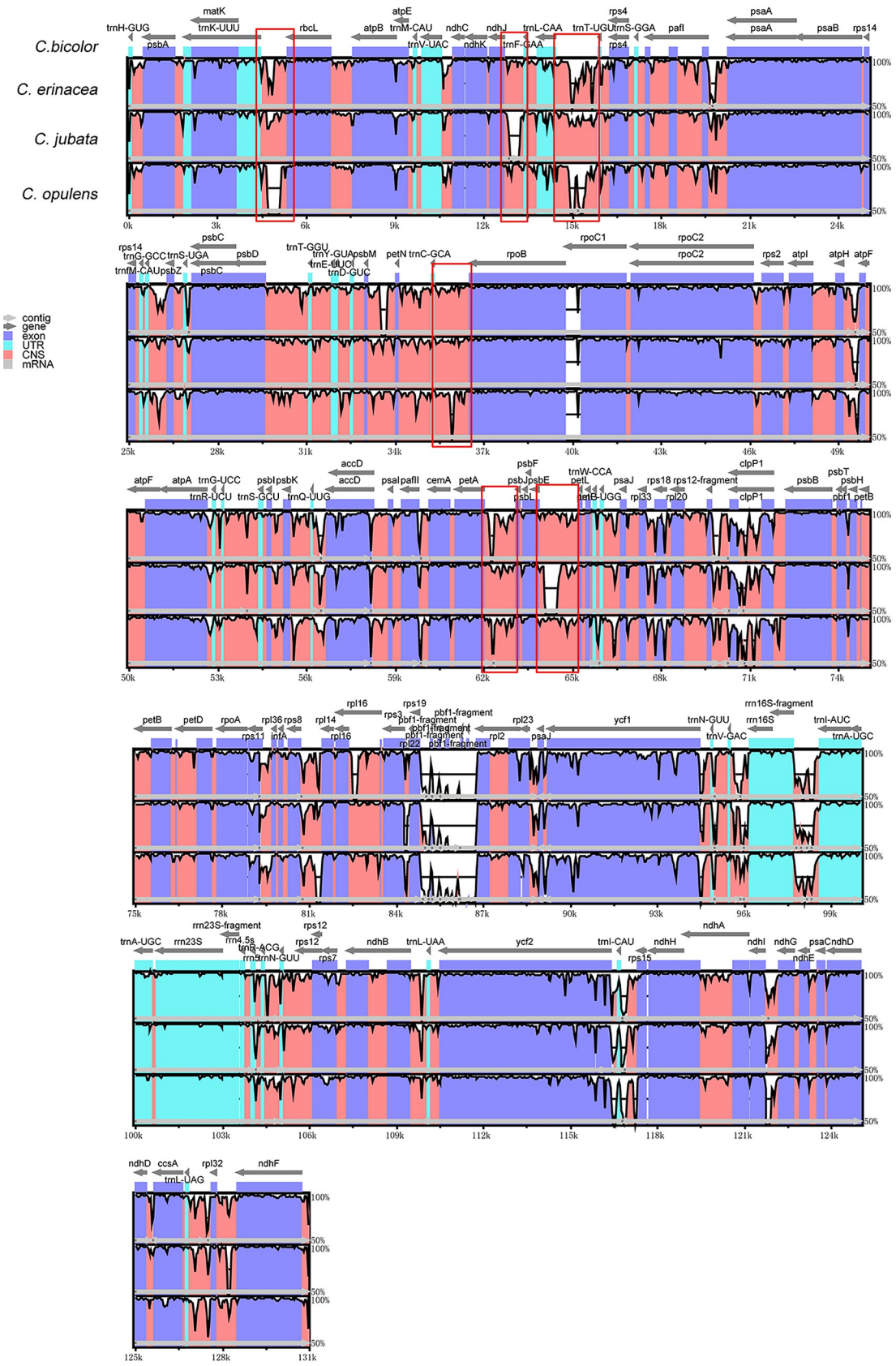
C.bicolor
C. erinacea
C. jubata
C. opulens
contig
gene
exon
UTR
CNS
mRNA

Fig 4. Comparative analyses of genomic differences in chloroplasts of four *Caragana* species. Gray arrows and thick black lines above the alignment indicate gene orientation. Purple bars represent exons; blue bars denote untranslated regions (UTRs); pink bars represent non-coding sequences (CNS), and gray bars denote mRNA. The y-axis represents percentage identity.

https://doi.org/10.1371/journal.pone.0272990.g004

and gene expression in *Caragana* by transcriptome has not been published. Therefore, further research into the role of introns in *Caragana* species is necessary.

Advances in phylogenetic analysis can reveal the evolution of chloroplast genomes, including nucleotide substitutions and structural changes [45, 46]. Our results of phylogenetic analysis were consistent with the status of the major taxa within the genus *Caragana* [1]. Species from the genus *Caragana* were monophyletic, and *C. jubata*, *C. erinacea*, *C. opulens*, and *C. bicolor* could be differentiated from other *Caragana* species. The current study demonstrated that chloroplast genomes can be used for the identification and classification of *Caragana* species. In addition, the phylogenetic relationship of *Caragana* species is related to elevation and geographical distribution (GD). For example, *C. kozlowii* (altitude: 3100–4300 m; GD: Qinghai, E Xizang) and *C. erinacea* (altitude: 2000–4000 m, GD: Qinghai, Xizang, Gansu, Ningxia, W Sichuan, NW Yunnan) are distributed at altitudes of 3100–4000 levels in Qinghai and Tibet. In addition, *C. microphylla* (altitude: 1800–3800 m, GD: Nei Mongol, Jilin, Liaoning) and *C. korshinskii* (altitude: 900–2400 m, GD: Nei Mongol, Gansu, Ningxia, Qinghai, W Shanxi) are distributed at altitudes of 1800–2400 levels in Nei Mongo. The phylogenetic relationship of *Caragana* species partially consistent with elevation and GD. This finding may provide a reference for further study on the relationship between distribution and evolution of *Caragana* species.

Caragana species have a large altitude span and wide GD, showing strong environmental adaptability[1–5]. Our results can provide valuable information for genetic transformation, the development of population genetic surveys, and evolutionary studies. Plastids contain a

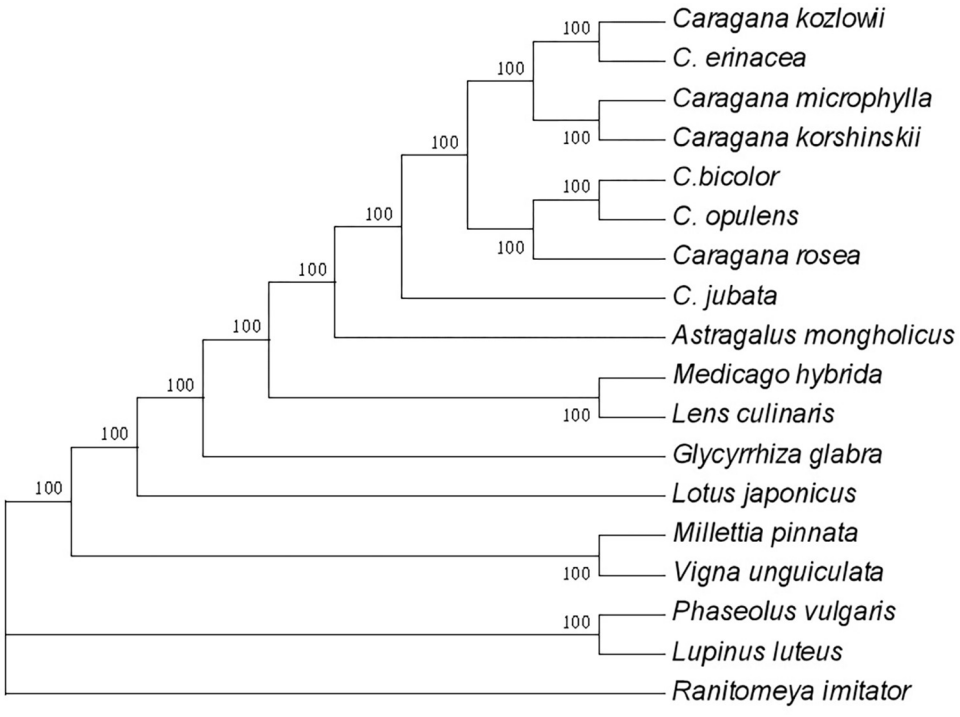

Fig 5. NJ tree based on chloroplast genomes of the Fabaceae family.

https://doi.org/10.1371/journal.pone.0272990.g005

range of genes associated with photosynthesis, and photosystem II is a key component of high temperature, drought stress, and many other stresses [47, 48]. However, the strong environmental adaptability mechanism of the genus *Caragana* remains unclear because of the lack of research and data[3, 12]. Our results can provide data for further investigation of the discovery of adaptability and strong adversity resistance genes of *Caragana* species. Our research data complement the database of herbgenomics [49, 50].

Author Contributions

Data curation: Maohua Yuan, Xianmei Yin.

Formal analysis: Maohua Yuan, Xianmei Yin.

Investigation: Bixing Gao.

Methodology: Maohua Yuan, Guihua Jiang.

Project administration: Rui Gu, Guihua Jiang.

Resources: Bixing Gao.

Validation: Maohua Yuan, Xianmei Yin.

Visualization: Xianmei Yin.

Writing – original draft: Maohua Yuan.

Writing – review & editing: Rui Gu, Guihua Jiang.

References

1. Wu Z. Flora of China. Beijing, China: Science Press; 2010. 528–45 p.
2. Meng Q, Niu Y, Niu X, Roubin RH, Hanrahan JR. Ethnobotany, phytochemistry and pharmacology of the genus Caragana used in traditional Chinese medicine. Journal of Ethnopharmacology. 2009; 124 (3). https://doi.org/10.1016/j.jep.2009.04.048 PMID: 19422901
3. Kang H-M, Chen K, Bai J, Wang G. Antioxidative system's responses in the leaves of six Caragana species during drought stress and recovery. Acta Physiologiae Plantarum. 2012; 34(6).
4. Xiao C-W, Sun OJ, Zhou G-S, Zhao J-Z, Wu G. Interactive effects of elevated CO2 and drought stress on leaf water potential and growth in Caragana intermedia. Trees Structure and Function. 2005; 19(6).
5. Khan S, Nazir M, Raiz N, Saleem M, Zengin G, Fazal G, et al. Phytochemical profiling, in vitro biological properties and in silico studies on Caragana ambigua stocks (Fabaceae): A comprehensive approach. Industrial Crops & Products. 2019; 131.
6. Luo HF, Zhang LP, Hu CQ. ChemInform Abstract: Five Novel Oligostilbenes from the Roots of Caragana sinica. ChemInform. 2001; 32(37).
7. Kakorin PA, Tereshkina OI, Ramenskaya GV. Potential Biological Activity and Chemical Composition of Caragana Jubata (Pall.) Poir. (Review). Pharmaceutical Chemistry Journal. 2018; 52(6).
8. A KP, V BI, O TY, V RG, A DT, G KV. [Hepatoprotective activity of aqueous extract from Caragana jubata (Pall.) Poir. shoots in the model of acute hepatitis induced by acetaminophen in rats]. Biomeditsinskaia khimiia. 2018; 64(3).
9. KP A., PI B., RE D., ÉK I., RG V., PL A., et al. Biologically Active Compounds in Aqueous Extracts of Caragana jubata (Pall.) Poir. Pharmaceutical Chemistry Journal. 2018; 51(11).
10. Liaquat A, Saima K, Mamona N, Naheed R, Saima N, Gokhan Z, et al. Chemical profiling, in vitro biological activities and Pearson correlation between phenolic contents and antioxidant activities of Caragana brachyantha Rech.f. South African Journal of Botany. 2021; 140.
11. Kamila Z, Gihwan L, Aizhamal B, Bari SA, Ho KJ, Yoon KJ, et al. Inhibitory mechanism of O-methylated quercetins, highly potent β-secretase inhibitors isolated from Caragana balchaschensis (Kom.) Pojark. Journal of Ethnopharmacology. 2021; 272.
12. Jiang M, Chen H, He S, Wang L, Chen AJ, Liu C. Sequencing, Characterization, and Comparative Analyses of the Plastome of Caragana rosea var. rosea. International Journal of Molecular Sciences. 2018; 19(5).

13. Zhang Z-l, Ma L-y, Yao H-b, Yang X, Luo J-h, Gong X, et al. Complete chloroplast genome of Caragana intermedia (Fabaceae), an endangered shrub endemic to china. Conservation Genetics Resources. 2016; 8(4).

14. Liu B-b, Duan N, Zhang H-l, Liu S, Shi J-w, Chai B-f. Characterization of the whole chloroplast genome of Caragana microphylla Lam (Fabaceae). Conservation Genetics Resources. 2016; 8(4).

15. Rehder A. Bibliography of Cultivated Trees and Shrubs Hardy in the Cooler Temperate Regions of the Northern Hemisphere. Arnold Arboretum of Harv. 1949.

16. O BJ, M RA, C ZP. Chloroplasts extend stromules independently and in response to internal redox signals. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112(32).

17. Wanjun L, Dapeng N, Yujun W, Junjie S, Xincun W, Dan Y, et al. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of Astragalus membranaceus. Scientific reports. 2016; 6(1).

18. D PJ, F TW. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell. 1982; 29(2).

19. Jamal S, Erika S, Nicholas E, Jin Z, A BN, Muhammed M, et al. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. Plant biotechnology journal. 2014; 12(6).

20. Domingos C, P dQL, Toby PR, C dLH, Emile F, F WM, et al. Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. American journal of botany. 2012; 99(12).

21. Yingchun H, Quan Z, Guangyuan R, Sodmergen. Occurrence of plastids in the sperm cells of Caprifoliaceae: biparental plastid inheritance in angiosperms is unilaterally derived from maternal inheritance. Plant & cell physiology. 2008; 49(6).

22. Daniell H, Kumar S, Dufourmantel N. Breakthrough in chloroplast genetic engineering of agronomically important crops. Trends in Biotechnology. 2005; 23(5). https://doi.org/10.1016/j.tibtech.2005.03.008 PMID: 15866001

23. Jun-Bo Y, De-Zhu L, Hong-Tao L. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Molecular ecology resources. 2014; 14(5).

24. D AR, J TJ, Ibai I-A, Ferdinand M, Carlos G-M, Elisa dlC-M, et al. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. Nature genetics. 2016; 48(3).

25. M LT, R ES. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids research. 1997; 25(5).

26. A FK, Lior P, Alexander P, M RE, Inna D. VISTA: computational tools for comparative genomics. Nucleic acids research. 2004; 32(Web Server issue).

27. G B. Tandem repeats finder: a program to analyze DNA sequences. Nucleic acids research. 1999; 27 (2).

28. Marc L, Oliver D, Ralph B. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Current genetics. 2007; 52 (5–6).

29. S K, V CJ, E O, C S, J S, R G. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic acids research. 2001; 29(22). https://doi.org/10.1093/nar/29.22.4633 PMID: 11713313

30. Sebastian B, Thomas T, Thomas M, Uwe S, Martin M. MISA-web: a web server for microsatellite prediction. Bioinformatics (Oxford, England). 2017; 33(16).

31. Seung-Bum L, Charalambos K, Robert J, Jessica H, Luke T, Christopher T, et al. The complete chloroplast genome sequence of Gossypium hirsutum: organization and phylogenetic relationships to other angiosperms. BMC Genomics. 2006; 7(1).

32. T C, M H, C G, R J. A generic intron increases gene expression in transgenic mice. Molecular and cellular biology. 1991; 11(6). https://doi.org/10.1128/mcb.11.6.3070-3074.1991 PMID: 2038318

33. Sajjad A, Latif KA, Aaqil KM, Muhammad W, Sang-Mo K, Byung-Wook Y, et al. Chloroplast genomes of Arabidopsis halleri ssp. gemmifera and Arabidopsis lyrata ssp. petraea: Structures and comparative analysis. Scientific reports. 2017; 7(1).

34. Wenpan D, Chao X, Tao C, Kui L, Shiliang Z. Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of saxifragales. Genome biology and evolution. 2013; 5(5).

35. Hao D, Zhang Y, Fu S. Molecular identification of Tibetan medicinal plants based on ITS2 sequence. Chinese Traditional and Herbal Drugs. 2019; 50(12):2967–75.

36. Koichiro T, Daniel P, Nicholas P, Glen S, Masatoshi N, Sudhir K. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution. 2011; 28(10).

37. Feng Y, Liying C, Claude d, Bernard M, Jijun T. Gene rearrangement analysis and ancestral order inference from chloroplast genomes with inverted repeat. BMC Genomics. 2008; 9(Suppl+1).

38. Downie SR, Llanas E, Downie DSK. Multiple Independent Losses of the rpoC1 Intron in Angiosperm Chloroplast DNA's. Systematic Botany. 1996; 21(2).

39. BR G. Alternative splicing: increasing diversity in the proteomic world. Trends Genet2001. 2 p.

40. Jiang X, Yang C, Baosheng L, Shuiming X, Qinggang Y, Rui B, et al. Panax ginseng genome examination for ginsenoside biosynthesis. GigaScience. 2017; 6(11).

41. Hir HL, Nott A, Moore MJ. How introns influence and enhance eukaryotic gene expression. Trends in Biochemical Sciences. 2003; 28(4). https://doi.org/10.1016/S0968-0004(03)00052-5 PMID: 12713906

42. Deng-Ke N, Yu-Fei Y. Why eukaryotic cells use introns to enhance gene expression: splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity. Biology direct. 2011; 6(1).

43. J C, M F, V W. Introns increase gene expression in cultured maize cells. Genes & development. 1987; 1 (10).

44. Shahram E, Dinah A, Ian K, B RA. The effects of a stimulating intron on the expression of heterologous genes in Arabidopsis thaliana. Plant biotechnology journal. 2013; 11(5).

45. Saski C, Lee S-B, Daniell H, Wood TC, Tomkins J, Kim H-G, et al. Complete Chloroplast Genome Sequence of Glycine max and Comparative Analyses with other Legume Genomes. Plant Molecular Biology. 2005; 59(2).

46. C HR, Matthew FH, L BJ, K JR. Extensive rearrangements in the chloroplast genome of Trachelium caeruleum are associated with repeats and tRNA genes. Journal of molecular evolution. 2008; 66(4).

47. Mulo P, Sakurai I, Aro EM. Strategies for psbA gene expression in cyanobacteria, green algae and higher plants: from transcription to PSII repair. Biochimica et biophysica acta. 2012; 1817(1):247–57. Epub 2011/05/14. https://doi.org/10.1016/j.bbabio.2011.04.011 PMID: 21565160.

48. Yamamoto Y, Aminaka R, Yoshioka M, Khatoon M, Komayama K, Takenaka D, et al. Quality control of photosystem II: impact of light and heat stresses. Photosynthesis research. 2008; 98(1–3):589–608. Epub 2008/10/22. https://doi.org/10.1007/s11120-008-9372-4 PMID: 18937045.

49. Xianmei Y, Baosheng L, Shuai G, Conglian L, Jin P, Jiang X, et al. The chloroplasts genomic analyses of Rosa laevigata, R. rugosa and R. canina. Chinese medicine. 2020; 15(1).

50. Shu-Dong Z, Jian-Jun J, Si-Yun C, W CM, E SD, Hong-Tao L, et al. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. The New phytologist. 2017; 214(3).