JSE Journal of Systematics and Evolution

doi: 10.1111/jse.13009

Research Article

DNA barcoding of recently diverging legume genera: Assessing the temperate Asian *Caragana* (Fabaceae: Papilionoideae)

Shabir A. Rather[1], Josphat K. Saina[1], Arjun Adit[2], Hong-Mei Liu[1]*, Zhao-Yang Chang[3], and Harald Schneider[1]

[1]Center for Integrative Conservation & Yunnan Key Laboratory for Conservation of Tropical Rainforests and Asian Elephants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Menglun 666303, Yunnan, China
[2]Centre for Functional Biodiversity, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg 3209, South Africa
[3]College of Life Sciences, Northwest A & F University, Yangling 712100, Shaanxi, China
*Author for correspondence. E-mail: liuhongmei@xtbg.ac.cn
Received 5 April 2023; Accepted 15 June 2023

Abstract Obtaining reliable species identification of the legume genus *Caragana* has been challenging. Until now, species identification was mostly carried out utilizing diagnostic morphological characteristics, in addition to some successful applications of secondary chemical compounds. This study was designed to establish a DNA barcoding protocol enabling unambiguous identification of 238 accessions belonging to 67 species of *Caragana*. The performance of four DNA barcoding regions nrITS, *trnH-psbA*, *matK*, and *rbcL* was explored using three analytical approaches, Pairwise Genetic Distance, Sequence Similarity and Phylogenetic Tree method. The chloroplast regions *rbcL* and *matK* showed lower discriminatory power compared with the nuclear region *internal transcribed spacer* (ITS) and the chloroplast region *trnH-psbA*. The nrITS outperformed the other regions in the resolution rate. The present study brings forth an efficient barcode locus for *Caragana*. A barcode based either on a single-locus nrITS or the combination of nrITS and *trnH-psbA* was found to be most suitable for species discrimination with distinctive barcoding gaps. An attempt has also been made to resolve taxonomic issues in the *Caragana opulens* complex. DNA barcoding tools when complemented with alpha taxonomic evidence can aid in solving complex systematic problems, especially when taxa are characterized by overlapping traits, such as species belonging to the Fabaceae family.

Key words: ITS, molecular taxonomy, species complex, species discrimination, systematics, *trnH-psbA*.

1 Introduction

The ecologically and pharmacologically important legume genus *Caragana* Fabr. contains ~91 species occurring throughout the boreal and temperate climate zones of the Northern Hemisphere, particularly in arid and semiarid areas (Zhao, 1993; Zhang, 1997; Zhang et al., 2009, 2016; Rather et al., 2021). With ~80 species, the Euro-Asian range, which extends toward Japan, Korea, and Siberia in the north and northeast, toward central Asia and Europe in the west and along the Himalayas toward Northern India, Bhutan, and Nepal in the South, is considered the center of diversity of the genus (Lock, 2005; Liu et al., 2010). However, China alone hosts more than 66 species (i.e., 70% of the total species) of *Caragana* (Zhao, 1993; Zhang et al., 1996, 2002; Zhang, 1997; Zhou et al., 2002; Duan et al., 2015, 2016; Rather et al., 2021).

Caragana plays a key role in the cold desert ecosystem throughout the Asian continent (Bhardwaj et al., 2013). The roots, flowers, pods and seeds of many species like *C. sinica*, *C. pygmaea*, *C. arborescens*, *C. brevispina*, and so forth, are consumed as food throughout China (Tanaka, 1976; Manandhar, 2002). Some of the species of *Caragana* (*C. korshinskii* in particular) have proved to be suitable for the restoration of eroded, desertified and degenerated lands in China due to the high soil holding capacity imparted through drought-tolerant root systems (Wang et al., 2019). The genus is, therefore, widely popular for ecorestoration projects where the species has been shown to modify the physiological properties of the substratum resulting in improved soil stability (Yin et al., 2022). Accurate identification of taxa in the genus is therefore imperative to avoid ambiguity and select the correct species for the rehabilitation of degraded ecosystems.

Further, the demand for authentic identification of *Caragana* is extremely important, as more than 20 species are recognized for pharmacological utility (Tanaka, 1976; Niu, 1988; Manandhar, 2002; Wang et al., 2004, 2005). Several species, such as *C. arborescens*, *C. bicolor*, *C. sinica*, *C. brevifolia*, *C. franchetiana*, *C. frutex*, *C. intermedia*, *C. jubata*, *C. leucophloea*, *C. microphylla*, *C. polourensis*, *C. pygmaea*, *C. rosea*, *C. tangutica*, *C. tibetica*, and *C. stenophylla*, are represented in Chinese, Indian (Ayurvedic, Homeopathic,

Unani, and Siddha), Mongolian and Tibetan traditional medicines due to their analgesic, anti-inflammatory and antipyretic properties. In Chinese folk medicines also, the genus is widely used for the treatment of cold, toothache, skin lesions, stomach ache, headaches, asthma, fever, wound, anemia, irregular menstruation, cough, nose bleeding, hypertension, anoxia, rheumatoid arthritis, breast cancer, uterine cervical cancer, rheumatic pains, leukorrhea, mastitis, metrorrhagia, and fatigue (Jia et al., 1997; Wang et al., 2005; Meng et al., 2009). As some herbal medicines are prone to adulterations or substitution with cheap alternatives, authentic identification of utilized plants with novel techniques like DNA barcoding becomes relevant to ensure reliable screening (Luo et al., 2012; Dai et al., 2018). DNA barcode has emerged as a tool for resolving taxonomic anomalies and can also be actively employed to detect adulterations in herbal products (Yang et al., 2019; Cui et al., 2020; Skjua et al., 2020).

The taxonomic treatment of *Caragana* has been challenging because of recent radiation in the genus (Zhang et al., 2016), resulting in several species complexes due to overlapping morphological characteristics. Furthermore, many species of *Caragana* show considerable morphological variability, while some have specific ecological and habitat preferences (Zhang et al., 2009; Duan et al., 2015, 2016; Rather et al., 2021). Unambiguous species identification is often hampered by the overlap of morphological characteristics and/or limited differences in leaf shape/size or flower morphology. A further challenge is the preparation of voucher specimens and maintaining all information required to achieve reliable identification. According to Pan & Zhu (2010), pubescence, the shape of stipules and leaflets and pods are basic diagnostic characters used for preparing key of *Caragana* species, while the length of pedicels, inflorescence, and seed number have negligible taxonomic value. The absence of flowers and fruits in herbarium specimens of the *Caragana* species causes misidentifications. Using morphological characters alone may incorporate an element of ambiguity, especially when applied to highly fragmented plant materials. Finally, convergent evolution and intraspecies morphological variations render identifying and classifying *Caragana* species laborious.

This study aims to establish a DNA barcoding standard to enable reliable identification of *Caragana* species that will help in quality control procedures especially for highly fragmented herbal materials. Since the inception of DNA barcoding to land plants, a constant challenge has been to determine a universal DNA barcode that works for all land plants. For example, a barcode combining fragments of the plastid genomic genes, such as *rbcL* and *matK*, which have been endorsed as universal barcodes for many land plants (CBOL Plant Working Group, 2009); however, subsequent research showed rather poor performance of these barcodes in many plant genera (Xiang et al., 2011; Yan et al., 2011; Clement & Donoghue, 2012; Li et al., 2012; Zhang et al., 2012). Therefore, alternative barcodes such as the plastid intergenic spacer *trnH-psbA* and the nuclear ribosomal internal transcribed spacer (ITS) are still a matter of scientific discussion (Chen et al., 2010; China Plant BOL Group, 2011; Li et al., 2011a; Amritha et al., 2020; Zhang & Jiang, 2020). In this study, four widely utilized regions, namely nrITS, *trnH-psbA*, *matK*, and *rbcL* either individually and/or in combinations have been considered to determine a DNA barcode for *Caragana*. The present study is the first attempt to utilize DNA barcoding for *Caragana* and its relatives. For our assessment, the most comprehensive taxon sampling of this genus was conducted that overachieved any sampling used in previous phylogenetic studies on *Caragana*. The following research questions have been addressed: (i) Determine the performance of standardized DNA fragments to distinguish closely related *Caragana*; (ii) testing the hypothesis that *Caragana* species sampled from China form a distinct clade within the genus; (iii) determine the accuracy of *Caragana* species identified utilizing DNA barcoding; and (iv) identify the most powerful barcode for the correct species identification in *Caragana*.

2 Material and Methods

2.1 Taxon sampling

In total, 238 accessions belonging to 67 species of *Caragana* were included in the present study. Approximately 74% of the extended diversity present in the genus was sampled. *Hedysarum alpinum* and *Astragalus coluteocarpus* were incorporated as outgroup taxa. Fresh and healthy leaves were collected and stored in silica from different localities across the known range of these legumes in China (Fig. 1). Voucher specimens were deposited at the Northwest A&F University Herbarium (WUH; Table S1). The taxonomic identity of each accession was confirmed by considering published taxonomic treatments (Zhao, 1993; Liu et al., 2010).

2.2 DNA isolation, PCR amplification, and sequencing

Genomic DNA was extracted from silica-dried leaves following the modified CTAB method (Doyle & Doyle, 1987) and DNA extraction kits (DNeasy Extraction Kits; Qiagen). Postisolation, the extractions were stored at −20 °C prior to amplification with one nuclear (nrITS) and three chloroplast DNA regions (*trnH-psbA*, *matK*, *rbcL*). These regions were selected based on recent discussions of DNA barcode selection in plants (Kress & Erickson, 2007; Taberlet et al., 2007; CBOL Plant Working Group, 2009; Kress et al., 2009; Chen et al., 2010; China Plant BOL Group, 2011; Hollingsworth et al., 2011; Kress, 2017). Amplification was performed in a 25-μL reaction volume, which consisted of 2.5 mL of 10× buffer containing 2 mM $MgCl_2$, 1 U of *Taq* DNA polymerase, 1 μL of dNTP (0.125 mM), 1 μL of each primer (5 pM), and 30–50 ng of total DNA. Nuclease-free water was added to attain the final volume in the reaction mixture. Following the PCR protocol described by Boatwright et al. (2008), we made slight modifications to our approach. Specifically, we utilized 10–40 ng of DNA template and incorporated 1–3 μL of trehalose-based additives in the PCR reactions to enhance the efficiency of the primers. The inclusion of these additives proved effective in significantly improving the PCR efficiency for all the primers used in our study (Samarakoon et al., 2013). The primer and PCR program information for each primer is provided in Table S2. PCR products (2 mL) were visualized on 0.8% agarose using gel electrophoresis. The amplified products were purified using the BioMed multifunctional DNA fragment purification recovery kits (Beijing, China). The processed samples were

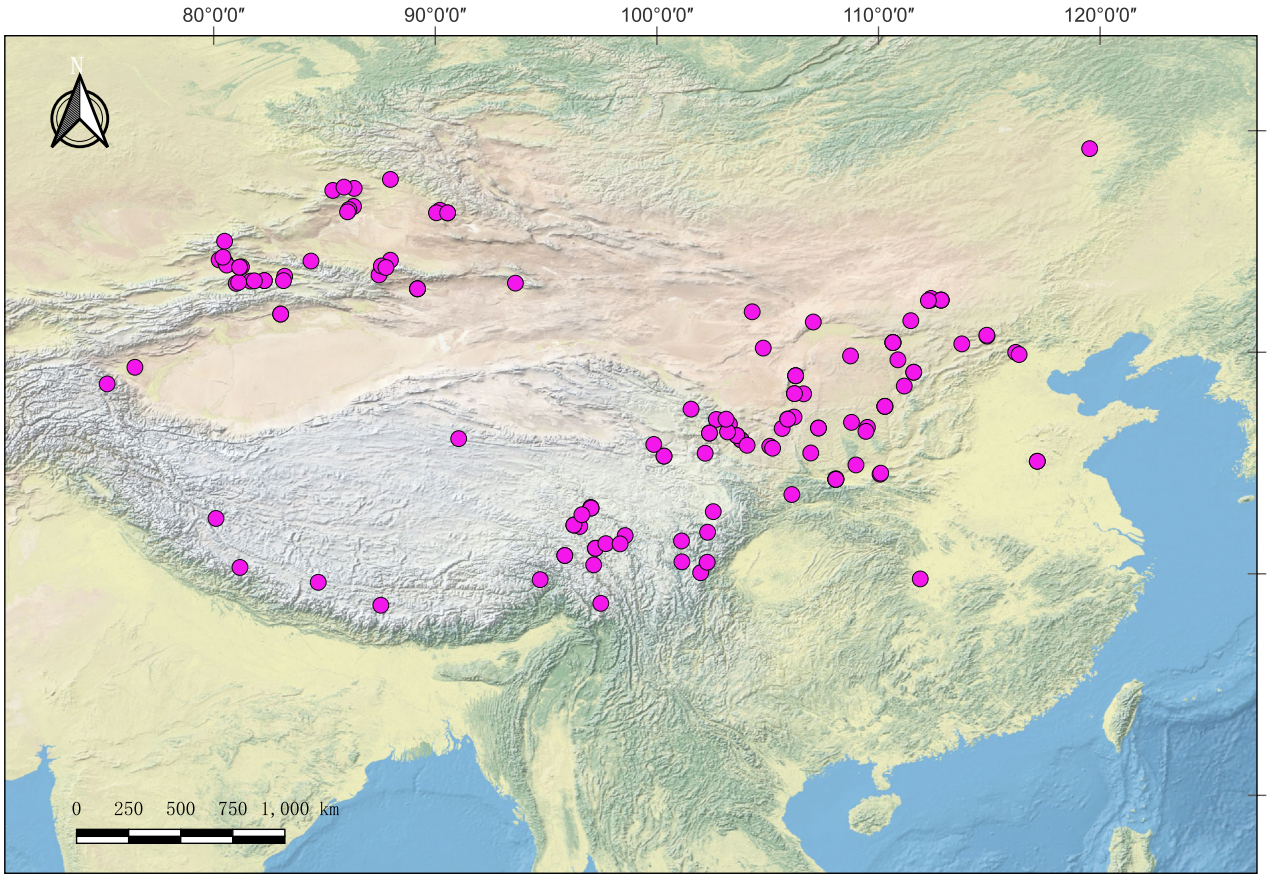


Fig. 1. Map illustrates the distribution pattern of the *Caragana* genus across different provinces in China. The pink dots on the map indicate the collection sites where various species of the *Caragana* genus have been found.

then sequenced in an ABI 3730 automated sequencer at Beijing Qingke Biological Technology Co. Ltd. (Applied Biosystems, Carlsbad, CA, USA). DNA segments that did not meet the quality requirements during the quality check were either resequenced or excluded from the final analysis.

2.3 Sequence acquisition from GenBank

The data set was expanded by tracking DNA sequences used in earlier studies (Zhang et al., 2016). The *rbcL* and *matK* sequences were retrieved from GenBank (ncbi.nlm.nih.gov). As NCBI sequence data are not necessarily linked with taxonomically validated voucher specimens, an examination of all available sequences downloaded from GenBank was carried out to ensure that correctly identified information was being used in the study.

2.4 Sequence alignment and data analysis

Each marker sequence was assembled and inspected using Sequencher 4.1 (Codes G. Sequencher), aligned with the MUSCLE (Edgar, 2004) as implemented in MEGA 7.0 (Kumar et al., 2016), and further checked manually using Se-Al version 2.0a11 (Rambaut, 2007). Gaps in the DNA sequences were treated following the suggestions of the China Plant BOL Group (2011). Both individual and all possible combinations of the four loci were used for the DNA barcoding survey. The base compositions, genetic distances, variable sites, and parsimony-informative site values were estimated using MEGA 7.0 as per the Kimura-2-parameter (K2P) model (Kimura, 1980). Three analytical approaches were employed to explore the performance of the four regions and their combination as a DNA barcode. The approaches used in these evaluations were: the Pairwise Genetic distance (PWG-distance), the Sequence Similarity (TaxonDNA), and a Phylogenetic Tree.

Intraspecific and interspecific divergences were calculated with the K2P model in MEGA 7.0 (Kumar et al., 2016). To detect the presence of a barcoding gap, the minimum interspecific and maximum intraspecific distances were compared (Meyer & Paulay, 2005; Zhang et al., 2015). To assess the accuracy of each individual barcode and all possible combinations for species discrimination, the functions of the best match, best close match, and all species barcodes functions were intended in the program TaxonDNA were used under the K2P-corrected distance model (Meier et al., 2006). The "best match" tool was used to search for the closest barcode match for each query. The identification was deemed successful only if both sequences were from the same species, while mismatched names were interpreted as failures. If there were several equally valid "best matches" from different species, they

were considered ambiguous (Meier et al., 2006; Zhang et al., 2012).

Phylogenetic analyses using combinations of nrITS and the three cpDNA markers were performed with the maximum likelihood (ML) method. The optimal fitting model was determined using MODELTEST v.3.7 (Posada & Crandell, 1998) by applying the Akaike information criterion (AIC; Posada & Buckley, 2004). The ML analysis was performed with IQ TREE v.1.6.0 (Nguyen et al., 2015). The degree of species resolution (identification) for the four DNA barcode regions was evaluated using the ML method. Species discrimination was considered successful when all conspecific individuals formed a clade in the ML tree with a bootstrap value of more than 60% (Zhang et al., 2012).

NCBI BLAST 2.2.29+ (Tao, 2010) was used for all sequences and analyzed using the "BLASTn" (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) command to index local reference databases. Successful discrimination was deemed when the species had the highest hit matching only a conspecific individual; to improve clarity, the query sequence was removed from the list of top hits (Meyer & Paulay, 2005).

3 Results

3.1 Amplification and sequence success

The amplification success percentage for four loci (nrITS, *trnH-psbA*, *matK*, and *rbcL*) was 100%; the sequencing success rates ranged from 100% for nrITS and *trnH-psbA* to 85% for *rbcL* to 61% for *matK* (Table 1). The *de novo* generated sequences combined with sequences obtained from GenBank to a data set bringing the total to 449 sequences consisting of 146, 147, 60, and 96 belonging to nrITS, *trnH-psbA*, *matK*, and *rbcL* respectively.

3.2 Characteristic analysis of each barcode locus

The sizes of aligned sequences varied from 1392 bp for *rbcL* to 397 bp for *trnH-psbA* (Table 1). nrITS contained the highest number of parsimony-informative characters and variable sites were found in ITS, followed by *trnH-psbA*, *matK*, and *rbcL*. Specifically, the nrITS region had 758 bp with 5 bp long indels, 52.9% GC content, 129 bp parsimony-informative sites and 171 bp variable sites distributed equally across the matrix. The *trnH-psbA* sequence region was 397 bp with 25.2% GC content and 6 bp long indels, whereas the distribution of parsimony-informative sites (77) and variable sites (95) were dispersive and sparse across the matrix. The size of the aligned *matK* region was 825 bp with 50 sparse parsimony-informative sites and 82 variable sites across the matrix and there were 3 bp long indels with 31% GC content. The *rbcL* region alignment generated a length of 1392 bp with 43.3% GC content and contained 35 dispersive parsimony-informative sites and 60 sparse variable sites across the sequence matrix without indels (Table 1). Among the nine barcodes, namely the 04 single regions, 03 two regions, 01 three regions and 01 four region barcodes, the pairwise intraspecific distances varied between 0 and 1.84 (Table 1). The mean intraspecific distances were found to be highest for nrITS (0.31) and lowest for *matK* + *rbcL* (0.003). Subsequently, pairwise interspecific distances varied from 0 to 3.94 (Table 1), with the average interspecific distance highest for ITS (0.89) and lowest for *rbcL* (0.007; Table 1). In summary, nrITS revealed the highest mean intra- and interspecific distances, number of parsimony-informative sites and variable sites (Table 1).

3.3 DNA barcoding gap analysis and discrimination of species

The evaluation showed clear barcoding gaps for nrITS, *trnH-psbA*, and the combination of ITS + *trnH-psbA* (Fig. 2). All other barcodes revealed overlaps without clear barcoding gaps (Fig. 2). The "Best match" and "Best close match" revealed distinct performances among the nine barcodes considered. *rbcL* yielded poor results for the single region analysis (BM and BCM: 43.69%; Table 2). nrITS showed the highest success in species identification (BM and BCM: 92.27%), followed by *trnH-psbA* (BM and BCM: 79.12%) and *matK* (BM and BCM: 54.68%). The combination of two barcodes, ITS + *trnH-psbA* showed the highest discrimination success (BM and BCM: 95.39%). A slightly higher species identification was achieved by concatenation of all four regions (ITS + *trnH-psbA* + *matK* + *rbcL*) compared with the combination of *trnH-psbA* + *matK* + *rbcL* regions using the best match and best close match method (73.42%; Table 2).

3.4 Phylogenetic Tree and similarity-based method

These two approaches showed highly congruent results by comparing the effectiveness of the four barcodes and their combinations considered (Table 3; Figs. 3, 4, S1–S7). Both the multilocus and single barcodes displayed varying levels of species discrimination, ranging from 19.29% to 95.52% (Table 3). Among the four barcodes, the best discrimination power was shown by the nrITS region (91.04%), followed by *trnH-psbA* (65.67%), and *matK* (38.46%), while the *rbcL* gene demonstrated the lowest level of discrimination (19.29%). The barcode combining nrITS and *trnH-psbA* revealed the maximum discrimination success (95.52%) compared with other barcode loci combinations (Table 3).

4 Discussion

4.1 PCR and sequencing efficiency

Discussion on plant DNA barcodes, including those by the CBOL Plant Working Group, suggests that a working DNA barcode is required to fit into the following criteria: reliable amplification and DNA sequencing with low error rates along with high species discrimination rates (Kress et al., 2005; Kress & Erickson, 2007; CBOL Plant Working Group, 2009). Previous research on DNA barcodes considered mostly single-marker or multimarker barcodes based on four regions, primarily including the nuclear ITS region and the chloroplast regions *rbcL*, *matK*, and *trnH-psbA* (China Plant BOL Group, 2011; Hollingsworth et al., 2011). Primer universality has been widely considered one of the criteria for selecting an ideal barcode (Kress & Erickson, 2007). The four barcode regions explored in the present study are universal markers with expected product length within a range easily amplified and sequenced (Kress, 2017). The efficiency in gene amplification and rate of sequencing are indicators for evaluating a DNA barcode. In the present study, ITS and *trnH-psbA* demonstrated the best amplification and sequencing results among all four regions, consistent with previous studies (Xu et al., 2015; Yan et al., 2015).

Table 1 Sequence characteristics of four DNA candidate loci and their combinations tested in the present study (outgroup excluded)

Characters	ITS	*trnH-psbA*	*mat*K	*rbcL*	ITS + *trnH-psbA*	ITS + *matK*	ITS + *rbcL*	*trnH-psbA* + *matK* + *rbcL*	ITS + *trnH-psbA* + *matK* + *rbcL*
Universality of primers	Yes	Yes	Yes	Yes	–	–	–	–	–
No. of species (individuals)	67/146	67/147	39/60	57/96	–	–	–	–	–
Percentage PCR success	100	100	100	100	–	–	–	–	–
Percentage sequencing success	100	100	61	85	–	–	–	–	–
Aligned sequence length (bp)	742	397	816	1392	1339	1558	2134	2605	3347
No. of parsimony-informative sites	129	77	50	35	206	299	190	242	418
No. of variable sites	171	95	82	60	259	403	332	309	993
Indel length	5	6	3	0	0	16	14	9	23
No. of conserved sites	541	302	725	1332	849	1121	1778	2289	2374
GC content (%)	52.9	25.2	31	43.3	42.5	46.4	48.3	35.1	41.4
Average intraspecific distance (range) (%)	0.31 (0–1.84)	0.21 (0–0.89)	0.003 (0–0.03)	0.003 (0–0.01)	0.06 (0–1.56)	0.02 (0–0.57)	0.006 (0–0.05)	0.01 (0–0.14)	0.03 (0–0.61)
Average interspecific distance (range) (%)	0.89 (0–3.94)	0.73 (0.30–1.79)	0.01 (0–0.05)	0.007 (0–0.01)	0.39 (0–3.13)	0.13 (0–1.21)	0.02 (0–0.25)	0.03 (0–0.31)	0.04 (0–0.38)

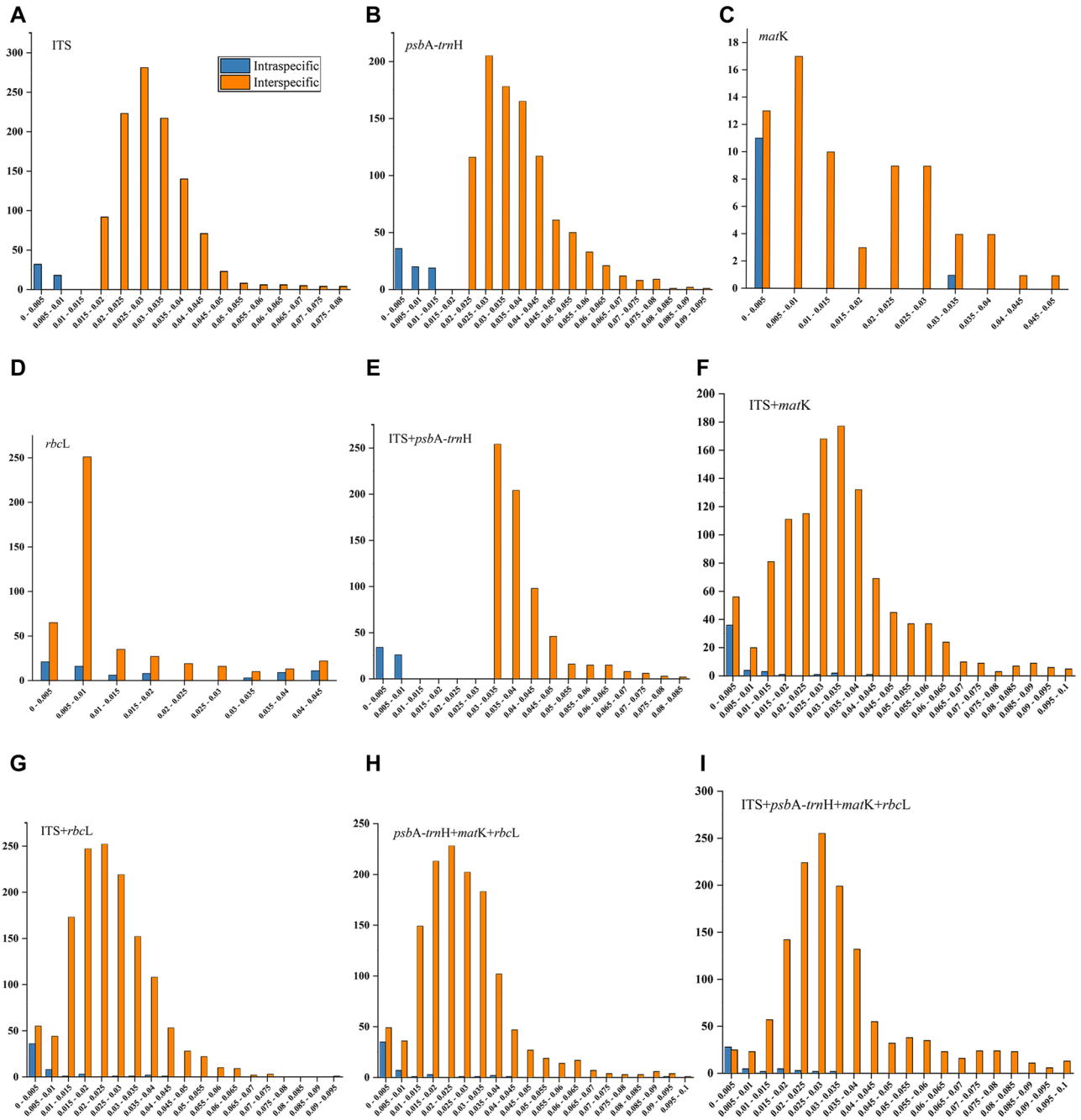


Fig. 2. Relative distribution of intraspecific and interspecific Kimura 2-parameter (K2P) distances among *Caragana* samples for the four candidate loci and their combinations. **A**, ITS. **B**, *trnH-psbA*. **C**, *matK*. **D**, *rbcL*. **E**, ITS + *trnH-psbA*. **F**, ITS + *matK*. **G**, ITS + *rbcL*. **H**, *trnH-psbA* + *matK* + *rbcL*. **I**, ITS + *trnH-psbA* + *matK* + *rbcL*.

4.2 The resolution of tested candidate barcodes in *Caragana*

The nuclear ITS has been considered a universal barcode for angiosperms (Li et al., 2011b; Xu et al., 2015; Yan et al., 2015). Few studies reject this proposal, since they have observed poor species resolution (Alvarez & Wendel, 2003; Chase et al., 2007; Hollingsworth et al., 2009; Starr et al., 2009). In contrast, other studies observed higher interspecific divergences and lower intraspecific variation in the ITS region compared with chloroplast markers (Chen et al., 2010; Yao et al., 2010; Li et al., 2011a; Zhu et al., 2017). In the present work, ITS showed the highest number of variation sites, parsimony-informative sites, accumulation of larger intra- and interspecific distances, most efficient amplification and high discriminatory power in comparison to the chloroplast markers (Table 1). Specifically, nrITS displayed the highest species-level resolution when employed individually or when combined with other barcoding regions (Figs. 3, 4, S4, S5, S7; Table 3). The ITS region was touted as a successful DNA barcode due to ITS rapid evolution rate, which was found to be, on average, three to four times faster than in chloroplast

Table 2 Identification success rates obtained using based the *best match*, *best close match*, and *all species barcodes* functions, as assessed in the program TaxonDNA

Region	Best match			Best close match			All species barcode			Threshold value
	Correct (%)	Ambiguous (%)	Incorrect (%)	Correct (%)	Ambiguous (%)	Incorrect (%)	Correct (%)	Ambiguous (%)	Incorrect (%)	
ITS	92.27	6.17	1.52	92.27	6.21	1.49	83.88	14.43	0.0	1.15
trnH-psbA	79.12	17.77	3.06	79.12	17.77	3.06	76.90	27.52	0.0	0.56
matK	54.68	34.37	10.93	54.68	34.37	10.93	37.37	59.95	1.21	1.42
rbcL	43.69	55.46	0.84	43.69	55.46	0.84	40.22	59.7	0.0	1.35
ITS + *matK*	74.29	12.61	3.87	74.29	4.92	9.97	67.35	32.29	0.0	1.21
ITS + *rbcL*	70.10	23.15	6.7	70.10	23.15	6.7	68.22	30.10	1.32	1.69
ITS + *trnH-psbA*	95.39	1.96	1.87	95.39	1.96	1.20	89.28	10.42	0.0	2.80
trnH-psbA + *matK* + *rbcL*	64.82	25.69	9.48	64.82	25.69	9.48	60.30	39.42	0.0	2.95
ITS + *trnH-psbA* + *matK* + *rbcL*	73.42	18.21	7.87	73.42	18.21	6.92	68.54	29.45	1.61	1.00

DNA regions (Kress et al., 2005; Chase et al., 2007; Liu et al., 2011; Zhang et al., 2012). In several taxonomy-based barcoding studies, this marker was utilized to identify taxa at the species rank, even in complex taxa, such as *Ficus* (Li et al., 2012), *Lysimachia* (Zhang et al., 2012), and *Viburnum* (Clement & Donoghue, 2012).

In the context of *Caragana*, ITS outperformed all four other barcoding regions considering the discrimination of species. Markedly, this region was also able to differentiate even taxonomically confusing species such as *C. opulens*, *C. licentiana*, and *C. kansuensis*. These results are consistent with the previous arguments considering ITS as in-disposable to the identification of plants used in traditional medicines due to the high level of variations and efficiency in identifying closely related species (CBOL Plant Working Group, 2009; Yao et al., 2010; Schoch et al., 2012; Yan et al., 2015; Liu et al., 2016; Wu et al., 2017; Ünsal et al., 2019; Gogoi et al., 2020). Because of low cost and high efficiency, ITS is recommended for the identification of *Caragana* species.

The chloroplast *trnH-psbA* spacer region displayed high variability. Previous studies demonstrated that *trnH-psbA* had a higher discriminating ability compared to *matK* + *rbcL* marker in angiosperm genera such as *Berberis* (Roy et al., 2010) and *Alnus* (Ren et al., 2010). The *trnH-psbA* region as a barcode is also realized to be a good candidate for large-scale DNA barcoding of some grasses and forage legumes. However, some issues affect its utilization for instance, extensive length variation in the *trnH-psbA* sequence resulted in alignment ambiguities. Moreover, inversions were found to occur in this region with the consequence of erroneous phylogenetic inferences due to sequence variation overestimation amongst closely related species and distantly placed taxa (Whitlock et al., 2010). Additionally, *psbA* mononucleotide (poly A/T) motifs in bidirectional reads have been identified to possess a negative effect on the ability to obtain full-sized sequences (Hollingsworth et al., 2009; Zhang et al., 2012). In the present case, the region provided satisfactory results, having a better discrimination power when used with ITS. Additionally, it has exhibited a high rate of amplification and sequencing success (Fig. 4; Table 3). This favorable outcome can be attributed to the application of standard DNA barcodes with standard barcoding PCR and cycle sequencing programs with modifications, which optimized the amplification and sequencing process. By employing these primers, we achieved the highest possible percentage of amplification and sequencing success, effectively mitigating any potential impact of *psb*A mononucleotide motifs. Our findings in this regard are consistent with recent studies conducted by Jiang et al. (2020) and Jin et al. (2023).

The *matK* and *rbcL* DNA barcodes employed in the study had previously been suggested as core regions for plant barcoding (Kress et al., 2005; CBOL Plant Working Group, 2009; Chen et al., 2010; Hollingsworth et al., 2011). However, they exhibited poor amplification, sequencing and species-level resolution in this study. Shifting the focus from universal toward *Caragana*-specific DNA barcodes, all plastid markers used in the present work were not recommended as DNA barcodes. The low resolving power of plastid markers at the species level had been reported for several

Table 3 Identification success rates obtained using maximum likelihood tree and local BLAST analysis methods for each candidate barcode and their combinations

S. No.	Single barcode locus and its combinations	ML tree method	Similarity-based method (BLAST)
1	ITS	91.04% (61/67)	91.04% (61/67)
2	*trnH-psbA*	65.67% (44/67)	65.67% (44/67)
3	*matK*	38.46% (15/39)	38.46% (15/39)
4	*rbcL*	19.29% (11/57)	19.29% (11/57)
5	ITS + *trnH-psbA*	95.52% (64/67)	95.52% (64/67)
7	ITS + *matK*	77.61% (52/67)	77.61% (52/67)
6	ITS + *rbcL*	68.65% (46/67)	68.65% (46/67)
8	*trnH-psbA* + *matK* + *rbcL*	53.73% (36/67)	53.73% (36/67)
9	ITS + *trnH-psbA* + *matK* + *rbcL*	74.62% (50/67)	74.62% (50/67)

angiosperm genera, such as *Berberis* (15.4%–23.1%), *Paphiopedilum* (25.74%), *Lysimachia* (47.1%–60.82%), *Viburnum* (53%), *Curcuma* (21.66%), and *Quercus* (0%) (Roy et al., 2010; Piredda et al., 2011; Clement & Donoghue, 2012; Zhang et al., 2012; Kim et al., 2014; Chen et al., 2015; Guo et al., 2015; Li et al., 2016; Liu et al., 2017). In our study, species identification efficiency was less with the *rbcL* and *matK* regions when compared with other examined regions (Figs. S2, S3; Tables 2, 3). However, levels of species resolution were higher in *matK* when compared with *rbcL*, which is consistent with the previous studies (Kress et al., 2009). Conversely, in some recent works, *rbcL* and *matK* showed low interspecific variations even in closely related species and, therefore, treated as an incongruous barcode marker (Kress et al., 2005; Steven & Subramanyam, 2009; Parveen et al., 2012, 2017; Singh et al., 2012; Kim et al., 2014; Rajaram et al., 2019; Worthy et al., 2022). In the present study too, these regions had a low number of variation sites, parsimony-informative sites, and species discrimination power, however, the discrimination rate slightly increased when both *matK* and *rbcL* were combined with ITS (Figs. S4, S5; Tables 1, 3). Overall, our results confirmed the expectation that *rbcL* and *matK* were not variable enough to assess interspecific variations in later derived angiosperm genera such as *Caragana* (Lahaye et al., 2008; Parveen et al., 2012; Singh et al., 2012).

An individual DNA barcode region lacks genetic variation to be a useful DNA barcode. Thus, combinations of multiple loci are deemed necessary to improve the species resolution ability. The concatenated barcodes have been known to exhibit higher species discrimination than single-locus barcodes (CBOL Plant Working Group, 2009; Li et al., 2011a, 2015; Yan et al., 2015). In this study, ITS + *trnH-psbA* showed the best discriminatory performance among the four tested loci combinations (Fig. 4; Tables 2, 3). The species identification rates utilizing the three cpDNA fragments were significantly lower (Fig. S6; Tables 2, 3). In conclusion, the study revealed that a combination of ITS + *trnH-psbA* DNA regions might be used as an effective plant DNA barcode.

4.3 DNA barcoding implications for the current taxonomy of *Caragana*

Few closely related species in *Caragana* are known to have overlapping morphological traits, which posed difficulty in their identification using the alpha taxonomic approach. One such case is the *C. opulens* species complex, which consists of three species, namely *C. opulens*, *C. kansuensis*, and *C. licentiana* (Rather et al., 2021). Although, diagnostic characteristics have been found to be inconsistent and labile, the three species have been differentiated by ovary/fruit pubescence, leaf shape, bract shape, and leaf pubescence (Moore et al., 2010). Further, the herbarium specimens of these species housed in major herbaria in China have been differently identified (Zhao, 1993). Extensive investigation of literature and comparison of specimens have revealed that these species have overlapping geographical distribution. Further, our barcoding and phylogenetic analysis revealed that these three species belong to a single and well-supported clade (Figs. 3, 4). Based on these pieces of two evidence, we hypothesize that these species are likely to be morphotypes of a single species, and need to be merged into a single taxon, *C. opulens*.

Misidentification, problems in cryptic species and discovery of new taxa have often been evaluated through barcoding (Burns et al., 2008; Liu et al., 2011; Saitoh et al., 2015). In our study, *C. opulens* Boufford 40 782 was probably misidentified by earlier researchers, not withholding geographical divergences. After extensive examination of the specimens, it has been found that it is *Caragana sinica* which is also supported by our study. It does not form a group with *C. opulens*, rather it is a part of the *C. sinica* clade in the phylogenies (Figs. 3, 4, S1–S7).

For any taxa, its robust taxonomic reconstruction relies on ecological, reproductive, molecular and morphological characterization along with geographical data. Nevertheless, delimitation is often difficult in genera having closely related species. In our study, despite the high success rate in *Caragana* spp., DNA barcoding had difficulty discriminating closely related species. For example, the three species (*C. sukiensis*, *C. conferta*, *and C. brevispina*) of the *C. sukiensis* species complex are morphologically similar, showing minute variations in diagnostic floral traits. The barcoding markers failed to differentiate the species in this complex either in the single or combined marker-based tree (Figs. 3, 4, S1–S7). Possible reasons for barcoding failure for closely related species are commonly associated with a slow molecular evolution rate, coupled with hybridization and introgression events, paralogy and incomplete sorting of ancestral polymorphisms (Funk & Omland, 2003; Hollingsworth et al., 2011;

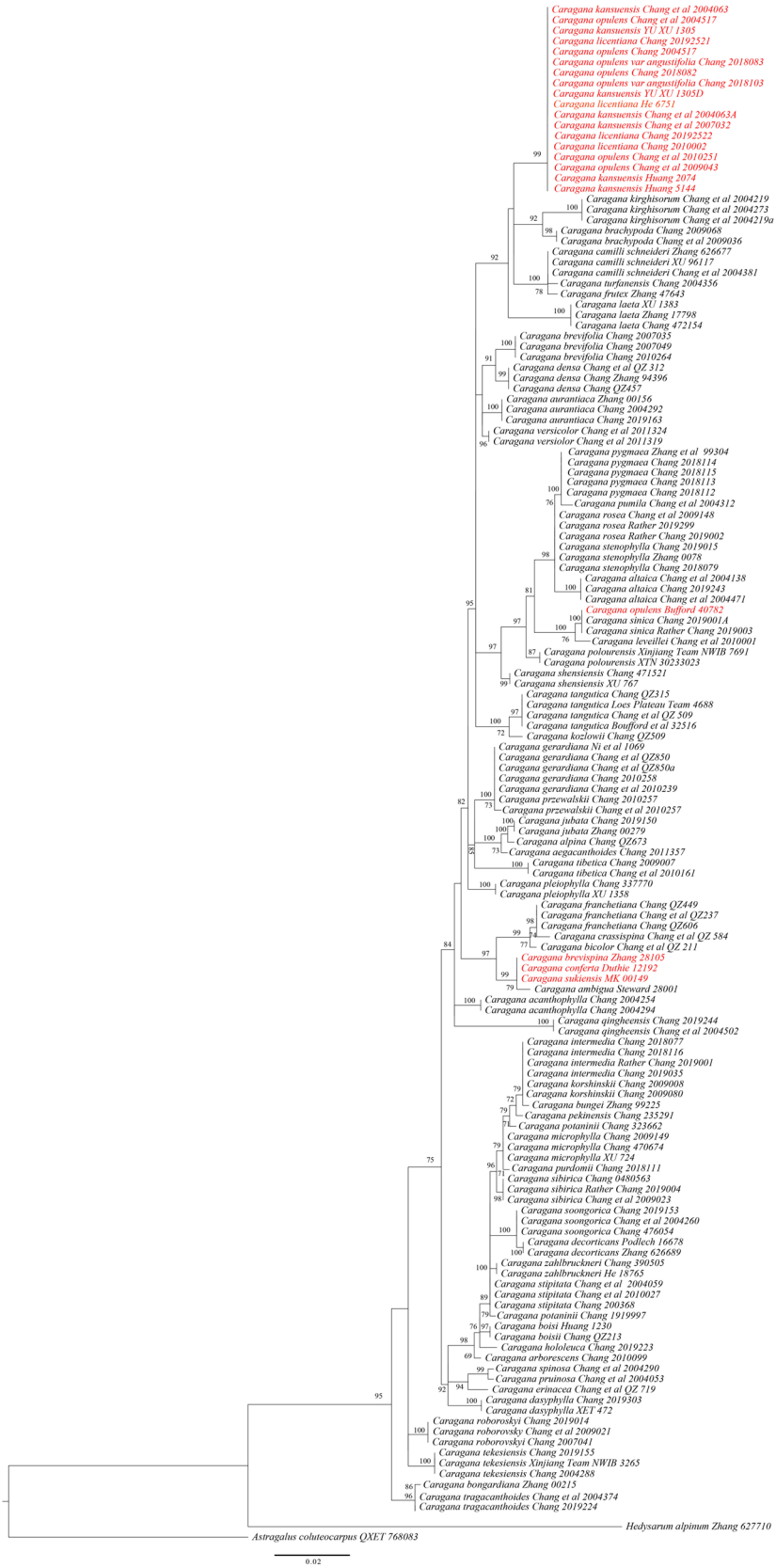
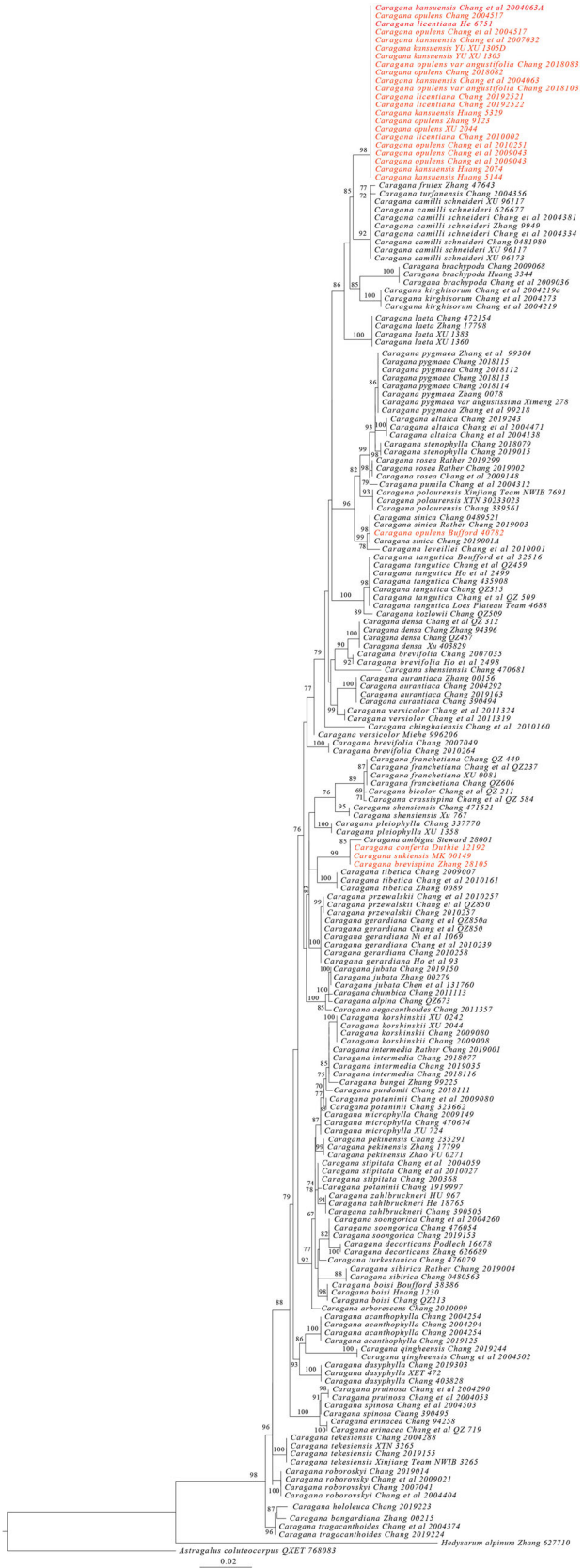


Fig. 3. Maximum likelihood phylogeny based on the ITS sequences in *Caragana*. Numbers on branches represent bootstrap values. Successfully identified species have bootstrap values of 60% or higher. Samples shown in red indicate misidentified and unresolved samples.



León-Romero et al., 2012). Additional molecular exploration, such as through single nucleotide polymorphism (SNP) and Satellite markers (SSR) may be required to develop DNA barcodes for assisting species identification in such cases (Liu et al., 2008; Zeng et al., 2012). The data generated provides useful insights into the DNA barcoding progress within the genus and facilitates the identification of economically important species for proper utilization and conservation purposes.

4.4 Conclusions

The present study reports a comparative analysis focusing on four potential barcoding DNA regions with the aim to identify accession of the legume genus *Caragana*. The nuclear ITS region is found to be the most accurate and efficient barcode for distinguishing *Caragana* accessions/species. This marker has the highest discriminatory rate and can differentiate between *Caragana* species when employed individually or in combinations with *trnH-psbA*. The previously recommended universal plant barcode *rbcL* + *matK* marker are ineffective as identifiers for *Caragana* species. By diverging from universal toward genera-specific barcodes, we recommend the ITS alone or the concatenation of two regions (*trnH-psbA* + ITS) as plant barcodes in *Caragana*.

Acknowledgements

We acknowledge the financial support provided by the National Natural Science Foundation of China and the Yunnan Postdoctoral Research Project funded by the Yunnan Science and Technology Department awarded to SAR (Grant Nos. 32250410305 and Y8BSH11008, respectively) and the Science and Technology Basic Work (Grant No. 2013FY112100). We are grateful to the curators of various herbaria (WUH, PE, KUN, IBSC, HITBC, PEY, and XJBI) for allowing us to consult the herbarium specimens.

Author Contributions

Shabir A. Rather designed the study. Chang Zhao-Yang and Shabir A. Rather collected the samples. Shabir A. Rather conducted laboratory experiments. Shabir A. Rather conducted bioinformatic and statistical analyses. Shabir A. Rather prepared the original draft. Shabir A. Rather, Arjun Adit, Hongmei Liu, Josphat K. Saina, Chang Zhao-Yang and Harald Schneider reviewed and edited the draft. All authors have approved the final version of the manuscript.

Fig. 4. Maximum likelihood phylogeny based on the ITS + *trnH-psbA* sequences in *Caragana*. Numbers on branches represent bootstrap values. Successfully identified species have bootstrap values of 60% or higher. Samples shown in red indicate misidentified and unresolved samples.

Conflict of Interest Statement

Authors declare no competing interests.

Consent to Participate

All authors approved their inclusion in the manuscript.

Consent to Publication

All authors gave their consent for publishing this work.

Code Availability

Not applicable.

Ethical Approval

Not applicable.

Data Availability Statement

All data generated during this study are included in this published article and its supplementary information files.

References

Alvarez I, Wendel JF. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics Evolution* 29: 417–434.

Amritha N, Bhooma V, Parani M. 2020. Authentication of the market samples of Ashwagandha by DNA barcoding reveals that powders are significantly more adulterated than roots. *Journal of Ethnopharmacology* 256: 112725. https://doi.org/10.1016/j.jep.2020.112725

Bhardwaj PK, Kapoor R, Mala D, Bhagwat G, Acharya V, Singh AK, Vats SK, Ahuja PS, Kumar S. 2013. Braving the attitude of altitude: *Caragana jubata* at work in cold desert of Himalaya. *Scientific Reports* 3: 1–9.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN. 2008. DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservación Guanacaste, Costa Rica. *Proceeding of National Academy of Sciences USA* 105: 6350–6355.

Boatwright JS, Le Roux MM, Wink M, Morozova T, Van Wyk BE. 2008. Phylogenetic relationships of tribe Crotalarieae (Fabaceae) inferred from DNA sequences and morphology. *Systematic Botany* 33: 752–761.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceeding of National Academy of Sciences USA* 106: 12794–12797.

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, Seberg O, Jørgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M. 2007. A proposal for a standardized protocol to barcode all land plants. *Taxon* 56: 295–299.

Chen J, Zhao J, Erickson DL, Xia N, Kress WJ. 2015. Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China. *Molecular Ecology Resources* 15: 337–348.

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5: e8613. https://doi.org/10.1371/journal.pone.0008613

China Plant BOL Group. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceeding of National Academy of Sciences USA* 108: 19641–19646.

Clement WL, Donoghue MJ. 2012. Barcoding success as a function of phylogenetic relatedness in *Viburnum*, a clade of woody angiosperms. *BMC Evolutionary Biology* 12: 73. https://doi.org/10.1186/1471-2148-12-73

Cui N, Liao BS, Liang CL, Li SF, Zhang H, Xu J, Li XW, Chen SL. 2020. Complete chloroplast genome of *Salvia plebeia*: Organization, specific barcode, and phylogenetic analysis. *Chinese Journal of Natural Medicines* 18: 563–572.

Dai W, Dong P, Tian S, Mei Q. 2018. A pharmacognostical study on *Ardisia gigantifolia* and ITS adulterants. *Medicinal Plants* 9: 39–44.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Duan L, Jun W, Yang X, Pei-Liang L, Arslan E, Ertuðrul K, Chang ZY. 2015. Phylogeny of *Hedysarum* and tribe Hedysareae (Leguminosae: Papilionoideae) inferred from sequence data of ITS, *matK*, *trnL-F* and *trnH-psbA*. *Taxon* 64: 49–64.

Duan L, Yang X, Peiliang L, Johnson G, Jun W, Chang ZY. 2016. A molecular phylogeny of Caraganeae (Leguminosae, Papilionoideae) reveals insights into new generic and infrageneric delimitations. *Phytokeys* 70: 111–137.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics* 34: 397–423.

Gogoi B, Wann SB, Saikia SP. 2020. DNA barcodes for delineating *Clerodendrum* species of North East India. *Scientific Reports* 10: 13490.

Guo YY, Huang LQ, Liu ZJ, Wang XQ. 2015. Promise and challenge of DNA barcoding invenus slipper (*Paphiopedilum*). *PLoS One* 11: e0146880. https://doi.org/10.1371/journal.pone.0146880

Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan R, Chase MW, Gaudeul M, Hollingsworth PM. 2009. Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources* 9: 439–457.

Hollingsworth PM, Graham SW, Little DP. 2011. Choosing and using a plant DNA barcode. *PLoS One* 6: e19254. https://doi.org/10.1371/journal.pone.0019254

Jia XY, Qian ZG, Zhang TX, Li BJ, Zhu CR. 1997. Study on species diversity of medicinal plants, genus *Caragana* Fabr in Yunnan. *Journal of Yunnan College of Traditional Chinese Medicine* 20: 8–12.

Jiang KW, Zhang R, Zhang ZF, Pan B, Tian B. 2020. DNA barcoding and molecular phylogeny of *Dumasia* (Fabaceae: Phaseoleae) reveals a cryptic lineage. *Plant Diversity* 42: 376–385.

Jin DP, Sim S, Park JW, Choi JE, Yoon J, Lim CE, Kim MH. 2023. Identification of the Plant Family Caryophyllaceae in Korea Using DNA Barcoding. *Plants* 12: 2060.

Kim HM, Oh SH, Bhandari GS, Kim CS, Park CW. 2014. DNA barcoding of Orchidaceae in Korea. *Molecular Ecology Resources* 14: 499–507.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111–120.

Kress WJ. 2017. Plant DNA barcodes: Applications today and in the future. *Journal of Systematics and Evolution* 55: 291–307.

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants, the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One* 2: e508. https://doi.org/10.1111/jse.12254

Kress WJ, Erickson DL, Jones FA, Swensond NG, Perez R, Sanjur O, Bermingham E. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceeding of National Academy of Sciences USA* 106: 18621–18626.

Kress WJ, Wurdac KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. *Proceeding of National Academy of Sciences USA* 102: 8369–8374.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870–1874.

Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. 2008. DNA barcoding the floras of biodiversity hotspots. *Proceeding of National Academy of Sciences USA* 105: 2923–2928.

León-Romero Y, Mejía O, Soto-Galera E. 2012. DNA barcoding reveals taxonomic conflicts in the *Herichthys bartoni* species group (Pisces: Cichlidae). *Molecular Ecology Resources* 12: 1021–1026.

Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW. 2011b. Comparative analysis of a large dataset indicates that an internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceeding of National Academy of Sciences USA* 108: 19641–19646.

Li DZ, Liu JQ, Chen ZD, Wang H, Ge XJ, Zhou SL, Gao LM, Fu CX, Chen SL. 2011a. Plant DNA barcoding in China. *Journal of Systematics and Evolution* 49: 165–168.

Li HQ, Chen JY, Wang S, Xiong SZ. 2012. Evaluation of six candidate DNA barcoding loci in *Ficus* (Moraceae) of China. *Molecular Ecology Resources* 12: 783–790.

Li XW, Yang Y, Henry RJ, Rossetto M, Wang YT, Chen SL. 2015. Plant DNA barcoding: From gene to genome. *Biological Review* 90: 157–166.

Li YL, Tong Y, Xing FW. 2016. DNA barcoding evaluation and ITS taxonomic implications in the recently evolved genus *Oberonia* Lindl. (Orchidaceae) in China. *Frontiers in Plant Science* 7: 1791.

Liu J, Möller M, Gao LM, Zhang DQ, Li DZ. 2011. DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species. *Molecular Ecology Resources* 11: 89–100.

Liu J, Yan HF, Ge XJ. 2016. The use of DNA barcoding on recently diverged species in the genus *Gentiana* (Gentianaceae) in China. *PLoS One* 11: e0153008. https://doi.org/10.1371/journal.pone.0153008

Liu YX, Chang ZY, Yakovlev GP. 2010. Caragana. In: Wu ZY, Hong DY, Raven PH eds. *Flora of China*. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 10: 528–545.

Liu Z, Meng J, Sun X. 2008. A novel feature-based method for whole-genome phylogenetic analysis without alignment: Application to HEV genotyping and subtyping. *Biochemical and Biophysical Research Communication* 368: 223–230.

Liu ZF, Ci XQ, Li L, Li HW, Conran JG, Li J. 2017. DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China. *PLoS One* 12: e0175788. https://doi.org/10.1371/journal.pone.0175788

Lock JM. 2005. Tribe Hedysareae. In: Lewis G, Schrire B, MacKinder B, Lock M eds. *Legumes of the world*. London: Kew Publishing. 489–495.

Luo K, Pei M, Hui Y, Jingyuan S, Keli C, Yimei L. 2012. Molecular identification of Fritillariae Cirrhosae Bulbus and ITS adulterants. *World Science and Technology* 1: 1153–1158.

Manandhar NP. 2002. *Plants and people of Nepal*. Oregon: Timber Press.

Meier R, Shiyang K, Vaidya G, Ng PK. 2006. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology* 55: 715–728.

Meng Q, Niu Y, Niu X, Roubin RH, Hanrahan JR. 2009. Ethnobotany, phytochemistry, and pharmacology of the genus *Caragana* are used in traditional Chinese medicine. *Journal of Ethnopharmacology* 124: 350–368.

Meyer CP, Paulay G. 2005. DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology* 3: e422. https://doi.org/10.1371/journal.pbio.0030422

Moore MJ, Soltis PS, Bell CD, et al. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceeding of National Academy of Sciences USA* 107: 4623–4628.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Niu XW. 1988. *Cultivation and utilization of peashrubs*. Taiyuan: Shanxi Science and Educational Press.

Pan B, Zhu XY. 2010. Taxonomic revision of *Dumasia* (Fabaceae, Papilionoideae). *Annales Botanici Fennici* 47: 241–256.

Parveen I, Singh HK, Raghuvanshi S, Pradhan UC, Babbar SB. 2012. DNA barcoding of endangered Indian *Paphiopedilum* species. *Molecular Ecology Resources* 12: 82–90.

Parveen I, Singh HK, Malik S, Raghuvanshi S, Babbar SB. 2017. Evaluating five different loci (*rbcL*, *rpoB*, *rpoC1*, *matK*, and ITS) for DNA barcoding of Indian orchids. *Genome* 60: 665–671.

Piredda R, Simeone MC, Attimonelli M, Bellarosa R, Schirone B. 2011. Prospects of barcoding the Italian wild dendro flora: Oaks reveal severe limitations to tracking species identity. *Molecular Ecology Resources* 11: 72–83.

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53: 793–808.

Posada D, Crandell KA. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformation* 14: 817–818.

Rajaram C, Yong C, Gansau JA, Go R. 2019. DNA barcoding of endangered *Paphiopedilum* species (Orchidaceae) of Peninsular Malaysia. *Phytotaxa* 387: 94–104.

Rambaut A. 2007. Se-al version2. 0a11 [online]. Available from http://tree.bio.ed.ac.uk/software/seal/ [Accessed 15 September 2022]

Rather SA, Wang S, Dwivedi MD, Zhaoyang C. 2021. Molecular phylogeny and systematic evaluation of the *Caragana opulens* species complex (Fabaceae, Papilionoideae) based on the molecular and morphological data. *Phytotaxa* 478: 179–200.

Ren BQ, Xiang XG, Chen ZD. 2010. Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers. *Molcular Ecology Resources* 10: 594–605.

Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, Datt B, Bag SK, Singh PK, Nair NK, Husain T. 2010. Universal plant DNA barcode loci may not work in complex groups: A case study with Indian *Berberis* species. *PLoS One* 5: e13674.

Saitoh T, Sugita N, Someya S, Iwami Y, Kobayashi S, Kamigaichi H, Higuchi A, Asai S, Yamamoto Y, Nishiumi. 2015. DNA barcoding reveals 24 distinct lineages as cryptic bird species candidates in and around the Japanese archipelago. *Molecular Ecology Resources* 15: 177–186.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceeding of National Academy of Sciences USA* 109: 6241–6246.

Singh HM, Parveen I, Raghuvanshi S, Babbar SB. 2012. The loci recommended as universal barcodes for plants on the basis of floristic studies may not work with congeneric species as exemplified by DNA barcoding of *Dendrobium* species. *BMC Research Notes* 5: 42. https://doi.org/10.1186/1756-0500-5-42

Samarakoon T, Wang SY, Alford MH. 2013. Enhancing PCR amplification of DNA from recalcitrant plant specimens using a trehalose-based additive. *Applications in Plant Sciences* 1(1): 1200236.

Skjua B, Mr C, Gss D, Vk A, Rus B, Gr E. 2020. DNA barcoding of *Momordica* species and assessment of adulteration in *Momordica* herbal products, an anti-diabetic drug. *Plant Gene* 22: 100227. https://doi.org/10.1016/j.plgene.2020.100227

Starr JR, Naczi RFC, Chouinard BN. 2009. Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources* 9: 151–163.

Steven GN, Subramanyam R. 2009. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Molecular Ecology Resources* 9: 172–180.

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E. 2007. Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* 35: e14. https://doi.org/10.1093/nar/gkl938

Tanaka T. 1976. *Tanaka's cyclopaedia of edible plants of the world*. Tokyo: Keigaku Publishing.

Tao T. 2010. Standalone BLAST setup for Windows PC. http://www.ncbi.nlm.nih.gov/books/NBK52637/

Ünsal SG, Çiftçi YÖ, Eken BU, Velioğlu E, Marco GD, Gismondi A, Canini A. 2019. Intraspecific discrimination study of wild cherry populations from North-Western Turkey by DNA barcoding approach. *Tree Genetics and Genomes* 15: 16.

Wang B, Zhao X, Liu Y, Fang Y, Ma R, Yu Q, An S. 2019. Using soil aggregate stability and erodibility to evaluate the sustainability of large-scale afforestation of *Robinia pseudoacacia* and *Caragana korshinskii* in the Loess Plateau. *Forest Ecology and Management* 450: 117491. https://doi.org/10.1016/j.foreco.2019.117491

Wang SG, Da Y, Hu CQ. 2004. Two new oligostilbenes from *Caragana sinica*. *Journal of Asian Natural Products Research* 6: 241–248.

Wang SG, Ma D, Hu CQ. 2005. Three new compounds from the aerial parts of *Caragana sinica*. *Helvetica Chimica Acta* 88: 2315–2321.

Whitlock BA, Hale AM, Groff PA, Joly S. 2010. Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. *PLoS One* 5: e11533. https://doi.org/10.1371/journal.pone.0011533

Worthy SJ, Bucalo K, Perry E, Reynolds A, Cruse-Sanders J, Pérez ÁJ, Burgess KS. 2022. Ability of *rbc*L and *mat*K DNA barcodes to discriminate between montane forest orchids. *Plant Systematics and Evolution* 308: 19. https://doi.org/10.1007/s00606-022-01809-z

Wu F, Ma JX, Meng YQ, Zhang DY, Muvunyi BP, Luo K, Di HY, Guo WL, Wang YR, Feng BC, Zhang JY. 2017. Potential DNA barcodes for *Melilotus* species based on five single loci and their combinations. *PLoS One* 12: e0182693. https://doi.org/10.1371/journal.pone.0182693

Xiang XG, Hu H, Wang W, Jin XH. 2011. DNA barcoding of the recently evolved genus *Holcoglossum* (Orchidaceae: Aeridinae): A test of DNA barcode candidates. *Molecular Ecology Resources* 11: 1012–1021.

Xu S, Li D, Li JW, Xiang X, Jin W, Huang W, Jin X, Huang L. 2015. Evaluation of the DNA barcodes in *Dendrobium* (Orchidaceae) from mainland Asia. *PLoS One* 10: e0115168. https://doi.org/10.1371/journal.pone.0115168

Yan HF, Hao G, Hu CM, Ge XJ. 2011. DNA barcoding in closely related species: A case study of *Primula* L. sect. Proliferae Pax (Primulaceae) in China. *Journal of Systematics and Evolution* 49: 225–236.

Yan HF, Liu YJ, Xie XF, Zhang CY, Hu CM, Hao G, Ge XJ. 2015. DNA barcoding evaluation and ITS taxonomic implications in the species-rich genus *Primula* L. in China. *PLoS One* 10: e0122903. https://doi.org/10.1371/journal.pone.0122903

Yang J, Feng L, Yue M, He YL, Zhao GF, Li ZH. 2019. Species delimitation and interspecific relationships of the endangered herb genus *Notopterygium* inferred from multilocus variations. *Molecular Phylogenetics and Evolution* 133: 142–151.

Yao H, Song JY, Liu C, Luo K, Han JP, Li Y, Pang XH, Xu HX, Zhu YJ, Xiao PG, Chen SL. 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One* 5: e13102. https://doi.org/10.1371/journal.pone.0013102

Yin X, Qian C, Yan X, Fang T, Fan X, Zhou S, Ma XF. 2022. Will the artificial populations be sustainable? A genetic assessment on *Caragana korshinskii* afforestation in the semiarid regions of North China. *European Journal of Forest Research* 141: 105–116.

Zeng X, Yuan Z, Tong X, Li Q, Gao W, Qin M, Liu Z. 2012. Phylogenetic study of Oryzoideae species and related taxa of the Poaceae based on *atpB-rbcL* and *ndhF* DNA sequences. *Molecular Biology Reports* 39: 5737–5744.

Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, Ge XJ. 2012. Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). *Molecular Ecology Resources* 12: 98–108.

Zhang D, Jiang B. 2020. Species identification in complex groups of medicinal plants based on DNA barcoding: A case study on *Astragalus* spp. (Fabaceae) from Southwest China. *Conservation Genetics Resources* 12: 469–478.

Zhang J, Chen M, Dong X, Lin R, Fan J, Chen Z. 2015. Evaluation of four commonly used DNA barcoding loci for Chinese medicinal plants of the family Schisandraceae. *PLoS One* 10: e0125574. https://doi.org/10.1371/journal.pone.0125574

Zhang ML. 1997. A reconstructing phylogeny in *Caragana* (Fabaceae). *Acta Botanica Yunnanica* 19: 331–341.

Zhang ML, Tian XY, Ning JC. 1996. Pollen morphology and its taxonomic significance of *Caragana* Fabr. (Fabaceae) from China. *Acta Phytotaxa Sinica* 34: 397–409.

Zhang M, Fritsch PW, Cruz BC. 2009. Phylogeny of *Caragana* (Fabaceae) based on DNA sequence data from *rbcL*, *trnS–trnG*, and ITS. *Molecular Phylogenetics and Evolution* 50: 547–559.

Zhang ML, Huang YM, Kang Y, Wang YW. 2002. Biodiversity and biogeography of legumes in Ordos Plateau, China. *Bulletin of Botanical Research* 22: 497–502.

Zhang ML, Xiang XG, Xue JJ, Sanderson SC, Fritsch PW. 2016. Himalayan uplift shaped biomes in Miocene temperate Asia: Evidence from leguminous *Caragana*. *Scientific Reports* 6: 1–7.

Zhao YZ. 1993. Taxonomic study of the genus *Caragana* from China. *Acta scientiarum naturalium Universitatis Intericris Mongolicae* 24: 631–653.

Zhou QX, Yang YP, Zhang ML. 2002. Karyotypes of fourteen species in *Caragana*. *Bulletin of Botanical Research* 22: 492–496.

Zhu RW, Li YC, Zhong DL, Zhang JQ. 2017. Establishment of the most comprehensive ITS2 barcode database to date of the traditional medicinal plant *Rhodiola* (Crassulaceae). *Scientific Reports* 7: 10051.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.13009/suppinfo:

Fig. S1. Maximum likelihood tree based on the *trnH-psbA* sequences in *Caragana*. Numbers on branches represent bootstrap values. Successfully identified species have bootstrap values of 60% or higher.

Fig. S2. Maximum likelihood tree based on the *matK* sequences in *Caragana*. Numbers on branches represent bootstrap values. Successfully identified species have bootstrap values of 60% or higher.

Fig. S3. Maximum likelihood tree based on the *rbcL* sequences in *Caragana*. Numbers on branches represent bootstrap values. Successfully identified species have bootstrap values of 60% or higher.

Fig. S4. Maximum likelihood tree based on the ITS + *matK* sequences in *Caragana*. Numbers on branches represent bootstrap values. Successfully identified species have bootstrap values of 60% or higher.

Fig. S5. Maximum likelihood tree based on the ITS + *rbcL* sequences in *Caragana*. Numbers on branches represent bootstrap values. Successfully identified species have bootstrap values of 60% or higher.

Fig. S6. Maximum likelihood tree based on the *trnH-psbA* + *matK* + *rbcL* sequences in *Caragana*. Numbers on branches represent bootstrap values. Successfully identified species have bootstrap values of 60% or higher.

Fig. S7. Maximum likelihood tree based on the ITS + *trnH-psbA* + *matK* + *rbcL* sequences in *Caragana*. Numbers on branches represent bootstrap values. Successfully identified species have bootstrap values of 60% or higher.

Table S1. Plant material used for molecular analysis of the genus *Caragana*. Original information with voucher information and GenBank accession numbers for all the samples of *Caragana* used in this study. All specimens have been deposited in the Northwest A & F University Herbarium (WUK).

Table S2. Details of primers used for amplification and subsequent sequencing in the present study.