UDC 575.630
https://doi.org/10.2298/ GENSR2401195T
Original scientific article

DNA BARCODING AND MORPHOLOGICAL IDENTIFICATION OF *Astragalus* IN IRAN: ADVANTAGES TO MOLECULAR APPROACH

Abolfazl TAHMASEBI[1] and Bakhtiar FATTAHI[2]

[1]Faculty of Agriculture and Natural Resources, Gonbad Kavous University, Gonbad, Iran
[2]Faculty of Natural Resources and Environment Engineering, Malayer University, Malayer, Iran

Tahmasebi A. and B. Fattahi (2024). *DNA barcoding and morphological identification of Astragalus in Iran: Advantages to molecular approach*.- Genetika, Vol 56, No.1, 195-211.

Astragalus is a crucial forage plant in the rangelands of northeastern Iran. However, identifying different species of *Astragalus* can be challenging due to their overlapping morphological features. In this study, we aimed to determine the relationships between 12 *Astragalus* species using both morphological and molecular traits. We also evaluated the effectiveness of DNA barcoding as a tool for identifying rangeland species of *Astragalus* in Northeast Iran. We examined five DNA barcodes, including three cpDNA regions (*trn*H-*psb*A, *rpl*32-*trn*L$_{(UAG)}$, and *mat*K) and two nuclear sequences (ITS and ETS). We found that the plastid markers (*rpl*32-*trn*L$_{(UAG}$, *mat*K) were the most effective in differentiating between species. We also found that morphological data, as represented by the Neighbor-Net network and UPGMA dendrogram, had the potential to separate *Astragalus* species. We identified *mat*K as the best and most accurate marker for barcoding, as *trn*H-*psb*A had some defects due to sequence size and alignment issues. Our study highlights the importance of using barcoding for quick and accurate recognition of plant species and shows the plastid markers were the most effective in differentiating between species.

Keywords: barcode, phylogeny, rangeland, species delimitation

INTRODUCTION

The complication within *Astragalus* is caused by species richness and high phenotypic plasticity. These challenges make it difficult to distinguish the border between species. In such

Corresponding author: Abolfazl Tahmasebi, Faculty of Agriculture and Natural Resources, Gonbad Kavous University, Gonbad, Iran. E-mail address: **ab_tahmasebi@gonbad.ac.ir**,
ORCID: 0000-0002-4477-1136, Z Bakhtiar Fattahi ORCID: 0000-0001-6187-2601

cases, species identification using morphological traits alone may be impossible (MINAEIFAR *et al*., 2016).

A new method for identifying plants based on the analysis of DNA barcode sequences has been suggested (HEBERT *et al*., 2003). DNA barcode sequence variation must be high enough between species to enable their differentiation. In practice, a DNA sequence from an unknown organism can be generated from standardized DNA regions and compared to a library of sequences sourced from described species. By comparing the sequence of the unknown organism to one of the reference sequences, a rapid and precise recognition can be achieved (CASIRAGHI *et al*., 2010; HEBERT *et al*., 2003).

DNA barcoding was presented as a novel system to enable wide-scale and quick species identification using certain gene sequences as molecular species-specific tags (HEBERT *et al*., 2003). Since then, the number of sequenced species has increased significantly, with DNA barcodes for almost 200,000 identified species available in international databases (RATNASINGHAM and HEBERT, 2007). DNA barcoding has enabled the identification of species at life stages that are difficult to distinguish morphologically, such as insect larvae (AHRENS *et al*., 2013; BONAL *et al*., 2011). This method has increased biodiversity and environmental monitoring and is a valuable tool in taxonomy, ecology, agriculture, and conservation, as well as in customs, police and feed control (BERGSTEN *et al*., 2012; JINBO *et al*., 2011; SAVOLAINEN *et al*., 2005). DNA barcoding is also useful for ensuring the safety of natural plant and animal products used in traditional medicine, revealing cryptic diversity, performing phylogenetic analysis, and ensuring food safety (PENTON *et al*., 2004; CHEN *et al*., 2010; NITHANIYAL *et al*., 2021).

The DNA barcoding initiative has been a significant historical invention, but this success does not imply that the method is without flaws (BERGSTEN *et al*., 2012; BERTHIER *et al*., 2011; DUBEY *et al*., 2009; NICHOLLS *et al*., 2012). One of these limitations is the decrease in identification accuracy as intraspecies genetic divergence increases (MEYER and PAULAY, 2005; BERGSTEN *et al*., 2012). As next-generation sequencing (NGS) technology advances, DNA barcoding is evolving into metabarcoding to overcome the limitations of single-species sequencing by enabling the simultaneous sequencing of multiple species in a single PCR (PIPER *et al*., 2019; GAO *et al*., 2019).

However, determining the most suitable regions for plant DNA barcoding, both in terms of identity and number, remains a challenging issue (HOLLINGSWORTH *et al*., 2009a, HOLLINGSWORTH *et al*., 2009b).

DNA barcoding and phylogenetic investigations of endemic species are vital and useful tools for the quick and accurate identification of species, and this method can also help discover new species. Molecular analyses have been used to study historical evolutionary research and species divergence (SHEIDAI *et al*., 2014). Additionally, molecular information can provide additional insights for the systematic classification of species based solely on morphological traits (MINAEIFAR *et al*., 2016).

The objectives of this study were to (1) differentiate between species using morphological characters, (2) determine the species relationships based on morpho-molecular data, and (3) explore the universal usage of the DNA barcoding approach to identify rangeland species of *Astragalus* in Northeast Iran.

MATERIALS AND METHODS

Plant material

In the present study, the researchers utilized 43 plant specimens representing 12 different species of *Astragalus* for both morphological and molecular analyses. The majority of the plant material was collected from natural populations located in Golestan province, which is situated in northeastern Iran (as shown in Figures 1, 2 and listed in Table 1).

The plant specimens were identified using established taxonomic keys, including Flora Iranica (PODLECH and ZARRE, 2013) and Flora of Iran (MAASSOUMI, 1998).

To identify the most suitable DNA barcode region for *Astragalus* species, the researchers experimentally tested five candidate genomic regions for DNA barcoding analyses across the 12 *Astragalus* species studied.

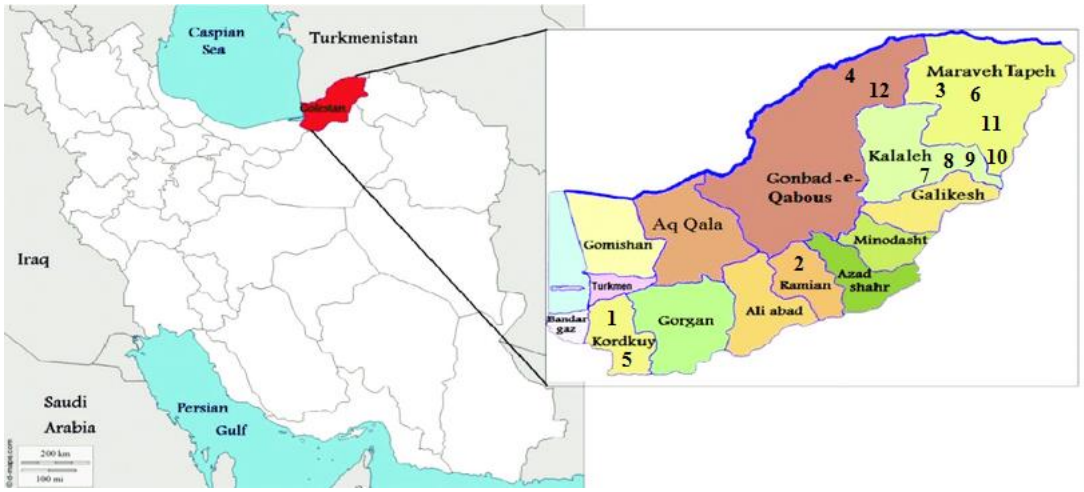

Figure 1. Mapped locations of specimens sampled in this study (1) *Astragalus askius*, (2) *A. aegobromus*, (3) *A. ackerbergensis*, (4) *A. filicaulis*, (5) *A. subalpinus*, (6) *A. podolobus*, (7) *A. hamosus*, (8) *A. glycyphyllos*, (9) *A. memoriosus*, (10) *A. jolderensis*, (11) *A. pseudoarvatensis* and (12) *A. rubromarginatus*

Morphometry

For morphological studies, 43 specimens of 12 species of *Astragalus* were used (Table 1) to examine 12 vegetative and reproductive morphological characters, including 9 qualitative and 3 quantitative (Table 3).

Figure 2. Samples of *Astragalus* species A: *Astragalus aegobromus*, B: *A. jolderensis*, C: *A. podolobus*

Table 1. Astragalus species and locality of species used for molecular analysis

Number in map	Taxa	Section	Collection data (all samples are from Iran)
1	*A. askius* Bunge	*Incani*	Golestan: Kordkoy, protected area of Jahan-nema, SW. slope of Jahan-nema
2	*A. aegobromus* Boiss.&Hohen	*Caprini*	Golestan: Ramian
3	*A. ackerbergensis* Freyn	*Incani*	Golestan: SE of Maraveh Tappeh
4	*A. filicaulis* Kar. & Kir.	*Sesamei*	Golestan: 26 km from Gonbade Kavous
5	*A. subalpinus* Boiss.&Buhse	*Incani*	Golestan: Kordkoy, protected area of Jahan-nema, SW. slope of Jahan-nema
6	*A. podolobus* Boiss.	*Ammodendron*	Golestan: Eslamabad to Maraveh Tappeh
7	*A. hamosus* L.	*Bucerates*	Golestan: Golestan forest, Tang-e Rah
8	*A. glycyphyllos* L.	*Glycophyllus*	Golestan: Golestan forest, inter Tang-e Rah et Tang-e Gol
9	*A. memoriosus* Pakravan, Nasseh & Maassoumi	*Caraganella*	Golestan: Golestan forest, Cheshmeh Khan
10	*A. jolderensis* B.Fedtsch.	*Incani*	Golestan: Golestan forest
11	*A. pseudoarvatensis* Podlech & Sytin	*Ammodendron*	Golestan: Maraveh-Tappeh, Ghazanghieh, eastern part
12	*A. rubromarginatus* Czerniak.	*Caprini*	Golestan: Inche-Boroon to Maraveh Tappeh
	Oxytropis aucheri Boiss.		GeneBank

Morphological analyses

In this study, the researchers coded the morphological characteristics of the plant specimens and standardized the data to have a mean of 0 and a variance of 1. They then used multivariate analyses, including Unweighted Paired Group using Average (UPGMA) and Ward's Minimum-Variance Method (Ward) clustering based on Euclidean distance and Gower distances, as well as principal coordinate analysis (PCoA) to analyze the data.

To identify the most variable morphological characters, the researchers used principal component analysis (PCA) (PODANI, 2000; SAFAEI *et al*., 2016). They also performed a Mantel test to assess the correlation between the geographical distance and the morphological distance of the studied populations (PODANI, 2000).

All data analyses were conducted using PAST version 2.17 (HAMMER *et al*., 2012). Additionally, the researchers used the Neighbor-Net method of networking after 100 times bootstrapping by Splits Tree 4 version 4.11.3 (HUSON and BRYANT, 2005) to visualize relationships among the studied populations.

DNA isolation and amplification

To extract DNA from the plant specimens, the researchers used a purification kit (Tiangen, Korean) and collected fresh and dried leaves (100 mg) from each of the 12 selected species. In order to compare the performance of different DNA markers, each DNA sample (listed in Table 1) was analyzed using five DNA barcoding regions. These regions included plastidial DNA regions such as the *mat*K coding region, the *rpl*32-*trn*$L_{(UAG)}$ intergenic spacer, and the *trn*H-*psb*A non-coding region, as well as two nuclear DNA regions (ITS and ETS) to evaluate the performance of the nuclear genome in comparison with the plastidial genome.

To amplify the DNA regions of interest, PCR was performed using the Applied Biosystem ABI Veriti system in a 25-µL reaction according to the manufacturer's instructions. The PCR cycles included an initial denaturation for 60 s at 94°C, followed by 37 cycles of denaturation (30 s at 94°C), annealing (50 s at different temperatures as listed in Table 2), and extension (1 min at 72°C), and a final extension at 72°C for 7 min. Details of the primers used for amplification are provided in Table 2.

The PCR products were sent for Sanger sequencing to Genetic Codon in Tehran, Iran. The nuclear and plastid datasets were aligned using the web-based version of MUSCLE (EDGAR, 2004, available at www.ebi.ac.uk/Tools/msa/muscle/) under default parameters, followed by manual adjustment.

Molecular analysis

In this study, the researchers conducted maximum parsimony, likelihood, and Bayesian analyses on the intergenic spacer of chloroplast sequences that were aligned and used different methods to examine the species relationships. Maximum parsimony (MP) analyses were conducted using PAUP* ver. 4.0a157 (SWOFFORD, 2002). Maximum likelihood analysis (ML) was performed on each dataset using RAxML ver. 8.2.10 (STAMATAKIS, 2014), as implemented in the CIPRES Science Gateway. Bayesian inference (BI) analyses were performed using MrBayes ver. 3.2 (RONQUIST *et al*., 2012) on the CIPRES Science Gateway (Cyber infrastructure

for Phylogenetic Research cluster) (MILLER *et al*., 2010, available at www.phylo.org) for the datasets. Additionally, the molecular clock test was performed in MEGA 7 (TAMURA *et al.*, 2011).

To evaluate the universality of the five candidate DNA markers, the researchers included regions that were amplified and sequenced in the maximum number of analyzed plants. For each selected DNA marker, only the most universal primer combinations (listed in Table 2) were examined to facilitate the interpretation of successes and failures. For all species and loci, PCR amplification was conducted in a two-stage experiment. Firstly, the standard PCR conditions were used, starting from 10 ng of the DNA template. The second step included only samples that either did not amplify or produced multiple PCR products. Samples of both types of failure were re-amplified using 1 and 25 ng of DNA template. The PCR process was repeated with a reduction of 5°C in the annealing temperature as explained in Table 2, and 40 PCR cycles for samples that failed to amplify. Each marker was assessed in terms of sequence length and alignment success in all the analyzed species.

Table 2 Details of primers used for amplification of the five candidate DNA barcoding markers

Locus	Primer name	Sequences (5′–3′)	Annealing temperature (°C)
Nuclear ITS	ITS5m ITS4	5'–GGAAGGAGAAGTCGTAACAAGG–3' 5'–TCCTCCGCTTATTGATATGC–3'	54
Nuclear ETS	ETS1 18S-IGS	5'–CCACAACTCCTTGCTGAGCTT–3' 5'–GAGACAAGCATATGACTAC–3'	52
Plastidial *rpL*32-*trn*L$_{(UAG)}$	rpL32–F *trn*L$_{(UAG)}$	5'–CAGTTCCAAAAAAACGTACTTC–3' 5'–CTGCTTCCTAAGAGCAGCGT–3'	54
Plastidial *mat*K	*mat*K-390f *mat*K-1326r	5'–CGATCTATTCATTCAATATTTC–3' 5'–TCTAGCACACGAAAGTCGAAGT–3'	50
Plastidial *trn*H-*psb*A	*psb*A *trn*H	5'–GTTATGCATGAACGTAATGCTC–3' 5'–CGCGCATGGTGGATTCACAATCC–3'	53

The researchers also assessed the comparative levels of variation and discrimination of each of the five markers using MEGA 4.0 to produce Kimura 2-parameter (K2P) distance matrices for each locus. In accordance with the guidelines of the Consortium for Life Barcoding, K2P distances were used to assess performance among barcoding

loci (http://www.barcoding.si.edu/protocols.html).) The discrimination values (K2P sequence divergence converted into percent) were calculated for each locus in all the studied taxa.

RESULTS

Morphometry

The UPGMA (Fig. 3), WARD tree (Figure not given) and PCA plot (Figure not given) of morphological characteristics clearly separated the studied species. In PCA analysis, the first three components contained about 39.82% of the total variance, in which Life form, Leaflet number, Corolla color, Bract color, Inflorescence, Embryo orient, Standard length and Peduncle made up 36.36% of the total variation. In species clustering according to morphological traits, UPGMA and Neighbor-Net network dendrogram presented similar results, except species of *A. askius*, *A. subalpinus* and *A. memoriosus* (Figs. 3 and 4). Generally, the studied species are readily distinguished from each other. The mantel test, after 10000 permutations, showed a significant correlation (r = 0.097, P = 0.0026) between geographical distance and morphological distance in the species. Therefore, as the species deviate from each other, they become more divergent in morphological features.

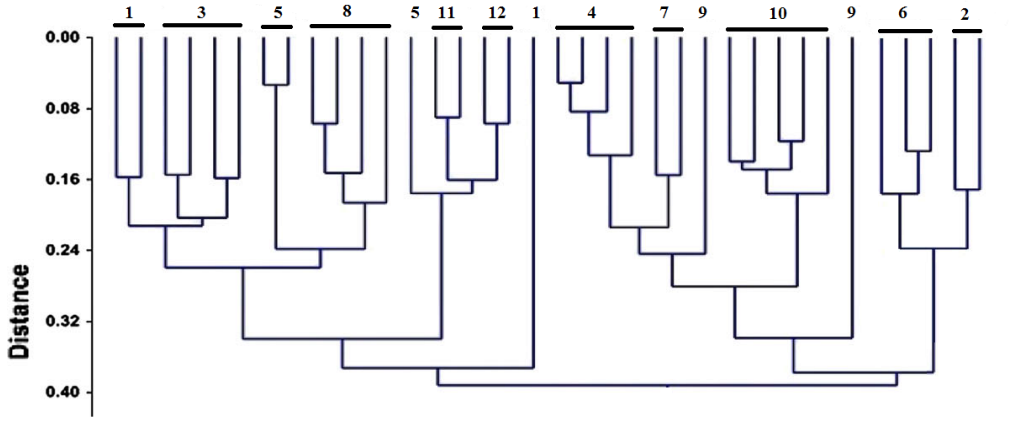

Figure 3. UPGMA dendrogram of morphological characters in *Astragalus* species. (1) *A. askius*, (2) *A. aegobromus*, (3) *A. ackerbergensis*, (4) *A. filicaulis*, (5) *A. subalpinus*, (6) *A. podolobus*, (7) *A. hamosus*, (8) *A. glycyphyllos*, (9) *A. memoriosus*, (10) *A. jolderensis*, (11) *A. pseudoarvatensis* and (12) *A. rubromarginatus*

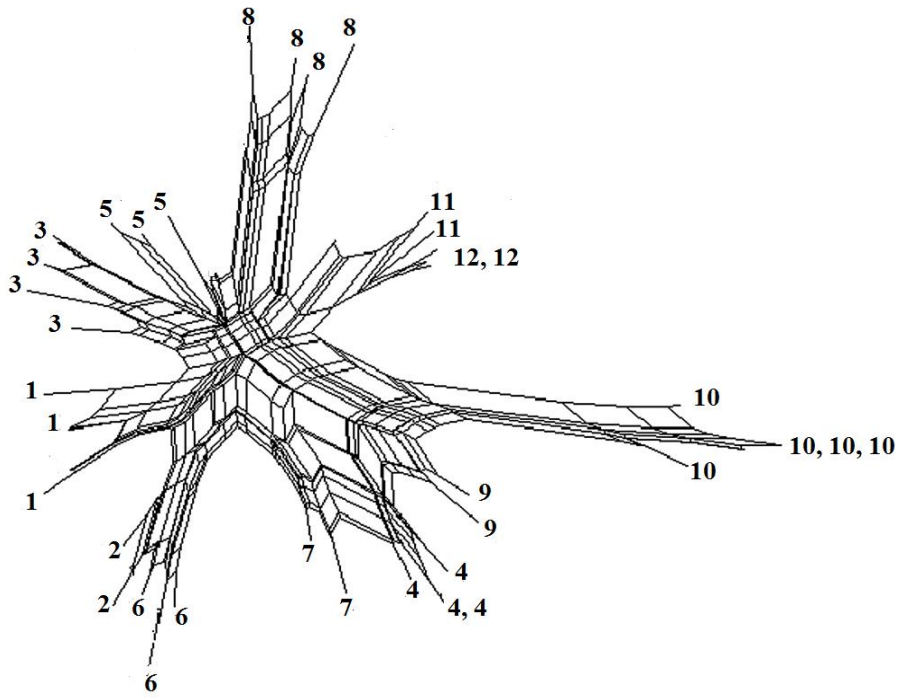

Figure 4. Neighbor Net network of morphological characters in *Astragalus* species. (1) *A. askius*, (2) *A. aegobromus*, (3) *A. ackerbergensis*, (4) *A. filicaulis*, (5) *A. subalpinus*, (6) *A. podolobus*, (7) *A. hamosus*, (8) *A. glycyphyllos*, (9) *A. memoriosus*, (10) *A. jolderensis*, (11) *A. pseudoarvatensis* and (12) *A. rubromarginatus*

Amplification and sequencing success

For 12 taxa (Table 1), high DNA quality was extracted from leaves of all the analyzed samples. The results obtained from DNA barcoding analysis across the selected species presented a considerable difference among the five tested loci with respect to PCR product length, amplification success and sequence quality. All three plastidial markers showed high PCR success using standard primers with 10 ng of DNA as template, and non-relevant differences were found among the species.

About nuclear markers, only the nrDNA ITS using standard primers and 10 ng of DNA as template, for the number of 9 species out of 12 studied species was successfully amplified, whereas ETS exhibits some amplification problems for several samples including from 1, 10 and 25 ng of DNA. All the PCR products for five DNA markers were successfully amplified except six out of 43. The assessment of sequence quality from the studied regions of DNA showed that high quality bidirectional sequences were usually obtained from five regions (ITS, ETS, *rpl*32-

*trn*L$_{(UAG)}$, *trn*H-*psb*A and *mat*K). The highest success rate for bidirectional sequences was observed for plastidial markers. The greatest problems for obtaining bidirectional sequences were observed in relation to ETS sequences, partly attributed to the high frequency of repetitions of some nucleotides, which interfered with disrupting individual sequencing reads. Therefore, it requires more manual editing.

Species relationship based on plastid sequences

The parsimony, likelihood and Bayesian analyses of plastid dataset produced almost congruent trees and gave similar results. Thus, only the Bayesian phylogenetic tree is discussed below (Fig. 5). There are four main groups called A-D in Figure 5. In the plastid tree, from the base of the tree, after the outgroup, there were four groups of all 12 species, and all the members of the species related to the clusters formed separate branches with very high support. The studied species were constituted of four groups. one group (A) included *Incani* section species. The next group (B) included species *of Caprini*. The two species belonging to the *Bucerates* and *Sesamei* sections and the sections of *Glycophyllus* and *Caraganella* together form a sister group called Group C. The next group (D) includes *A. podolobus* and *A. pseudoarvatensis* from sect. *Ammodendron*.

Morphological and molecular phylogenetic trees

The topology resulting from the Bayesian inference of the plastid sequences confronted with the tree obtained from the morphological data have been shown is presented in Figure 6. The pattern of species clustering based on plastid sequences was not similar to that was obtained by morphological traits. For example, according to the morphological characteristics (Table 3), *A. subalpinus* and *A. glycyphyllus* as well as *A. aegobromus* and *A. podolobus*, were closely together, but in the molecular data tree, they were grouped separately.

Table 3. The qualitative and quantitative morphological characters studied in Astragalus species

No	Character	State of character and their codes
1	Life form	erect= (0); caespitose = (1)
2	Leaflet number	3-5 pairs= (0); 1 pairs= (1)
3	Bracteole	present= (0); absent= (1)
4	Corolla color	yellow= (0); yellow-violet= (1); violet = (2)
5	Hair color of calyx	black= (0); black and white= (1); white= (2)
6	Bract color	white= (0); brown= (1); red- white= (2); red= (3)
7	Bract hair	> 3 mm= (0); < and =3 mm= (1)
8	Standard length	> 16 mm= (0); < and =16 mm= (1)
9	Inflorescence	densely flowered (>5) = (0); remotely flowered (1-5) = (1)
10	Calyx hair	absent= (0); present= (1)
11	Embryo orient	vertical (0), horizontal (1)
12	Peduncle	> 4 mm= (0); < and =4 mm= (1)

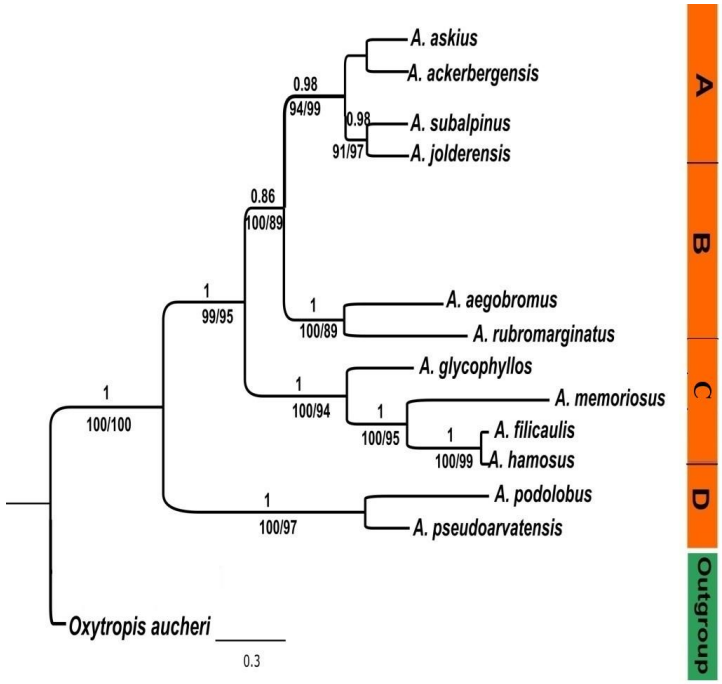

Figure 5. 50% majority rule consensus tree resulting from Bayesian phylogenetic analysis of the plastid dataset. Numbers above and below the branches are posterior probability and likelihood as well as parsimony bootstrap values, respectively. Values < 50% are not shown.

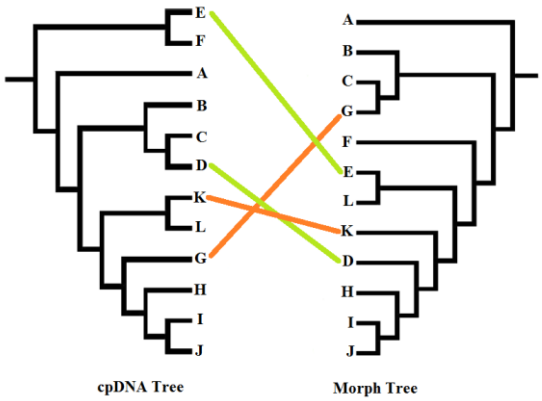

Figure 6. Tanglegram showing the combined trees obtained from cpDNA (*rpl*32-*trn*L$_{(UAG)}$, *mat*K and *trn*H-*psb*A) and morphology data. (A) *A. askius*, (E) *A. aegobromus*, (B) *A. ackerbergensis*, (I) *A. filicaulis*, (C) *A. subalpinus*, (K) *A. podolobus*, (J) *A. hamosus*, (G) *A. glycyphyllos*, (H) *A. memoriosus*, (D) *A. jolderensis*, (L) *A. pseudoarvatensis* and (F) *A. rubromarginatus*

Evaluation of the discrimination value of the five candidate loci

To determine the best DNA barcode marker for identification of the species of *Astragalus*, sequence divergence values were calculated for each sequence (Table 4). Among the plastid markers, the *mat*K gene showed the highest genetic variability among species; however, the divergence was very low among some species: *A. subalpinus* Boiss. & Buhse and *A. jolderensis* B.Fedtsch. exhibited the same *mat*K sequence (Table 4).

In comparison to other plastid and nuclear regions, the *trn*H-*psb*A had the lowest divergence value in studied species. However, the high variability of this marker did not allow to align the sequences properly for some species and to compute the genetic diversity values (Table 4).

The *rpl*32-*trn*L$_{(UAG)}$ sequences showed higher sequence divergences than the other examined markers. But, the efficiency of this marker is very low in several species such as, *A. ackerbergensis* Freyn and *A. hamosus* L. (Table 4).

Nuclear markers revealed consistent genetic diversity among all the studied species. In any case, we emphasize that ETS was amplified in a few species and the sequences of this gene presented a large conserved character. On the contrary, the ITS showed wide genetic diversity in all studied species and amplified well.

The results from sequence analysis in GenBank using BLAST analysis indicated that *mat*K and *rpl*32-*trn*L$_{(UAG)}$ were successful for the most samples, respectively (data not shown). This shows that information databases such as NCBI or EMBL have already cataloged several sequences of *mat*K and *rpl*32-*trn*L$_{(UAG)}$ referable to plant species.

Two loci, including ITS and *rpl*32-*trn*L$_{(UAG)}$, showed six exact matches, while no relevant sequence for the ETS was observed in GenBank.

Table 4. Summary of the proportion of individuals successfully amplified and sequenced and divergence values from five candidate barcoding regions

	ETS	ITS	*rpl*32-*trn*L$_{(UAG)}$	*mat*K	*trn*H-*psb*A
Amplification success	10/12	9/12	12/12	12/12	9/12
Sequencing success	11/12	10/12	12/12	12/12	10/12
Sequence length	580-630	570-620	680-850	788–850	210–472
Alignment success	9/12	11/12	12/12	12/12	10/12
Alignment length	605	610	780	787	240
% of divergence	3.5	5.8	6.3	6.8	3.4

DISCUSSION

In this research, both morphological characteristics and molecular markers were used to identify and determine the delimitation of different species of *Astragalus*. Determining the boundaries of species is important in various biological sciences, including population genetics, ecology, biogeography, biodiversity, and phylogeny (SCHLUTER, 2001; HEDREN, 2004;

RODRIGUEZ *et al*., 2007; DUMINIL *et al*., 2009). The main aim of this research was to provide a method for rapid and accurate identification of unidentifiable rangeland plants based on DNA markers. By using a standardized DNA sequence, researchers can quickly and accurately identify a species and determine its relationship to other species. This can be especially helpful in cases where traditional morphological identification is difficult or impossible, or when the species in question is rare or endangered. Molecular markers have many advantages, including their ability to determine the boundaries of related species, investigate the probable date of the origin of species, and the path of migration and distribution of species in the world (DUMINIL *et al*., 2009; KRAK *et al*., 2013; MINAEIFAR *et al*., 2016; SKUZA *et al*., 2019). Additionally, natural interspecific hybridization can produce new interesting subjects to study the evolution process (ZHU and GAO, 2015; KOOHDAR *et al*., 2018).

The researchers in this study investigated the effectiveness of five different DNA barcode regions, including three plastidial markers and two nuclear markers. In this study, the researchers tested three chloroplast loci (*rpl*32-*trn*L $_{(UAG)}$, *mat*K, and *trn*H-*psb*A) and two nuclear loci (ITS and ETS) as suggested barcodes in plants for differentiation of rangeland species in the genus *Astragalus*.

In species clustering according to morphological traits, UPGMA and Neighbor-Net network dendrogram presented similar results, except species of *A. askius*, *A. subalpinus* and *A. memoriosus*. Generally, the studied species are readily distinguished from each other. Our results proposed that morphology can clarify relationships in *Astragalus* species to a certain extent.
In the plastid tree, there were four groups of all 12 species, and all the members of the species related to the clusters formed separate branches with very high support. The studied species were constituted of four groups. One group (A) included *Incani* section species. The next group (B) included species *of Caprini*. The two species belonging to the *Bucerates* and *Sesamei* sections and the sections of *Glycophyllus* and *Caraganella* together form a sister group called Group C. The next group (D) includes *A. podolobus* and *A. pseudoarvatensis* from sect. *Ammodendron.* However, when analyzing the data, the researchers found that the topology of the Tanglegram tree was very different in terms of morphological and chloroplast data. The limitations of morphological characters can cause incongruence between molecular and morphological trees and need to be carefully considered and analyzed to accurately determine the relationships between species.

The results of the study showed that two plastidial markers presented good results, while only the ITS nuclear marker exhibited effective performance in the analyzed species. Based on these findings, the researchers suggest that an ideal DNA barcode should be selected for the species with standard PCR conditions (HEBERT *et al*., 2003; CHASE *et al*., 2007), and they excluded the ETS nuclear marker among the studied regions and this is in agreement with our study. The use of DNA barcoding can provide a useful tool for identifying plant species quickly and accurately, which can be especially helpful in cases where traditional morphological identification is difficult or impossible. By selecting an appropriate DNA barcode region and using standard PCR conditions, researchers can obtain reliable and consistent results for species identification.

In the analyzed species, all three plastidial markers worked with a single set of PCR conditions. However, the non-coding *trn*H-*psb*A intergenic spacer marker showed low resolution

in species differentiation and was considered inappropriate for plant identification. In contrast, the *mat*K was found to be the most suitable marker for the investigated plants. The *mat*K had PCR priming sites within highly conserved flanking sequences, combined with a non-coding region that showed high numbers of substitutions, making it very appropriate as a plant barcode.

Overall, the selection of an appropriate DNA barcode region is crucial for accurate and reliable species identification. The *mat*K appears to be a promising candidate for plant barcoding, but further studies are needed to determine its effectiveness across a wider range of plant species. The results of the present study confirm the suitability of the *rpl*32-*trn*L$_{(UAG)}$ and *mat*K gene as DNA barcodes for plant identification, in agreement with the studies of KRESS and ERICKSON (2007 . The *rpl*32-*trn*L$_{(UAG)}$ was found to have high resolution in differentiating the analyzed species, while the *mat*K gene showed a high potential for differentiation and presented easy amplification and alignment in the studied species. LAHAYE *et al*. (2008) identified this gene as a suitable and universal barcode for flowering plants and have also identified *mat*K as a suitable and universal barcode for flowering plants. The CBOL Plant Working Group has recommended the combination of *mat*K and *rbc*L as a universal barcode for plant identification, although *rbc*L was not tested in this study due to its low differential power (FAZEKAS *et al*., 2008; HOLLINGSWORTH *et al.*, 2009a; NEWMASTER and RAGUPATHY, 2009). Among the nuclear markers, the ITS marker was found to be suitable for DNA barcoding.

Overall, the results of this study confirm the advantage of DNA barcoding for identifying *Astragalus* species despite the challenges posed by hybridization, reticulation, and chloroplast capture between species.

In this study, *mat*K marker was found to be the most appropriate candidate among the nuclear sequences for plant barcoding. This finding is in contrast to the study of AGHAYEVA *et al*. (2021), which also identified ITS as a potential universal plant barcode. While the results of this study suggest that *mat*K can be a useful DNA barcode for identifying plant species, it is important to note that further evaluation of its performance in larger datasets is necessary. This is because DNA barcoding is a rapidly evolving field, and new methods and technologies are constantly being developed to improve the accuracy and reliability of species identification.

CONCLUSION

In this research, both morphological characteristics and molecular markers were used to determine the delimitation of different species of *Astragalus*. The analysis of chloroplast regions suggested *mat*K as the best and most accurate marker for barcoding. Our study highlights the importance of using barcoding for quick and accurate recognition of plant species and also shows the plastid markers were the most effective in differentiating between species.

ACKNOWLEDGEMENTS

The authors would like to thank all of persons who helped us in this research work. The present study was financially supported by Gonbad Kavous University, Gonbad (Iran).

Received February14, 2023
Accepted December 18, 2023

REFERENCES

AGHAYEVA, P., S., COZZOLINO, D., CAFASSO, V., ALI-ZADE, S., FINESCHI, D., AGHAYEVA (2021): DNA barcoding of native Caucasus herbal plants: potentials and limitations in complex groups and implications for phylogeographic patterns. Biodiversity Data Journal, *9*: e61333.

AHRENS, D., S., FABRIZI, P., ŠÍPEK, PK., LAGO (2013): Integrative analysis of DNA phylogeography and morphology of the European rose chafer (*Cetonia aurata*) to infer species taxonomy and patterns of postglacial colonisation in Europe. Mol. Phylogenetics. Evol., *69*: 83–94.

BERGSTEN, J., D.T., BILTON, T., FUJISAWA, M., ELLIOTT, M.T., MONAGHAN, M., BALKE, L., HENDRICH, J., GEIJER, J., HERRMANN, G.N., FOSTER, I., RIBERA, A.N., NILSSON, T.G., BARRACLOUGH, A.P., VOGLER (2012): The effect of geographical scale of sampling on DNA barcoding. Syst. Biol., *61*:851-69.

BERTHIER, K, M.P., CHAPUIS, S.M., MOOSAVI, D., TOHIDI-ESFAHANI, G.A., SWORD (2011): Nuclear insertions and heteroplasmy of mitochondrial DNA as two sources of intra-individual genomic variation in grasshoppers. Syst. Entomol., *36*: 285–299.

BONAL, R, A., MUÑOZ, A.P., VOGLER (2011): Complex selection on life history traits and the maintenance of variation in exaggerated rostrum length in acorn weevils. Oecologia *167*: 1053–1061.

CASIRAGHI, M, M., LABRA, E, FERRI, A., GALIMBERTI, F., DE MATTIA (2010): DNA barcoding: a six-question tour to improve users' awareness about the method. Brief Bioinform.

CHASE, M.W., R.S., COWAN, P.M., HOLLINGSWORTH, C., VAN DEN BERG, S., MADRIÑÁN, G., PETERSEN *et al*. (2007): A proposal for a standardised protocol to barcode all land plants. Taxon, *56*:295–299.

CHEN, S., H.,YAO, , J. HAN, , C. LIU, , J., SONG, , L., SHI, , Y., ZHU, , X., MA, , T., GAO, , X., PANG, , K., LUO, , Y., LI, , X., LI, , X., JIA, , Y., LIN, , & C., LEON (2010): Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE, *5*(1): 1–8.

DUBEY S, J., MICHAUX, H., BRÜNNER, R., HUTTERER, P., VOGEL (2009): False phylogenies on wood mice due to cryptic cytochrome-b pseudogene. Mol. Phylogenetics. Evol., *50*: 633–641.

DUMINIL, J., M., DI MICHELE (2009): Plant species delimitation: a comparison of morphological and molecular markers. Plant. Biosyst. *143*:528–542.

EDGAR, R. C. (2004): Muscle: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res., *32*: 1792–1797.

FAZEKAS, A.J., K.S., BURGESS, , P.R., KESANAKURTI, , S.W., GRAHAM, , S.G., NEWMASTER, , B.C., HUSBAND, , D.M, .PERCY, , M., HAJIBABAEI, , S.C., BARRETT (2008): Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE, *3*:e2802.

GAO, Z., Y., LIU, , X., WANG, , X., WEI, , J., HAN, (2019): DNA Mini-Barcoding: A Derived Barcoding Method for Herbal Molecular Identification. Front. Plant Sci., 10(August), 00987.

HAMMER, Ø., D.A.T., HARPER., P.D., RYAN (2012): PAST: Paleontological Statistics software package for education and data analysis. Paleontol Electron., *4*:9.

HEBERT, P.D.N, A, .CYWINSKA, S.L., BALL, J.R., DEWAARD (2003): Biological identification through DNA barcodes. Proc. R. Soc. Lond. [Biol], *270*:313–321.

HEDREN, M. (2004): Species delimitation and the partitioning of genetic diversity—an example from the Carex flava complex (Cyperaceae). Biodivers, Conserv., *13*:293–316.

HOLLINGSWORTH, M.L., CLARK, A., FORREST, L.L., RICHARDSON, J.E., PENNINGTON, R.T., LONG, D., COWAN, R., CHASE, M.W., M., GAUDEUL, P.M., HOLLINGSWORTH (2009a): Selecting barcoding loci for plants: evaluation of seven

candidate loci with species level sampling in three divergent groups of land plants. Mol. Ecol. Resour., *9*:439–457.

HOLLINGSWORTH, P.M., L.L., FORRESTA, J.L., SPOUGEB *et al*. (2009b): A DNA barcode for land plants. Proc. Natl. Acad. Sci. USA., *106*:12794–12797.

HUSON, D.H., D., BRYANT (2005): Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol., *23*(254):267.

JINBO, U. , T., KATO, , M., ITO (2011): Current progress in DNA barcoding and future implications for entomology. Entomol. Sci.. *14*:107-124.

KOOHDAR, F., F., ATTAR, S.M., TALEBI, M., SHEIDAI (2018): Contemporary interspecific hybridization between *Dracocephalum kotschyi* and *Dracocephalum oligadenium* (Lamiaceae): Evidence from morphological, anatomical and molecular data. Acta. Biol. Szeged. *62*:123–129.

KRAK, K, P., CAKLOVA, J., CHRTEK, J., FEHRER (2013): Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (Hieracium, Asteraceae). Heredity, *110*:138–151.

KRESS, W. J., AND D. L., ERICKSON (2007): A two-locus global DNA barcode for land plants: the coding rbcL gene complements the noncoding trnH-psbA spacer region. PLoS One, *6*: e508.

LAHAYE, R., M., VAN DER BANK, D., BOGARIN, *et al*. (2008): DNA barcoding the floras of biodiversity hotspots. Proc Natl. Acad. Sci. USA, *105*:2923–2928.

MAASSOUMI, A.A. (1998): Old world check-list of *Astragalus*. – Research Inst. of Forests and Rangelands, Tehran.

MEYER, C.P. and G., PAULAY (2005): DNA Barcoding: Error Rates Based on Comprehensive Sampling. PLoS Biol., *3*(12): 422.

MILLER, M.A., W., PFEIFFER, T., SCHWARTZ (2010): Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, Louisiana. Piscataway: IEEE, 45-52.

MINAEIFAR, A.A., M., SHEIDAI, F., ATTAR, Z., NOORMOHAMMADI, S., GHASEMZADEH- BARAKI, (2016): Biosystematic study in the genus *Cousinia* Cass. (Asteraceae), section *Cousinia*. Biochem. Syst. Ecol., *69*:252–260.

NEWMASTER, S.G., S., RAGUPATHY (2009): Testing plant barcoding in a sister species complex of pantropical Acacia (Mimosoideae, Fabaceae). Mol. Ecol. Resour., *9*:172–180.

NICHOLLS, J. A. R. J., CHALLIS, S., MUTUN, G. N., STONE (2012): Mitochondrial barcodes are diagnostic of shared refugia but not species in hybridizing oak gall wasps. Mol. Ecol., *21*: 4051–4062.

NITHANIYAL, S., S., MAJUMDER, S., UMAPATHY, M., PARANI (2021): Forensic application of DNA barcoding in the identification of commonly occurring poisonous plants. J. Forensic Leg. Med.,*78*(February): 102126.

PIPER, A. M., J., BATOVSKA, N. O. I., COGAN, J., WEISS, J. P., CUNNINGHAM, B. C., PODANI, J. (2000): Introduction to the exploration of multivariate data [english translation]. Backhuyes, Leide Ltd.

PENTON, E. H., P. D. N., HEBERT, T. J., CREASE (2004): Mitochondrial DNA variation in North American populations of Daphnia obtusa: Continentalism or cryptic endemism? Mol. Ecol., *13*(1): 97–107.

PODLECH, D. and S., ZARRE (2013): Taxonomic revision of the genus *Astragalus* L. (Leguminosae) in the Old World, Vol. II. – Naturhistorisches Museum Wien.

RATNASINGHAM, S. and P.D.N., HEBERT (2007): BOLD:thebarcode oflifedatasystem(www.barcodinglife.org). Mol Ecol Notes *7*:355–364.

RODRIGUEZ, J, J., BALCH, K., RODRIGUEZ-CLARK (2007): Assessing extinction risk in the absence of species-level data: quantitative criteria for terrestrial ecosystems. Biodivers. Conserv, *16*:183–209.

RONQUIST, F., M., TESLENKO, P., MARK, D.L., AYRES, A., DARLING, S., HÖHNA, B., LARGET, L., LIU, M.A., SUCHARD, J.P., HUELSENBECK (2012): MrBayes 3.2:efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol., *61*:539-542.

SAFAEI, M., M., SHEIDAI, B., ALIJANPOOR, Z., NOORMOHAMMADI (2016): Species delimitation and genetic diversity analysis in Salvia with the use of ISSR molecular markers. Acta. Bot. Croat., *75*: 45-52.

SAVOLAINEN, V, S., COWAN ROBYN, P., VOGLER ALFRIED,, K., RODERICK GEORGE, LANE RICHARD (2005): Towards writing the encyclopaedia of life: an introduction to DNA barcoding 360Phil. Trans. R. Soc. B.

SCHLUTER, D. (2001): Ecology and the origin of species. Trends. Ecol. Evol., *16*:372–380

SHEIDAI, M., F., AFSHAR, M., KESHAVARZI, S.M., TALEBI, Z., NOORMOHAMMADI, T., SHAFAF (2014): Genetic diversity and genome size variability in *Linum austriacum* (Lineaceae) populations. Biochem. Syst. Ecol. *57*:20–26.

SKUZA, L., I., SZUC´KO, E., FILIP, T., STRZAŁA (2019): Genetic diversity and relationship between cultivated, weedy and wild rye species as revealed by chloroplast and mitochondrial DNA non-coding regions analysis. PloS one, *14*(2):e0213023.

STAMATAKIS, A. (2014): RAxML ver. 8: a tool for phylogenetic analysis and post analysis of large phylogenies. Bioinformatics, *30*: 1312–1313.

SWOFFORD, D.L. (2002): PAUP*: phylogenetic analysis using parsimony (*and other methods), ver. 4.0b10. – Sinauer Associates, Sunderland.

TAMURA, K., D., PETERSON, N., PETERSON, G., STECHER, M., NEI, S., KUMAR (2011): MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol., *28*:2731–2739.

DNK BARKODIRANJE I MORFOLOŠKA IDENTIFIKACIJA *Astragalusa* U IRANU: PREDNOSTI MOLEKULARNOG PRISTUPA

Abolfazl TAHMASEBI[1*], Bakhtiar FATTAHI[2]

[1] Fakultet za poljoprivredu i prirodne resurse, Univerzitet Gonbad Kavous, Gonbad, Iran
[2]Fakultet prirodnih resursa i inženjerstva životne sredine, Univerzitet Malajer, Malajer, Iran

Izvod

Astragalus je ključna krmna biljka u pašnjacima severoistočnog Irana. Međutim, identifikacija različitih vrsta Astragalusa može biti izazovna zbog njihovih preklapajućih morfoloških karakteristika. U ovoj studiji, imali smo za cilj da utvrdimo odnose između 12 vrsta *Astragalusa* koristeći i morfološke i molekularne osobine. Takođe smo procenili efikasnost DNK barkodiranja kao alata za identifikaciju vrsta pašnjaka *Astragalusa* u severoistočnom Iranu. Ispitali smo pet DNK bar kodova, uključujući tri cpDNK regiona (trnH-psbA, rpl32-trnL$_{(UAG)}$ i matK) i dve nuklearne sekvence (ITS i ETS). Otkrili smo da su markeri plastida (rpl32-trnL(UAG, matK) najefikasniji u razlikovanju vrsta. Takođe smo otkrili da morfološki podaci, predstavljeni Neighbor-Net mrežom i UPGMA dendrogramom, imaju potencijal da razdvoje vrste *Astragalus*. Identifikovali smo matK kao najbolji i najtačniji marker za barkodiranje, pošto je trnH-psbA imao neke nedostatke zbog veličine sekvence i problema sa poravnanjem. Naša studija naglašava važnost upotrebe barkodiranja za brzo i tačno prepoznavanje biljnih vrsta i pokazuje da su markeri plastida bili najefikasniji u razlikovanju između vrsta.

Primljeno 14.II.2023.
Odobreno 18. XII.2023.