Pak. J. Bot., 56(5): 1963-1971, 2024. DOI: http://dx.doi.org/10.30848/PJB2024-5(33)

BARCODING OF ASTERACEAE PLANTS OF JUNIPER ECOSYSTEM ZIARAT, BALOCHISTAN

SHAZIA IRFAN[1], SHAHJAHAN SHABBIR AHMED RANA[2], MUHAMMAD MUSHTAQ[2], IMRAN ALI SANI[2], RAZIA KANWAL[1], SAADULLAH[2], DAWOOD SHAHID[2], RUKHSANA JABEEN[1] AND NAZEER AHMED[2]*

[1]Department of Botany, Sardar Bahadur Khan Women's University, Quetta
[2]Department of Biotechnology, Balochistan University of Information Technology, Engineering and Management Sciences (BUITEMS), Quetta
**Corresponding author's email: nazeer.ahmed@buitms.edu.pk*

Abstract

The Asteraceae family, which includes some economically and medicinally important plant species such as sunflower (*Helianthus annus*), salad leaves (*Lactuca sativa*) and *Artemisia* species has been a focus of DNA barcoding research due to the challenges in identifying these species using traditional morphological methods. The present study aimed to develop and evaluate the ability of the DNA barcodes targeting chloroplast genomic regions i.e., Riboluse-1,5- Biphosphate Carboxylase/ Oxygenase Large subunit (*rbcL*), and Maturase K (*matK*) for several plant species belonging to family Asteraceae collected from Juniper ecosystem Ziarat, Pakistan. The overall results indicated that both *matK* and *rbcL* markers successfully amplified the targeted regions and sequenced. During species resolution, *matK* (70% and 82%) showed better results compared to *rbcL* (42% and 48%) through BLASTN analysis and Neighbor Joining tree analysis respectively. Interspecific and Intraspecific divergence showed 100% discrimination by both markers while *matK* showed 63% success in discriminating species within a genus compared to *rbcL* (50%). A total of 22 novel barcodes which were found effective in discriminating and identifying the evolutionary links among the species under study were finally developed and submitted to the Barcode of Life Database (BOLD).

Key words: Asteraceae, *matK, rbcL,* BLASTN, NJ, NN

Introduction

Asteraceae is considered one of the biggest families of flowering plants comprising over 1,600 genera and 23,600 species (Panero & Crozier, 2016). The Asteraceae family comprises approximately 770 species, making it the largest plant family in Pakistan (Abid & Qaiserd, 2021). This family contains plants of aesthetic and therapeutic value. The Balochistan province of Pakistan also harbors a number of species of Asteraceae spreading over in different areas like Ziarat, Kalat, Rodenjo, Nimargh, Mangocher, Harboi, Nichara, Surab, Zehri, and Khuzdar (Tareen *et al*., 2010).

Ziarat Biosphere Reserve is a valuable herbaceous genetic reserve that spans an area of 110,000 hectares at altitude ranging from 5500- 9500 feet. Remarkable endemic and unique plants including medicinal plants are found in this region because of their extraordinary conformity/compatibility and adaptation to the environment. The Ziarat locals employ 90 species of different genera from 35 families as medicines for a variety of illnesses (Sarangzai *et al*., 2013). These plants include *Hertia intermedia* used for acne and wood is utilized as fuel. Children with stomachaches are given decoction of *Achillea wilhemsii*. Various components, including fruits, leaves, stems, and bark, are used to treat a variety of conditions, including coughs, colds, stomach cramps, asthma, diuretics, carminatives, stimulants, dropsies, gonorrhea, leucorrhea, and skin problems. Certain plants are used in cosmetic goods, and gin are popularly flavored with juniper berry oil (Sarangzai *et al*., 2013).

Identification of Juniper Forest plant species is essential for both resource development and preservation. Geographical variations make it challenging to classify plants merely based on their morphological character, and therefore necessitates the use of professional taxonomists (Chase & Fay, 2009). Identification at the species level is frequently challenging due to the diversity of plants in the Asteraceae family (Bayer & Starr, 1998). A simple and reliable way of authenticating Asteraceae species is essential for its conservation and other downstream uses including international trade and medicinal use. A thorough understanding of phylogenetic and systematics is also hindered by the incomplete knowledge of some lineages of large genera, such as *Artemisia, Tanacetum, Centaurea* and *Taraxacum.* For example, *Artemisia scoparia* and *Serriphedium maritima* (syn *Artemisia maritima*) and *Serriphedium quettensis* are the most difficult species to identify morphologically.

In such circumstances, The Consortium for the Barcode of Life (CBOL) plant working group recommended *rbcL* and *matK* as two strong markers as plant barcodes (Group, 2009, Dormontt *et al*., 2018). These markers are commonly used in plant DNA barcoding studies and have been shown to be effective in identifying plant species (Kress, 2017) compared to 5' region of the mitochondrial cytochrome c oxidase I (COI) gene, which is an ideal DNA barcode for animals only (Howard *et al*., 2020). The choice of marker depends on the taxonomic group being studied, and the availability of reference data (Mahadani *et al*., 2022). In addition, building DNA barcode databases for plants has historically been ad hoc and with a relatively narrow taxonomic focus, which limits the usefulness of DNA barcoding for plants (Dormontt *et al*., 2018). A number of markers have, therefore, been suggested as plant barcode markers (Kress & Erickson, 2007, Song *et al*., 2009, Chen *et al*., 2010, Hollingsworth *et al*., 2011, Hosein *et al*., 2017).

To realize the potential of DNA barcoding for plants, systematic sequencing of the entire flora of a region is necessary, and appropriate archiving and publication of

data are important for effective plant DNA barcode databases (Kress, 2017). DNA barcoding and genomics both place a strong focus on gathering vast amounts of genetic data, which enables insights into genotypic - phenotypic concerns that are outside the purview of the conventional taxonomic disciplines.

Plant DNA barcoding research has made significant progress in recent years. It is a method of identifying biological specimens and assigning them to their respective species using a standardized DNA sequence (Bhargava & Sharma, 2013, Mahadani *et al*., 2022). This technique has been useful for community ecologists seeking to understand the factors affecting plant diversity (Howard *et al*., 2020) as well as to find out undiscovered species that have not been recorded yet (Mahadani *et al*., 2022). A DNA barcode consists of a standard short DNA sequence (400–800bp) that is easy to create and would be different for every species on the planet.

This study aims to assess the efficacy of the *rbcL* and *matK* DNA barcodes, both individually and in combination, for the tentative identification and classification of examined plant species of the family Asteraceae from Juniper ecosystem of Ziarat, Pakistan and also to construct phylogenetic tree to assess the evolutionary relationships among the Asteraceae by constructing Neighbor Joining trees and conducting barcode gap analysis.

Material and Methods

Study Area and Sample Collection: The study area is located between 30°18’52.9" to 30° 27' 5.04"N and 067°21’26.5 to 67° 48' 10.188" E, and the elevation is between 6531 to 9176 feet. Thirty-three (33) plant samples belonging to family Asteraceae were collected in different months during the springs of 2016 to 2018 from different altitudes of Juniper ecosystem Ziarat (Fig. 1). The pool of collected plant samples contained at least one individual per species.

Identification of plants: Fresh leaves were put in silica gel beads after collection till before DNA extraction. All the plants were preliminary identified after the evaluation of morphological features after reference to e-floras (https://openherbarium.org). The images of the herbarium specimens as well as specimen data, including voucher details, taxonomy, and locality for each sample, were posted to BOLD (Barcode of Life Database) after genomic identification and barcoding under the DNA Barcoding of Flora of Ziarat Paksitan (DBFZP) project (http:// http://www.barcodinglife.org/DBFZP/Asteraceae).

DNA extraction, amplification and sequencing: Genomic DNA was extracted from silica gel-dried plant leaves using modified CTAB (Cetyl trimethylammonium bromide) protocol (Li *et al*., 2013) and different kits including Plant Universal Genomic DNA kit (TIANGEN Biotech Beijing Co., China), QIAGEN DNeasy Plant Mini Kit (QIAGEN Inc., Valencia, CA, USA) and WizPrepTM Plant DNA Mini Kit (50 PREP) (https://www.wizbiosolution.com/product-page/plant-dna-mini-kit). Kit based extraction was carried out according to the manufacturer’s protocol.

For PCR amplification, the target DNA sequences were two chloroplast regions consisting of coding genes *matK* and rbcL (Table 1). Amplification was performed in 20µl reactions containing 10µl PCR Master mix, 8µl PCR water, 0.5µl each primer and 1µl DNA template. Primer pairs for PCR and sequencing used in this study are listed in (Table 1).

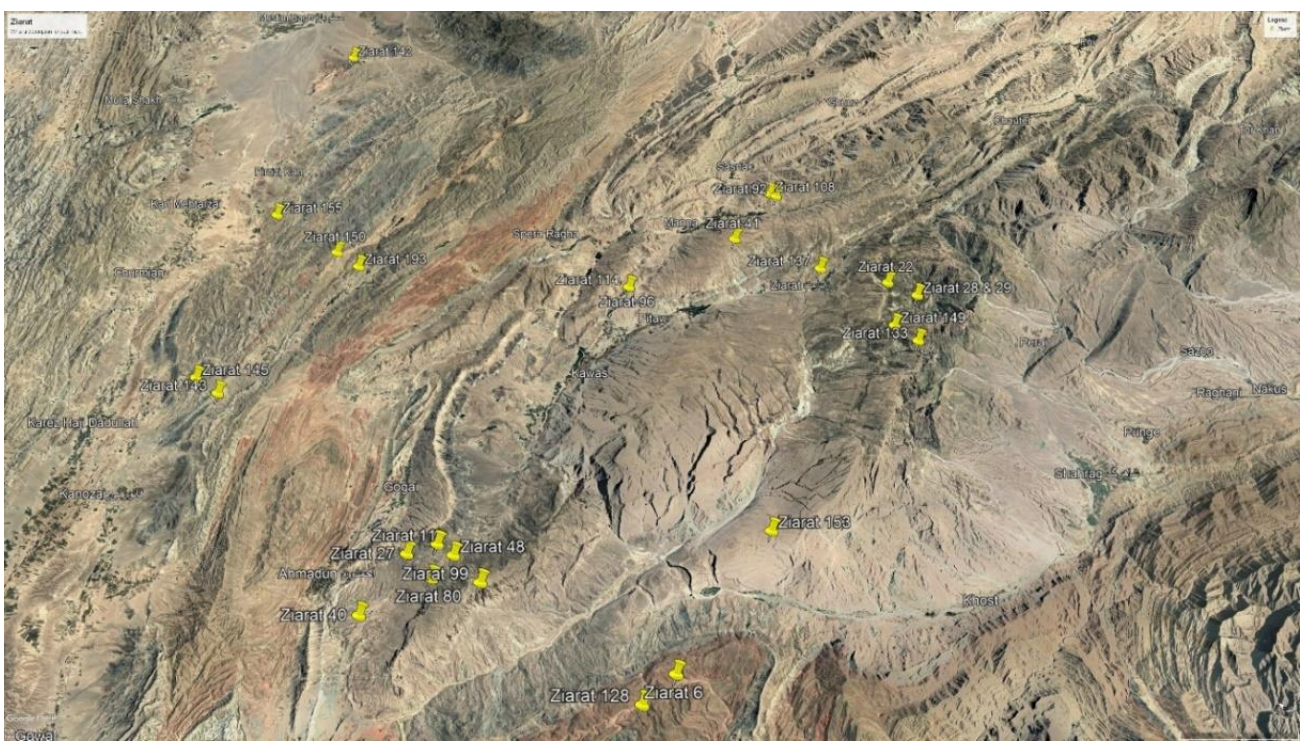

Fig. 1. Sample coordinates of studied area of Juniper Ecosystem Ziarat.

Table 1. Details of molecular markers utilized in the study.

Primer sets	Primers F/R	Primer Sequence (5′-3′)	Expected amplicon size range	References
MatK	*MatK*-1RKim F	ACCCAGTCCATCTGGAAATCTTGGTTC	840bp	Ki-Joong Kim (unpublished)
MatK	*MatK*-3F-Kim R	CGTACAGTACTTTTGTGTTTACGAG		Ki-Joong Kim (unpublished)
RbcL	*rbcL*a-F	ATGTCACCACAAACAGAGACTAAAGC	550bp	Levin *et al*., 2003
RbcL	*rbcL*a-R	GTAAAATCAAGTCCACCRCG		Kress & Eickson, 2007

The PCR Program for studied barcoding markers were as follows:

***rbcL*:** Initial denaturation at 94°C for 5 min followed by 30 cycles of denaturation at 94°C for 45sec, annealing at 61.5-63°C for 45 sec, elongation at 72°C for 60 sec, and final extension at 72°C for 4 min.

***matK*:** Initial denaturation at 94°C for 5 min followed by 35 cycles of denaturation at 94°C for 60sec, annealing at 52°C for 45 sec, elongation at 72°C for 60 sec, and final extension at 72°C for 5 min.

The PCR products were visualized in 2% agarose gel through gel electrophoresis which was run for 40 minutes at 110 volts. Amplicons were sequenced through Sanger Sequencing method in forward direction by Tsingke Biological Technology through ASKER & Co. Lahore (http://www.tsingke.net).

Sequence analysis: The sequence data was visualized with Finch TV Version 1.40 (TV). Nucleotide Basic Local Alignment Search Tool (BLASTN) was used for similarity search. The potential DNA barcode's sequences were aligned using Clustal W, and the Kimura 2-Parameter (K2P) model was used to determine genetic distances in MEGA 7 (Kumar *et al*., 2016). Barcode Gap Analysis was performed and DNA barcodes were generated in Barcode of Life Database (BOLD).

The amplification and sequencing success of *rbcL* and *matK* was calculated when either one of the markers was amplified or sequenced. The ratio of successful amplification for the 2 barcoding markers was determined by dividing the total number of amplified samples by the total number of tested samples.

Species resolution: Asteraceae species resolution of both markers were analyzed using three different methods: Basic Local Alignment Search tool (BLASTN) analyses, Neighbor Joining (NJ) tree analysis and Nearest Neighbor (NN) analysis.

Basic local alignment search tool (BLASTN) analysis: For the dual locus (*rbcL*and *matK*) barcode, in a BLAST analysis, a taxon was scored as fully discriminated if it was resolved by *rbcL* and/or *matK*. A species or genus was scored as fully resolved if the top score for each sample was the species in question itself or another ebst match if the top hits shared numerous species including the correct one, and incorrect/mismatch if all of the top results belonged to a species other than the query species.

Neighbor joining (NJ) analysis: Three separate Neighbor Joining trees (*rbcL*, *matK*, and *rbcL+matK*) were created in MEGA 7 (Kumar *et al*., 2016) and were examined on the basis of formation of monophyletic or paraphyletic/ polyphyletic clads with a strong (≥50%) or weak (<50%) bootstrap support.

Nearest neighbor (NN) analysis: The "Barcode Gap Analysis" of *matK* and *rbcL* sequences was performed on BOLD. The distributions of intra-specific v/s. inter-specific variability were compared using the DNA "barcode gap." A species is considered distinct from its nearest neighbor if its minimum inter-specific distance was greater than its maximum intra-specific distance.

Observation and Results

Taxonomic identification: Out of 33 plant samples from the Juniper Ecosystem Ziarat, 24 plant species of the family Asteraceae were taxonomically recognized (Fig. 2).

Amplification and sequencing success: The success of the PCR amplification and the recoverability of the sequence are crucial factors in determining the barcode efficiency. Our collection represented 19 genera and 24 species of family Asteraceae in the studied area. The amplification success for *matK* (100%) was better than *rbcL* and *rbcL+matK* (91%). The *rbcL* marker amplified 30 samples out of 33 resulting in identification of 21species out of total 24. On The other hand *matK* marker was successfully amplified all samples and identified all 24 species. The sequencing success rate, on the other hand, was recorded as 100% (21/21 species) for *rbcL* and *rbcL+matK* (20/20species) each and 96% (23/24) for *matK*. The average product sizes of the two barcode primers, *rbcL* and *matK*, for all species were around 525 bp and 800 bp, respectively.

Species resolution

Species identification through BLASTN analysis: As a result of BLASTN analysis tool offered by the NCBI a total of 24 species belonging to 19 genera were identified. The identification at the species or generic level was based on the similarity with corresponding sequences (Table 2).

The BLASTN analysis of the 29 sequences obtained from *rbc*L resulted in 21 species from 16 genera. Out of 21 species 5 sequences were fully resolved and identified corroborating to morphological identification. However, 7 species showed another best match, whereas 9 species have no data record (Table 2).

On the other hand, the BLAST analysis of 32 sequences obtained from *matK* resulted in identification of 23 species form 18 genera. Seven sequences were fully resolved corroborating the morphological identification. Whereas, 3 species remained unresolved and 13 species had no data record. Thus, the species resolution based on *matK* locus using BLAST method was 70% (7 out of 10, Table 2).

Fig. 2. Images of Asteraceae species in the study area.

Phylogenetic relationship through neighbor joining (NJ) analysis: The NJ analysis of species of family Asteraceae from Juniper Ecosystem Ziarat and reference sequences showed proper placement of all species in their respective genera in three trees of *rbcL*, *matK* and *rbcL+matK* (Figs. 3-5). However, in terms of node support value, the barcodes, individually and in combination, behaved differently. For example, out of 21 *rbcL* sequences of studied plants and 7 reference sequences, 13 nodes (52%) had node support value of < 50% and 12 nodes (48%) had node support value of ≥50%. Similarly, 5 nodes, out of total 23 *matK* sequences of plants of studied area and 11 reference sequences, 5 nodes (17%) had node support value < 50% while 25 species (83%) had node support value ≥50%. When we combined *rbcL+matK,* 3 nodes (18%) had node support value < 50% and 17 nodes (82%) had node support value ≥50%. Overall, on the basis of NJ analysis, *matK* showed better species discrimination compared to *rbcL+ matK* and *rbcL*.

Barcode gap analysis through nearest neighbor (NN) analysis: The barcode gap analysis was performed through Nearest Neighbor analysis using Kimura-2-Parameter and Kalign for *rbcL* (table 3) and *matK* (table 4) for all the plants of family Asteraceae. Out of 21 species with *rbcL* barcode, 06 species belonged to two genera i.e., *Artemisia* and *Taraxacum*. With 0.00 interspecific distance, this barcode failed to discriminate 3 species of *Artemisia* like *A. maritima*, *A. scoparia* and *Artemisia* sp. except *A. scoparia* which was discriminated with *A. sieversiana* due to 0.19 interspecific distance. When we analyzed the remaining 2 species of genus *Taraxacum*, here barcode performed well with 0.57 interspecific distance between both species. On the other hand, 12 species belonged to different genera with single specimen and remaining 3 species contained multiple specimens. It may be mentioned here that in case of species with multiple specimens had no intraspecific distance indicating that all belonged to same species. The barcode performed well in case of all 18 out of 21 identified species (interspecific distance 0.19 to 2.11) except 3 species of *Artemisia.*

Except *Rhaponticum repens* all remaining 23 species had *matK* barcode. Among 23 species, 8 species belonged to 3 genera. Four species belonged to genus *Artemisia,* two belonged to genus *Himalaiella* and remaining two species were the part of the genus *Taraxacum*. With 0.13 interspecific distance, this barcode discriminated the three species of *Artemisia* i.e., *Artemisia maritima*, *A. scoparia* and *A. sieversiana*. One species of *Artemisia* discriminated with 0.13 interspecific distance between *Artemisia* sp. and *A. maritima*. In all 4 species of *Artemisia* intraspecific distance was not mentioned due to absence of specimens of each species. When we analyzed the remaining two species of genus *Taraxacum*, here barcode performed well with 0.12 interspecific distance between both species. The two species of genus *Himalaiella* had zero interspecific distance between *Himalaiella deltoidea* and *H. heteromalla*, which indicated that *matK* barcode did not perform well. The intraspecific distance was not applicable since all 8 species under study comprised single specimen each. In this family, there were two species with zero intraspecific distance while in case of remaining 13 species intra specific distance was not applicable due to single specimen under study. In case of all 15 species the barcode performed well and discriminated all species by showing the range of interspecific distance from 0 to 4.48. So, on the basis of above-mentioned results, this barcode performed well with 53% (5/8 species) success rate in the identification of species belonging to same genus but in the species of different genera barcode showed 100% (15/15) success rate by discriminating other 15 species.

Discussion

Universality of the two candidate barcodes: This study aimed to inventory the Asteraceae plant species within the Ziarat Juniper ecosystem, sampled through a random sampling method during multiple visits to the study site. The Asteraceae family, known for its vast diversity and ecological significance, poses challenges in identification using traditional morphological methods. To address this, the study evaluates the utility of the *rbcL* and *matK* DNA barcodes, both individually and in combination, for accurate identification and classification of the randomly selected Asteraceae species. This allowed for a broader understanding of the genetic diversity within this unique ecological setting. Through this inclusive approach, the project provides insights into the genetic makeup of Asteraceae species, contributing to the Barcode of Life Database (BOLD. This genetic data enhances the resources available for future plant identification and biodiversity studies, supporting conservation efforts and sustainable management practices in the region.

In the current study, we investigated comparative utility of DNA barcodes (*matK* and *rbcL*) of members of Asteraceae family (Table 2, Fig. 2) to decipher, among others, the genetic links among the chosen species. A barcode needs to be readable by a variety of species in order to be deemed universal. It should ideally be somewhat facilitating amplification, and sequencing. In terms of amplification efficiency, *matK* performed better compared to *rbcL* (100% and 91% respectively), however, in terms of sequencing success rbcL barcode showed better success (100%) than matK barcodes (96%). *matK* was the most easily amplified and sequenced barcode in 23 species except *Rhaponticum repens* while *rbcL* followed with 21 species except *Himalaiella heteromalla, Himalaiella deltoidea* and *Cichorium intybus* of Juniperus Forest Ziarat. Prior studies revealed that the success rate of *matK*-based amplification ranged between 40% to 97% (Kress & Erickson 2007, Zhang *et al*., 2012). Our research is consistent with studies on *Solanum nigrum, Euphorbia helioscopia* and *Dalbergia sissoo* (Wattoo *et al*., 2016). Similar trends were observed in a study on Philippines ethnomedicinal Apocynace plants (Cabelin & Alejandro, 2016). Over all the results obtained from the study of Asteraceae plants in the Juniper Forest ecosystem in Ziarat provide insights into the success rates and challenges encountered in the identification of species using molecular techniques like DNA barcoding (*matK* and *rbcL*).

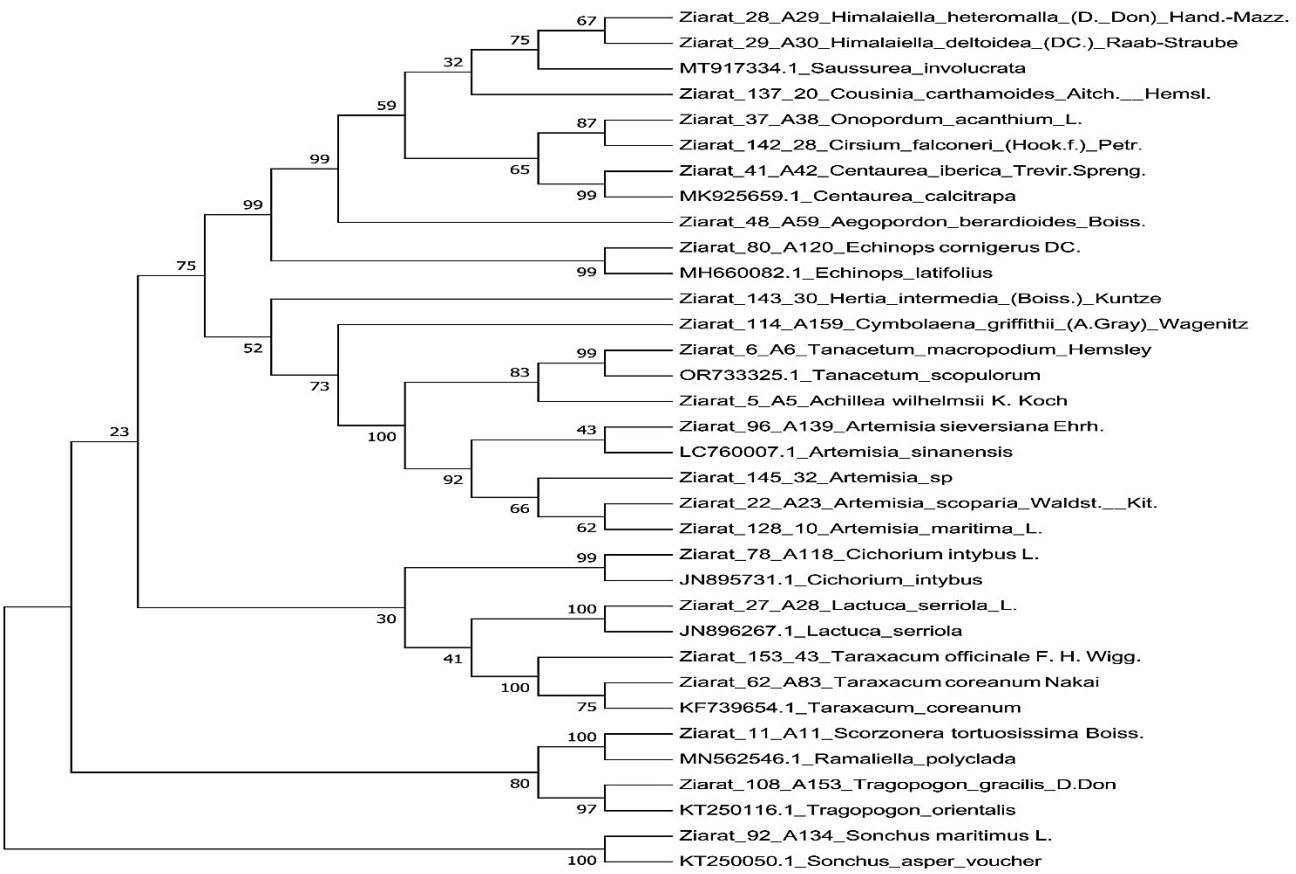

Fig. 3. Evolutionary relationships of taxa of Asteraceae in the Juniper Forest ecosystem Ziarat of *matK* barcode using Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Kimura 2-parameter method and were in the units of the number of base substitutions per site. Evolutionary analyses were conducted in MEGA7.

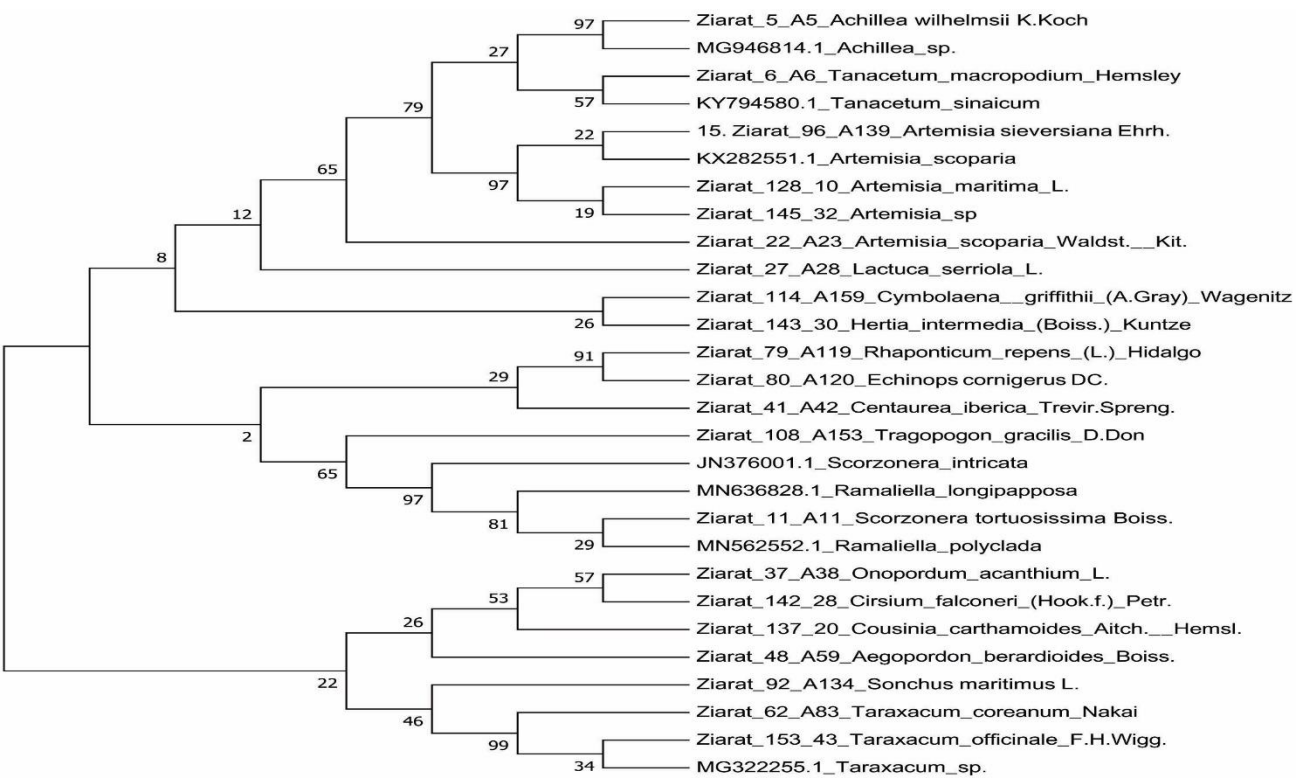

Fig. 4. Evolutionary relationships of taxa of Asteraceae plants in the Juniper Forest ecosystem Ziarat of *rbcL* barcode using Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Evolutionary analyses were conducted in MEGA7.

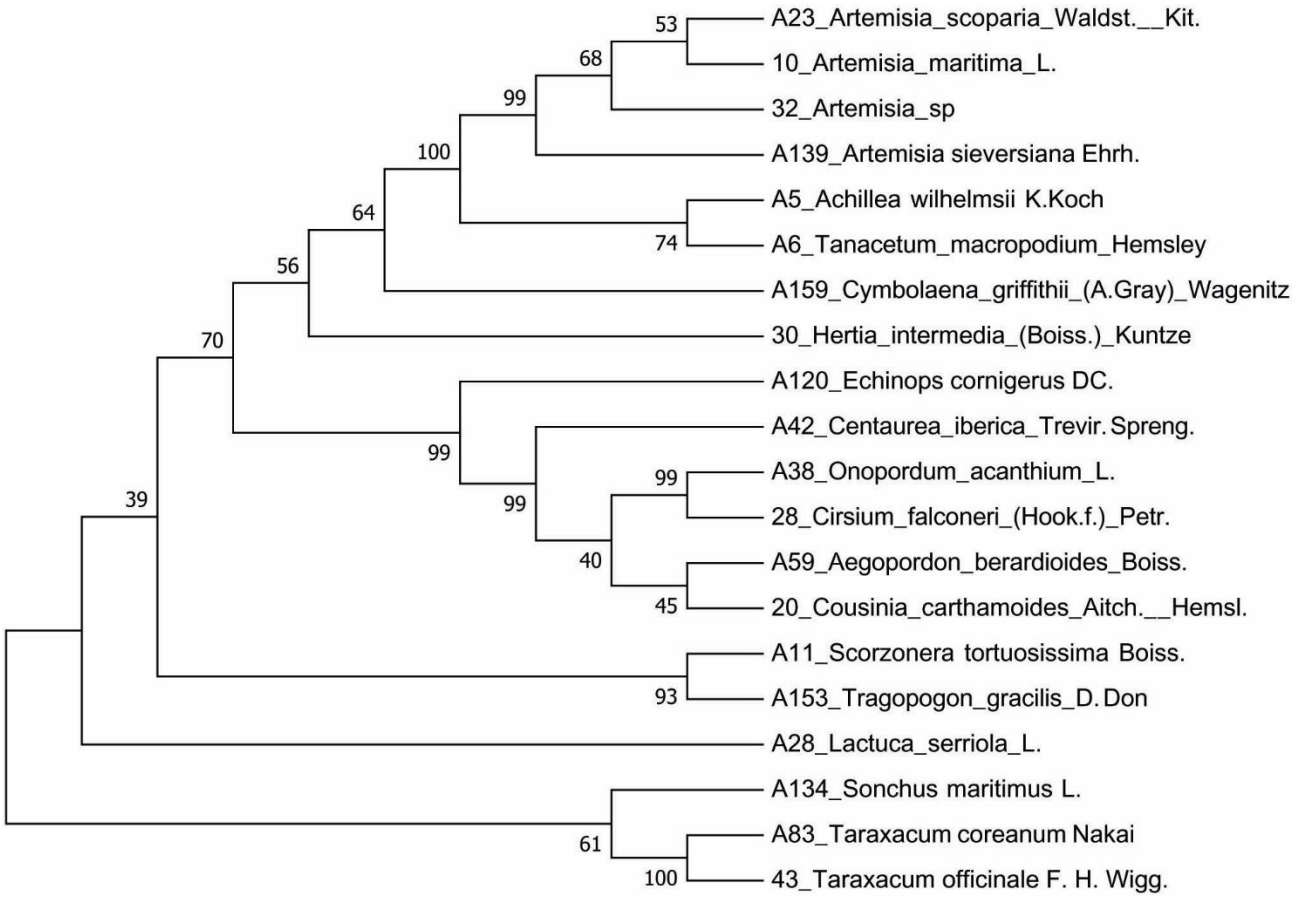

Fig. 5. Evolutionary relationships of taxa of Asteraceae plants in the Juniper Forest ecosystem Ziarat of *rbcL+matK* barcode using Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Evolutionary analyses were conducted in MEGA7.

Species resolution of the two cp DNA markers: The present study contributed 22 novel *matK* and *rbcL* barcodes to sequence database namely BOLD. Such additions in the reference sequence library are supposed to facilitate the future researchers who want to further explore species of family Asteraceae. Furthermore, databases like GenBank and BOLD have scanty published sequence availability particularly for endemic species and species of relatively unexplored areas like Juniper ecosystem Ziarat.

The resolving power of barcoding markers, to identify unknown species, is usually determined by comparing a specific query barcode sequence to available database sequences. For this purpose, most commonly used methods are BLAST and NJ. Positive outcomes include correct matches in BLAST and resolutions in NJ (i.e., monophyly of specimens with higher bootstrap values). According to Table 2, *matK* has the highest rates of correct identification in BLAST, both at the generic and species level followed by dual locus barcode. This confirm the earlier findings establishing superiority of *matK* barcodes over *rbcL* barcodes in identifying species (Hollingsworth *et al*., 2011). In another study conducted by Amandita *et al*., (2019), *matK* resolved 78% plants accurately corroborating with our findings where *matK* resolved 70% species correctly (Amandita *et al*., 2019).

Few of the other studies on 1667 angiosperms (Lahaye *et al*., 2008), Orchidaceae species (Kress & Erickson, 2007) and Rosaceae species from 22 genera (Pang *et al*., 2011) also authenticated better resolving ability of *matK*. Similarly, Gao *et al*., (2021) discovered that the Family Asteraceae barcode had good resolution in 63 species and 48 genera. However, a study on the tropical cloud forest achieved almost similar species resolution 42.88% ± 2.59% with *matK* and 41.50% ± 2.81% with *rbcL* (Kang *et al*., 2017).

Similar trends were observed in NJ analysis (Figs. 3-5), where *matK* performed better compared to *rbcL* and dual locus barcode showing monophyly with good node support values. In comparison to *rbcL* (48%) and *rbcL+matK* (79%) the overall *matK* NJ analysis demonstrated improved species discrimination with 83% node support. This observation is consistent with erstwhile reports (Asahina *et al*., 2010, Mahadani *et al*., 2013).

It is reported that *rbcL* fits all the requirements for barcoding because it is simple to amplify, sequence, and align in most terrestrial plants (Hollingsworth *et al*., 2016). Unfortunately, as shown by our data, it only has minor discriminatory power. Low genetic distances across most species may be the cause of the poor resolution in tree topologies with low bootstrap scores and polytomies found for *rbcL* sequences (Kress *et al*., 2009). Because *rbcL* is known to have inadequate nucleotide sequence diversity to discriminate between closely related species, several researchers have observed low resolution in this gene (Kress & Erickson, 2007, Newmaster *et al*., 2013).

Neighbor joining (NJ) analysis: The study of phylogenetic and evolutionary relationships among the studied taxa of Asteraceae involves analyzing genetic data to understand their phylogenetic relationships/relatedness. The study employed molecular markers like *matK* and *rbcL* to elucidate phylogenetic relationships. These markers are often utilized in plant phylogenetic due to their conserved regions and variable sites, aiding in reconstructing evolutionary histories. NJ analysis demonstrated that most species were placed correctly within their respective genera, but certain species had lower node support values (<50%), indicating potential uncertainties in their placements. *matK* showed a higher percentage (83%) of species with node support values ≥50% compared to *rbcL*+*matK* (79%) and *rbcL* (48%).

The NJ analysis likely revealed the evolutionary divergence among species within the Asteraceae family. It showcased how these plants had evolved distinct genetic lineages over time. The trees might have shown the clustering of species into genera and higher taxonomic groups, providing a framework to understand their evolutionary relatedness and shared ancestry. Examination of the phylogenetic trees might have indicated whether certain genera or groups are monophyletic (sharing a common ancestor) or polyphyletic (having multiple evolutionary origins). Monophyletic groups suggest closer evolutionary relationships. It's possible that the analysis highlighted regions or species within the ecosystem that serve as evolutionary hotspots, indicating higher diversification or unique evolutionary trajectories. The study might have inferred instances of adaptive radiation, where certain lineages diversified rapidly to exploit various ecological niches within the Juniper Forest ecosystem.

Nearest neighbor (NN) analysis: The minimum inter-specific distance and the maximum intra-specific distance were used to describe intra-specific variation, whereas the maximum intra-specific distance and average intra-specific distance were utilized to characterize inter-specific divergence. When *matK* and *rbcL* potential DNA regions were compared for their inter-specific divergences, *matK* had the largest maximum inter-specific distance and superior congeneric species discrimination than *rbcL*. Both markers discriminating non-congeneric species efficiently. Our results were consistent in the study on Poaceae family plants (Ullah *et al*., 2016) *Clinacanthus nutans* (Ismail *et al*., 2018) and Dendrobium (Orchidaceae) species (Mahadani *et al*., 2022).

In nutshell, the findings of this study have important implications in the field of plant, barcoding and its potential applications. The use of DNA barcoding markers, particularly *rbcL* and *matK*, can provide a simple and reliable way of authenticating Asteraceae species, which is essential for their conservation, international trade, and medicinal use. Additionally, the use of these markers in combination and conducting barcode gap analysis can improve the accuracy of species identification and contribute to the understanding of the evolutionary relationships among Asteraceae species. These findings can have significant implications in the broader context of plant biodiversity research, molecular ecology, and taxonomy.

There are few limitations encountered during the current research that should be taken into consideration. For example, the study only used two DNA barcoding markers, *rbcL* and *matK*. While these markers are widely used for plant barcoding, they may not be sufficient for all plant species. It is possible that other markers or a combination of markers could provide more accurate results. Besides, the study was conducted in a specific geographic location, and the results may not be applicable to other regions with different environmental conditions or species composition.

Conclusions

To sum up, this study generated 22 new barcode sequences that were not previously found in the BOLD databases. DNA barcoding allows for distinguishing between closely related and distantly related species. The results show that the effectiveness of barcoding depends on factors such as the representation of taxonomically appropriate sequences in the reference database, the genetic distance between sequences in this database, the characteristics of the studied species, the specific barcode used, and the level of monogenic variation within a species.

Acknowledgements

Authors would like to express sincere gratitude to Prof. Mehboor-ur Rehman for helping in plant species identification.

References

Abid, R. and Qaiser. 2021. In: *Fl. Pakistan* 224: 138, Nom. Nov.

Amandita, F.Y., K. Rembold, B. Vornam, S. Rahayu, I.Z. Siregar, H. Kreft and R. Finkeldey. 2019. DNA barcoding of flowering plants in Sumatra, Indonesia. *Ecol. Evol*., 9(4): 1858-1868.

Asahina, H., J. Shinozaki, K. Masuda, Y. Morimitsu and M. Satake. 2010. Identification of medicinal *Dendrobium* species by phylogenetic analyses using *matK* and *rbcL* sequences. *J. Nat. Med*., 64: 133-138.

Bhargava, M. and A. Sharma. 2013. DNA barcoding in plants: evolution and applications of in silico approaches and resources. *Mol. Phylogen. Evol*., 67(3): 631-641.

Cabelin, V.L.D. and G.J.D. Alejandro. 2016. Efficiency of *matK*, *rbcL*, trnH-psbA, and trnL-F (cpDNA) to molecularly authenticate Philippine ethnomedicinal Apocynaceae through DNA barcoding. *Pharmacog. Mag.,* 12(Suppl 3): S384.

Chase, M.W. and M.F. Fay. 2009. Barcoding of plants and fungi. *Science,* 325(5941): 682-683.

Chen, S., H. Yao, J. Han, C. Liu, J. Song, L. Shi, Y. Zhu, X. Ma, T. Gao and X. Pang. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS One,* 5(1): e8613.

Dormontt, E.E., K.J. Van Dijk, K.L. Bell, E. Biffin, M.F. Breed, M. Byrne, S. Caddy-Retalic, F. Encinas-Viso, P.G. Nevill and A. Shapcott. 2018. Advancing DNA barcoding and metabarcoding applications for plants requires systematic analysis of herbarium collections an Australian perspective. *Front. Ecol. Evol*., 6 (Article No. 134):1-12.

Gao, T., H. Yao, J. Song, Y. Zhu, C. Liu and S. Chen. 2010. Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. *BMC Evol. Biol*., 10(1): 1-7.

Group, C.P. W., P.M. Hollingsworth, L.L. Forrest, J.L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M.W. Chase, R.S. Cowan and D.L. Erickson. 2009. A DNA barcode for land plants. *PNAS,* 106(31): 12794-12797.

Hollingsworth, P.M., D.Z. Li, M. van der Bank and A.D. Twyford. 2016. Telling plant species apart with DNA: from barcodes to genomes. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.,* 371(1702): 20150338.

Hollingsworth, P.M., S.W. Graham and D.P. Little. 2011. Choosing and using a plant DNA barcode. *PloS One,* 6(5): e19254.

Hosein, F.N., N. Austin, S. Maharaj, W. Johnson, L. Rostant, A.C. Ramdass and S.N. Rampersad. 2017. Utility of DNA barcoding to identify rare endemic vascular plant species in Trinidad. *Ecol. Evol.*, 7(18): 7311-7333.

Howard, C., C. Lockie-Williams and A. Slater. 2020. Applied barcoding: The practicalities of DNA testing for herbals. *Plants,* 9(Article No. 1150): 1-24**.**

Ismail, N.Z., H. Arsad, M.R. Samian, M.R. Hamdan and A.S. Othman. 2018. Assessment of three plastid DNA barcode markers for identification of *Clinacanthus nutans* (Acanthaceae). *3 Biotech*, 8: 1-8.

Kang, Y., Z. Deng, R. Zang and W. Long. 2017. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Sci. Rep*., 7(1): 1-9.

Kress, W.J. 2017. Plant DNA barcodes: Applications today and in the future. *J. Syst. Evol.,* 55(4): 291-307.

Kress, W.J. and D.L. Erickson. 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(6): e508.

Kress, W.J., D.L. Erickson, F.A. Jones, N.G. Swenson, R. Perez, O. Sanjur and E. Bermingham. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci.* USA, 106(44): 18621-18626.

Kumar, S., G. Stecher and K. Tamura. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, 33(7): 1870-1874.

Lahaye, R., M. Van der Bank, D. Bogarin, J. Warner, F. Pupulin, G. Gigot, O. Maurin, S. Duthoit, T. G. Barraclough and V. Savolainen. 2008. DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci.* USA, 105(8): 2923-2928.

Lassmann, T., and E.L. Sonnhammer. 2005. Kalign–an accurate and fast multiple sequence alignment algorithm. *BMC Bioinform*., 6(1): 1-9.

Li, J., S. Wang, J. Yu, L. Wang and S. Zhou. 2013. A modified CTAB protocol for plant DNA extraction. *Chin. Bull. Bot.,* 48(1): 72-78.

Mahadani, P., G.D. Sharma and S.K. Ghosh. 2013. Identification of ethnomedicinal plants (Rauvolfioideae: Apocynaceae) through DNA barcoding from northeast India. *Pharmacog. Mag.,* 9(35): 255-263.

Mahadani, P., M. Dasgupta, J. Vijayan, C. S. Kar and S. Ray. 2022. DNA Barcoding in Plants: Past, Present, and Future. In: (Eds.): Singh, R.L., S. Mondal, A. Parihar and P. K. Singh. Plant Genomics for Sustainable Agriculture. Springer, Singapore, 331-349.

Mahadani, P., S. Chakrabarti, R. Pal, S.K. Jain and A.M. Sevanthi. 2022. Application of DNA barcoding tool for authentication of traditionally important medicinal Dendrobium (Orchidaceae) species from Sikkim Himalaya. *S. Afr. J. Bot*., 146: 905-911.

Newmaster, S.G., M. Grguric, D. Shanmughanandhan, S. Ramalingam and S. Ragupathy. 2013. DNA barcoding detects contamination and substitution in North American herbal products. *BMC Med.*, 11: 1-13.

Panero, J.L. and B.S. Crozier. 2016. Macroevolutionary dynamics in the early diversification of Asteraceae. *Mol. Phylogen. Evol*., 99: 116-132.

Pang, X., J. Song, Y. Zhu, H. Xu, L. Huang and S. Chen. 2011. Applying plant DNA barcodes for Rosaceae species identification. *Cladistics*, 27(2): 165-170.

Sarangzai, A.M., A. Ahmed and S.K. Laghari. 2013. Traditional uses of some useful medicinal plants of Ziarat District Balochistan, Pakistan. *FUUAST J. Biol.*, 3(1): 101-107.

Song, J., H. Yao, Y. Li, X. Li, Y. Lin, C. Liu, J. Han, C. Xie and S. Chen. 2009. Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. *J. Ethnopharm*., 124(3): 434-439.

Tareen, R.B., T. Bibi, M.A. Khan, M. Ahmad, M. Zafar and S. Hina. 2010. Indigenous knowledge of folk medicine by the women of Kalat and Khuzdar regions of Balochistan, Pakistan. *Pak. J. Bot*., 42(3): 1465-1485.

TV, G.I.F. version 140 Geospiza. *Inc, Seattle, WA*.

Ullah, S., Z. Khan and M. Ashfaq. 2016. Identification of the grass family (Poaceae) by using the plant dna barcodes rbcl and matK. *J. Biodiv. Environ. Sci*., 08: 175-186.

Wattoo, J.I., M.Z. Saleem, M.S. Shahzad, A. Arif, A. Hameed and M.A. Saleem. 2016. DNA Barcoding: Amplification and sequence analysis of *rbcL* and *matK* genome regions in three divergent plant species. *Adv. Life Sci.,* 4(1): 03-07.

Zhang, C.Y., F.Y. Wang, H.F. Yan, G. Hao, C.M. Hu and X.J. Ge. 2012. Testing DNA barcoding in closely related groups of *Lysimachia* L.(Myrsinaceae). *Mol. Ecol. Resour.,* 12(1): 98-108.

(Received for publication 22 February 2023)