MOLECULAR CHARACTERIZATION OF *AMYGDALUSSPINOSISSIMA* COLLECTED FROM THE JUNIPER ECOSYSTEM OF BALOCHISTAN

RASHID ABDULLAH[1], SHAHJAHAN SHABBIR AHMED RANA*[1], IFTEKHAR AHMED BALOCH[3], NAZEER AHMED[1], IMRAN ALI SANI[1], SAADULLAH KHAN[1], DAWOOD SHAHID[1] AND SAADULLAH LEGHARI[2]

[1]Department of Biotechnology, Faculty of Life Sciences, Balochistan University of Information Technology Engineering and Management Sciences (BUITEMS), Quetta Pakistan.
[2]Department of Botany, University of Balochistan, Pakistan.
[3]Department of Botany, University of Turbat, Pakistan
**Corresponding author email: *Shahjahan.shabbir@buitms.edu.pk*

خلاصہ

امیگدالس اسپینوسیما (جنگلی بادام) فرینچ اور اسی فیملی سے تعلق رکھتا ہے جو بلوچستان کے جونیپر ایکٹو سسٹم میں پائے جانے والے ادویات بنانے والے پودوں میں شمار ہوتا ہے ڈین این اے (DNA) بار کوڈنگ کی تکنیک نے متعدد جاندار اقسام کی نشاندہی کرنے کیلئے مولیکیولر کردار کا مطالعہ کیا ہے کیونکہ یہ ایک قابل اعتماد طریقہ ہے اس ضمن میں دو بار کوڈنگ پرائمرز Matk اور RbcL کی تشخیص کی گئی ہے۔ امیگدالس اسپینوسیما کی شناخت کرنے کے لئے (DNA bases) بیس میں تبدیلی کا تعین، اس تبدیلی کی شناخت اور فائیلوجینیٹک تجزیہ میں نیبر جوئینگ (NJ) کا طریقہ اور کائی میر ۲ K2P کا ماڈل استعمال کیا گیا ہے۔

ہمارے مطالعے سے انکشاف ہوا ہے کہ 3 خاص تبدیلیوں RbcL اور ایک خاص تبدیلی Matk کے نمونوں میں آئی ہے DNA bases کی ترتیب (Sequences alignment) اور فائیلوجینیٹک تجزیہ کے مطالعے سے ہم اس نتیجے پر پہنچے ہیں کہ امیگدالس اسپینوسیما کی شناخت کرنے کے لئے RbcL Matk, کے مقابلے میں زیادہ قابل اعتماد بار کوڈ پرائمر ہے۔

Abstract

Amygdalus spinosissima (Bunge) French. belongs to the family Rosaceae and is considered among the medicinal plants found in the juniper ecosystem of Balochistan. The technique of DNA barcoding was implied to identify several species, as it offers a reliable tool for molecular characterization. In this study two barcoding markers rbcL and matK were evaluated for the identification of *A. spinosissima*. The Determination of base substitutions, assessment of identified substitutions, and construction of phylogenetic trees using the Neighbor-Joining (NJ) method under the K2P distance model were carried out. Our study revealed 3 unique substitutions in rbcL samples and a single unique base substitution in matK samples. Based on phylogenetic analysis and sequence alignment we reportrbcL to be a more significant barcode maker as compared to matK for the identification of *A. spinosissima*.

Keywords: *Amygdalus spinosissima*, DNA barcoding, matK, rbcL, Medicinal plants, Neighbor Joining, Juniper ecosystem, Phylogenetic analysis.

Introduction

The province of Balochistan has a unique juniper ecosystem found in its Ziarat district and its juniper forest is also referred to as "The Living Fossils" because of their long life span of >3000 years and slow growth (Achakzai *et al*., 2013). It is found in the studies that 141,000 hectors of Balochistan are covered by the Juniper forest (Sarangzai *et al*., 2010). The Juniper forests are also found in the Districts of Kalat and Loralai, in the Zarghoon region near Quetta, and in the Herboi region of Hernai. There are also several varieties of medicinal plants found in that area, most of which are still not molecularly characterized.

Based on the taxonomic classification *Amygdalus spinosissima* (Bunge) French. (wild/thorny almond) belongs to the genus Prunus, its subgenus is *Amugdalus* which belongs to the family of Rosaceae, and its subfamily and tribe are Amygdaloideae and Amygdaleae, respectively. It is found in the areas which are inhabited by mountains and have cold and dry weather (Vadafar *et al*., 2010). This plant grows on the stony slopes of mountains or valleys. The maximum height of the plant is found to be 3 to 4 meters, it has pink flowers with rounded petals, the leaves are curved almost forming a "V-shaped" structure and are green in color. The flowering season of the plant is in April and May. The stem is grayish-white in color and has long spines (Dzhangaliev *et al*., 2003).

The technique of DNA barcoding implies the use of short genomic portions (barcodes) that are standardized and help in the total detection of species, which is very helpful in cases where the detection of species is difficult

on the basis of morphological data. It is reported in the study conducted by Hebert *et al*., (2003) that cytochrome c oxidase 1 (CO1); a mitochondrial gene to be the barcode for animals, while for plants a mixture of two-locus; chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene (rbcL) and maturase K gene (matK) are suggested by Consortium for the Barcode Of Life (CBOL) as a core barcode markers (Schindel and Miller 2005).

There is no detailed study conducted on the molecular characterization of *Amygdalus spinosissima*, therefore in the present study two DNA barcoding regions rbcL and matK were utilized as a tool to assess their efficiency and reliability for the identification purpose. These two regions were also used for the identification of genus *Suaeda* (amaranthaceae) species by (Munir *et al*., 2015).

Materials and Methods

Plant Material

The specimen of *Amygdalus spinosissima* was collected from three different locations covering two districts of Balochistan; Quetta (Shaban area, Zarghoon range) and Ziarat (Kawas). The samples were collected randomly and instantly immersed in liquid nitrogen. The list of samples along with the name of species and barcoding marker is given in Table 1.

DNA Extraction

Fresh leaf samples of *Amygdalus spinosissima* were used to extract the total genomic DNA by combining two reported DNA extraction protocols (Lodhi *et al*., 1994) and (Porebski *et al*., 1997). 5 microliters of 1% mercaptoethanol solution was poured in each tube along with 300ul of DNA extraction buffer, and it was then ground with metal pestle followed by the addition of 30 micro-liters of 6% PVP solution and 165ul micro-liter of 7.5 molar solution of ammonium acetate (pH 5.2). Prior to the final washing with 90% ethanol, the genomic DNA was washed with a 24:1 solution of chloroform and isoamyl. The DNA pellet was air-dried and was dissolved in 30ul of T.E. buffer.

PCR Amplification

For PCR matK region was amplified using primer pair (forward):5'CGATCTATTCATTCAATATTTC3', (reverse):5'GTAAAATCAAGTCCACCRCG-3' (Asahina *et al*., 2010), and the rbcL region was amplified using primer pair (forward): 5'-ATGTCACCACAAACAGAGACTAAAGC-3' (reverse): 5'-GTAAAATCAAGTCCACCRCG-3' (Munir *et al*., 2015). The amplification was performed in 20ul reaction volume containing 10ul of PCR mix (Thermo Fisher Scientific, USA), total 1ul of both forward and reverse primers, 1ul of genomic DNA (25ng), and the rest was adjusted with 8ul of PCR water. For matK region, thermo-cycles were optimized as follows: Initial denaturation for 5min at 94^{o}C, followed with 35 cycles at 52oC for 45sec, 1min at 72^{o}C, with a final extension for 5min at 72^{o}C. For the rbcL region, the annealing temperature was 61.5^{o}C for 45sec followed with 30 cycles. Other conditions were same as for matK.

Gel Electrophoresis

For the confirmation of genomic DNA 1%, agarose gel was prepared in 50ml T.B.E buffer by adding 0.5g of a garose gel and the PCR products were examined using 2% a garose gel, prepared by adding 1g of agarose in 50ml T.B.E buffer at 90 volts for 30 min. The results were examined using Ultra Violet Visualization system (Cleaver Scientific Ltd, UK.) The size of PCR products was evaluated using 100bp DNA ladder (Thermo Fisher Scientific, USA).

Sequencing and Post-sequencing Analysis

The PCR products were sent to the commercial lab (Macrogen, Korea) for sequencing. For matK and rbcL products, the sequencing was performed using the same primer pairs (*i.e.* matK and rbcL). The alignment of each DNA region was performed by the CLUSTALW tool using Bioedit V.7.0 (Hall, 2004). The pairwise genetic distances for both rbcL and matK were calculated using MEGA V.5.0 (Munir *et al*., 2015). The phylogenetic tree was constructed using the Neighbor-Joining (NJ) method under the Kimura 2-Parameter (K2P) distance model in MEGA V.5.0 (Tamura *et al*., 2007). To check the quality of the nucleotide chromatogram,Finch TV DNA analysis program was used (Kusumanjali *et al*., 2012).

Statistical Analysis

SPSS version 22 was implied to calculate central tendency and variance analysis. All means were spread on one-way analysis of variance at α 0.05 significance level.

Results and Discussion

The current study showed that DNA extraction by using standard CTAB protocol failed to yield successful results as compared to the combination of two separate manual DNA isolation protocols which produced positive results for genomic DNA and good quality PCR product. The product sizes of rbcL and matK region were found to be 700bp and 800bp in length respectively. All the obtained products were successfully sequenced. For a comparison, DNA based genetic analysis multiple sequence alignment (MSA) was performed

among collected sequences and reference sequence of *P.spinosissima* for both rbcL (Accession No.HQ235564) and matK (Accession No.HQ235281) (Fig.1 and Fig.2) The CLUSTALW based alignment revealed 3 unique base substitutions among all the collected samples of rbcL (Fig.2) and 1 unique base substitution among the samples of matK (Fig.1). The regions of identified substitutions were compared with 15 species (*Prunus dulcis, Prunus kansuensis, Prunus persica, Prunus tangutica, Prunus kuramica, Prunus davidiana, Prunus armeniaca, Prunus argentea, Prunus mume, Prunus pumila, Prunus japonica, Prunus geniculata, Prunus consociiflora, Prunus mandshurica,* and *Prunus alleghaniensis*) that appeared in the BLAST results and shared similar regions, however, the 3 substitutions of rbcL samples stand unique (Fig.2) while the substitutions of matK samples were found to be identical in 4 species (*Prunus dulcis, Prunus kansuensis, Prunus persica,* and *Prunus mume*).

The identified substitutions were checked for their accuracy by evaluating the particular peaks of chromatogram. Among all the collected samples of rbcL and matK and showed Q-values above the standard threshold of Q20 (Table.4 and Table.5). The minimum and maximum sequence divergence among the samples of *A.spinosissima* sequenced with rbcL marker and collected from different geographical locations varied from 0.011% to 0.027% respectively. The samples sequenced with matK showed a minimum of 0.031% to a maximum 0.082% of divergence. The phylogenetic tree constructed for rbcL region (Fig. 3) using its reference sequence (Accession No. HQ235564) as a root showed that the samples rbcL_SA2 (Shaban) and rbcL_KA1 (kawas) shared a common ancestor and the other sample rbcL_SA1 (shaban) is closely related to rbcL_SA2 (Shaban). The other samples rbcL_KA2 (kawas), rbcL_Za4 (ziarat) and rbcL_Z2 (ziarat) shared a clade with each other. This tree was different from that of matK region (Fig. 4) where matK_3 (kawas) and matK_1(ziarat) shared common ancestors while matK_K (kawas) was closely related to the reference sequence of *P.spinosissima* (Accession No. HQ235281).

The wild almond or *Amygdalus spinosissima* which is known for its medicinal use by the indigenous population of Balochistan (Pakistan) grows 400 to 1500m above sea level. The standard CTAB protocol by Krishna *et al*., (2012) was utilized for the DNA Isolation and it failed to produce successful results, that might be due to the fact that there was no use of β-mercaptoethanol in the protocol which serves as an antioxidant agent.

By the combination of two separate DNA extraction protocols Lodhi *et al*., (1994) and Porebski *et al*., (1997) positive results of genomic DNA and good quality PCR product was obtained. Attempt for the molecular characterization of subgenus Prunus by Wen *et al*., (2008) using Internal Transcribed Spacer ITS sequences of nuclear ribosomal DNA revealed the subgenus Amygdalus to be within the subgenus Prunus, and the relationship between the 2 sampled taxa of the subgenus remained ambiguous. Ribulose-bisphosphate carboxylase (rbcL) is the gene of a large subunit located on the chloroplast genome and is considered as an appropriate choice for the inference of phylogenetic studies at higher taxonomic levels (Mondal *et al*., 2013). It was utilized along with Maturase-K gene (Mat K) which also showed high levels of discrimination among angiosperm species when it comes to plastid coding regions (Hollingsworth *et al*., 2009).

In this study identification of 3 unique substitutions by rbcL barcoding marker suggest it to be a strong candidate for DNA barcoding-based identification of species. It comes in accordance with the study by Munir *et al*., (2015) on genus *Suaedas* suggesting rbcL as a potential source of barcoding based identification. The identification of a total of 4 unique substitutions in this study using rbcL and matK agrees with the study by Huang *et al*., (2014) on germplasm identification of *Amomum villosum* based on DNA barcoding, where they concluded that the combination of rbcL and matK overall provided not more than one to four substitutions, suggesting that their resolution is limited, and provide only a few polymorphisms to differentiate individuals within species.

Table 1. List of samples used in this study with their codes and area of collection

S.No.	Sample ID	Species Name	Collection Area	Barcoding Marker
1.	MatK_A1	*Amygdalus spinosissima*	Shaban	*matK*
2.	MatK_A2		Shaban	
3.	MatK_K		Kawas	
4.	MatK_3		Kawas	
5.	MatK_Z		Ziarat	
6.	MatK_1		Ziarat	
7.	rbcL_SA1	*Amygdalus spinosissima*	Shaban	*rbcL*
8.	rbcL_SA2		Shaban	
9.	rbcL_KA1		Kawas	
10.	rbcL_KA2		Kawas	
11.	rbcL_Za4		Ziarat	
12.	rbcL_Z		Ziarat	

Table 2. Pairwise genetic distances in *matK* barcoding region

	MatK_A1	MatK_A2	MatK_K	MatK_3	MatK_Z	MatK_1
MatK_A1	-	-	-	-	-	-
MatK_A2	0.044	-	-	-	-	-
MatK_K	0.034	0.036	-	-	-	-
MatK_3	0.076	0.082	0.061	-	-	-
MatK_Z	0.050	0.042	0.031	0.053	-	-
MatK_1	0.058	0.050	0.044	0.039	0.034	-

Table 3. Pairwise genetic distances in *rbcL*barcoding region

	rbcL_SA1	rbcL_SA2	rbcL_KA1	rbcL_KA2	rbcL_ZA4	rbcL-Z2
rbcL_SA1	-	-	-	-	-	-
rbcL_SA2	0.011	-	-	-	-	-
rbcL_KA1	0.014	0.018	-	-	-	-
rbcL_KA2	0.016	0.018	0.023	-	-	-
rbcL_ZA4	0.014	0.018	0.021	0.014	-	-
rbcL_Z2	0.020	0.021	0.027	0.020	0.014	-

Table 4. Finch TVchromatogram analysis based on Q value of unique substitutions identified in *matk*barcodingregion.

Sample Name	**Sample ID**	**Q Value**	**Position**	**Change**
A.spinosissima (Shaban)	MatK_A1 MatK_A2	30 40	433 433	A>T A>T
A.spinosissima (Kawas)	MatK_K MatK_3	52 30	430 436	A>T A>T
A.spinosissima (Ziarat)	MatK_Z MatK_1	30 30	436 434	A>T A>T

Table 5. Finch TV chromatogram analysis based on Q value of 3 unique substitutions identified in *rbcL* barcodingregion

Sample Name	**Sample ID**	**Q Value**	**Position**	**Change**
A.spinosissima (Shaban)	SA1	55 52 45	120 397 555	A>G C>A T>C
A.spinosissima (Shaban)	SA2	55 55 52	120 397 555	A>G C>A T>C
A.spinosissima (Kawas)	KA1	59 62 58	122 399 557	A>G C>A T>C
A.spinosissima (Kawas)	KA2	55 55 45	117 394 554	A>G C>A T>C
A.spinosissima (Ziarat)	Z2	55 55 55	119 396 554	A>G C>A T>C
A.spinosissima (Ziarat)	Za4	55 55 52	124 401 559	A>G C>A T>C

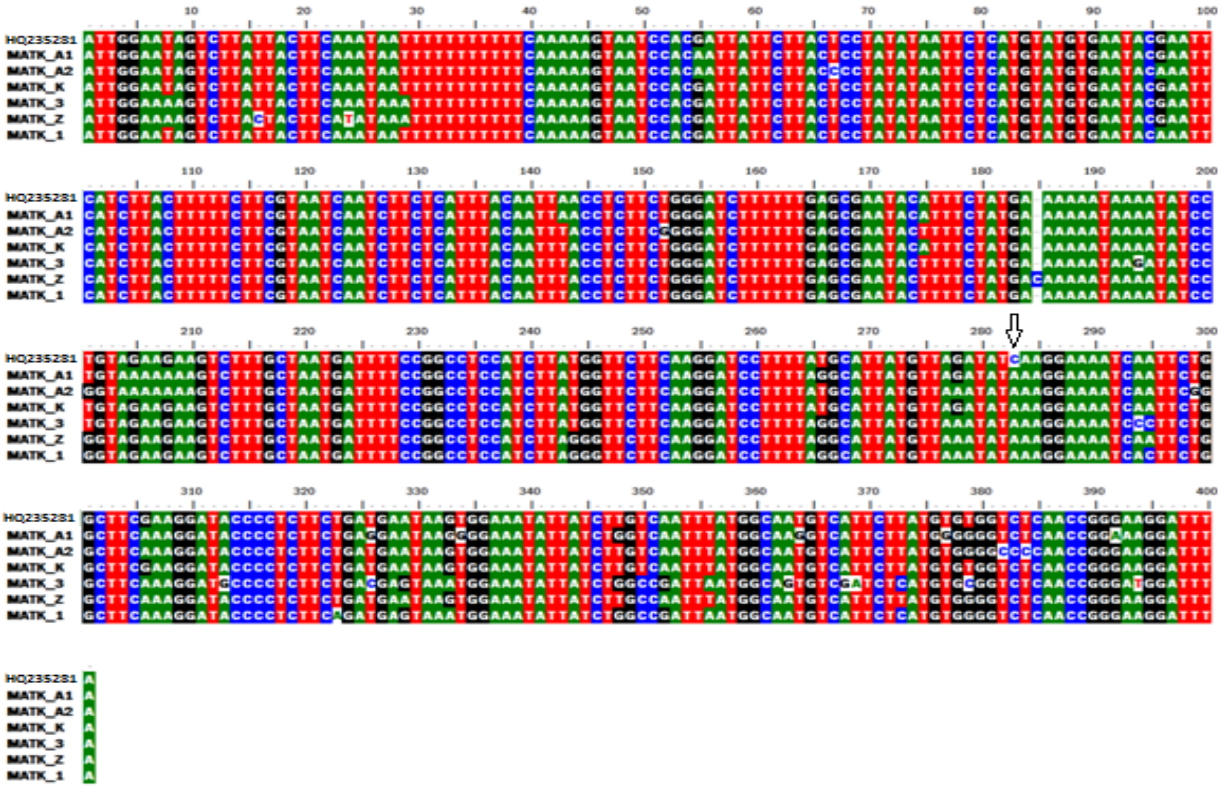

Fig.1. Result of multiple sequence alignment with the *P.spinosissima matK*region(Accession No.HQ235281)on the top followed by the sequences of *Amygdalus spinosissima* samples. The arrow is poniting out the region of unique substitutions.

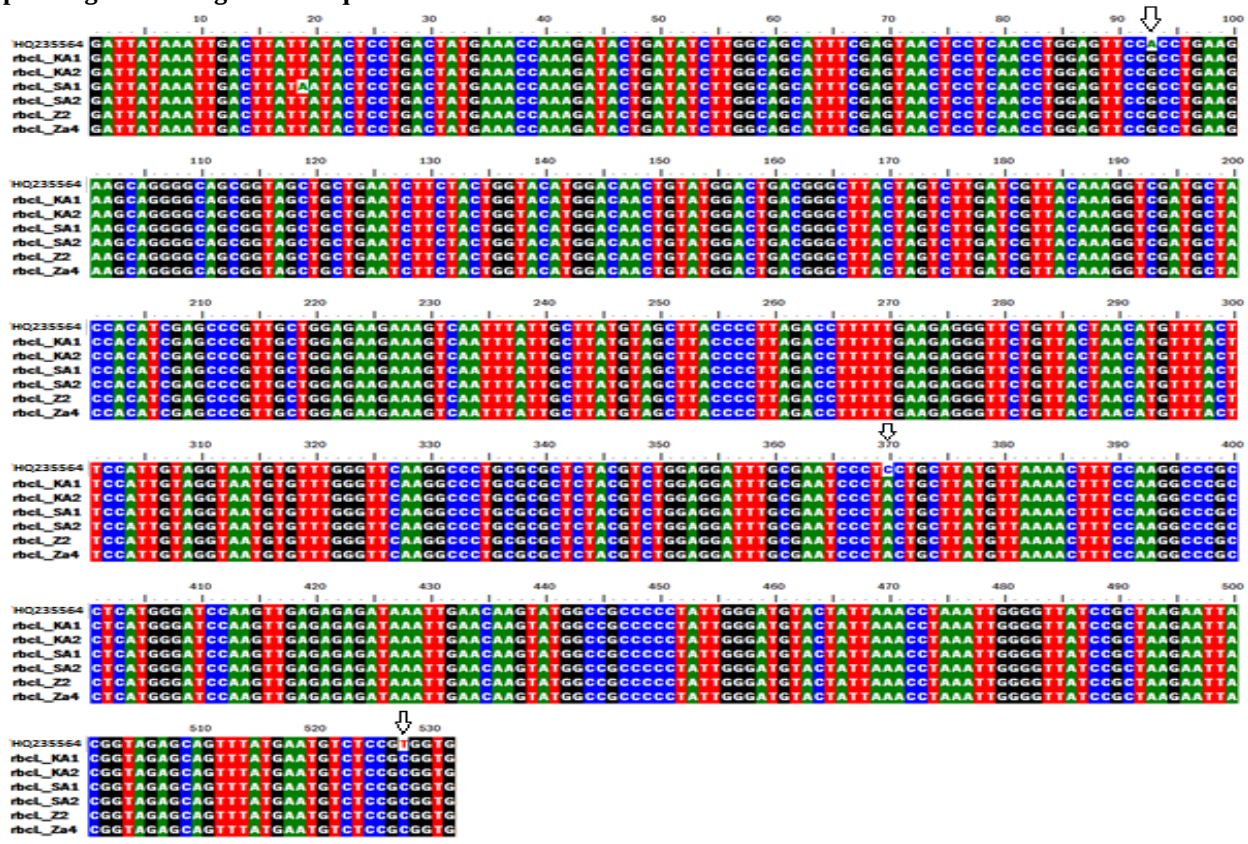

Fig.2. Result of multiple sequence alignment with the *P.spinosissimarbcL*region (Accession No.HQ235564) on the top followed by the sequences of *Amygdalus spinosissima* samples. The arrow is pointing out the region of unique substitutions.

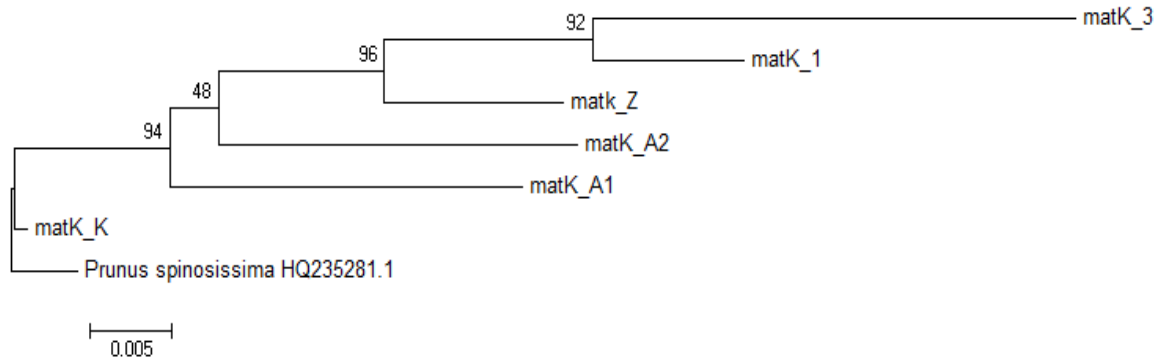

Fig.3. NJ tree based on *matK* sequences along with bootstrap values on relevant branches

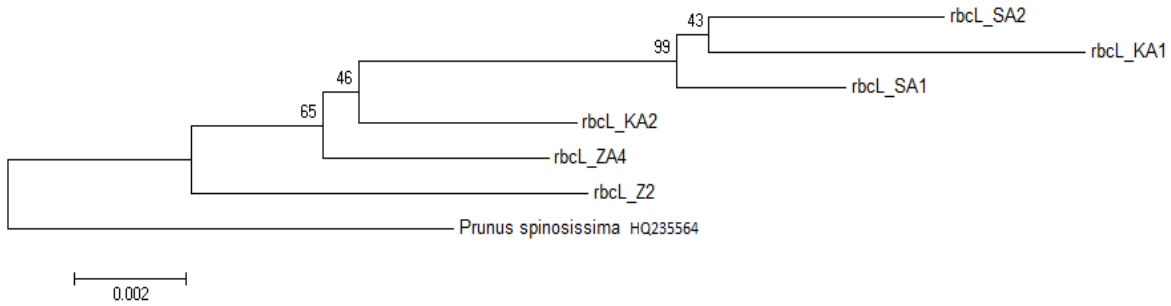

Fig.4. NJ tree based on *rbcL* sequences along with bootstrap values on relevant branches

The barcoding markers used in this study (rbcL and matK) have been recommended as core plant barcodes by Consortium for Barcode of Life (CBOL) Schindel and Miller (2005) however according to Munir *et al*., (2015) the efficiency of rbcL alone is greater (85%) than that of rbcL and matK combined (75.0%).

Neighbor-Joining method (NJ) was utilized with Kimura two-parameter (K2P)for phylogenetic tree construction. (Tamura *et al*., 2007). NJ method was also utilized by Khan *et al*., (2013) for the identification of endemic plant species based on DNA barcoding. However, in this study high level of divergence was found among the species collected from the same regions, in the tree generated by both rbcL and matK sequences. This ambiguity in the relationship can be clarified by further investigation along with the increase in sample size and the use of more molecular markers.

Conclusion

Based on our findings, it is concluded that to clarify the ambiguity in the molecular status of *Amygdalus spinosissima*, the use of rbcL and other barcoding markers are suggested along with the increase in sample size for a better phylogenetic analysis. Molecular analysis of various unreported medicinal plants exit in the Juniper ecosystem of Baluchistan would open wide range of scientific data significantly important for the pharma industry of Pakistan.

Acknowledgement

We acknowledge Office of Research, Innovation and Commercialization (ORIC); BUITEMS, for providing financial support. The authors also declare no conflict of interest in the conducted study.

References

Achakzai, A. K. K., Batool, H., Aqeel, T. and Bazai, Z.A. (2013) A comparative study of the deforestation and regeneration status of Ziarat Juniper forest. *Pak J Bot* 45 (4):1169-1172

Asahina, H., Shinozaki, J., Masuda, K., Morimitsu, Y. and Satake, M. (2010). Identification of medicinal Dendrobium species by phylogenetic analyses using matK and rbcL sequences. *Journal of Natural Medicines* 64 (2):133-138

Dzhangaliev, A., Salova, T. and Turekhanova, P. (2003). The wild fruit and nut plants of Kazakhstan. *HORTICULTURAL REVIEWS-WESTPORT THEN NEW YORK*- 29:305-372

Hall, T. (2004) BioEdit version 7.0. 0. Distributed by the author, website: www mbio ncsu edu/BioEdit/bioedit html

Hebert, P.D., Cywinska, A. and Ball, S.L. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B: *Biological Sciences* 270 (1512):313-321

Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M.W., Cowan, R.S., Erickson, D.L. and Fazekas, A.J. (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* 106 (31):12794-12797

Huang, Q., Duan, Z., Yang, J., Ma, X., Zhan, R., Xu, H. and Chen, W. (2014). SNP Typing for Germplasm Identification of Amomum villosum Lour. Based on DNA Barcoding Markers. *PloS one* 9 (12):e114940

Khan, S., Al-Qurainy, F., Nadeem, M. and Tarroum, M. (2013) Selection of chloroplast DNA markers for the development of DNA barcode and reconstruction of phylogeny of senecio asirensis boulos and jri wood. *Pak J Bot* 45 (2):703-710

Krishna, R.B., Reddy, S.R.R., Javangula, H., Swapna, D. and Reddy, K.J. (2012) An easy and simple method of isolation and purification of genomic DNA from the leaves of Gymnema sylvestre an anti-diabetic plant. *Int J Life Sci & Pharma* Res 2 (1):15-20

Kusumanjali, K., Kumari, G., Srivastava, P. and Das, S. (2012). Sequence conservation and divergence in miR164C1 and its target, CUC1, in Brassica species. *Plant biotechnology reports* 6 (2):149-163

Lodhi, M.A., Ye, G-N., Weeden, NF. and Reisch, B.I. (1994). A simple and efficient method for DNA extraction from grapevine cultivars andVitis species. *Plant Molecular Biology Reporter* 12 (1):6-13

Mondal, S.K., Shit, S. and Kundu, S. (2013) A comparative computational study of the'rbcL' gene in plants and in the three prokaryotic families--Archaea, cyanobacteria and proteobacteria. *Indian Journal of Biotechnology* 12 (1):58-66

Munir, U., Perveen, A. and Qamarunnisa, S. (2015). THE UTILITY OF RBCL AND MATK REGIONS FOR DNA BARCODING ANALYSIS OF THE GENUS SUAEDA (AMARANTHACEAE) SPECIES. *PAKISTAN JOURNAL OF BOTANY* 47 (6):2329-2334

Porebski, S., Bailey, L.G. and Baum, B.R. (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant molecular biology reporter* 15 (1):8-15

Sarangzai, A.M., Khan, N., Wahab, M. and Kakar, A. (2010) New spread of dwarf mistletoe (Arceuthobium oxycedri) in Juniper forests, Ziarat, Balochistan, *Pakistan. Pak J Bot* 42 (6):3709-3714

Schindel, D.E. and Miller, S.E. (2005) DNA barcoding a useful tool for taxonomists. Nature 435 (7038):17-17

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution* 24 (8):1596-1599

Vadafar, M., Attar, F. and Maroofi, H. (2010) Trichome micromorphology in drupe of Amygdalus L.(Rosaceae) from Iran. Acta Botanica Croatica 69 (1):93-105

Wen, J., Berggren, S.T., Lee, C-H., Ickert-Bond, S., Yi, T-S., Yoo, K-O., Xie, L., Shaw, J. and Potter, D. (2008) Phylogenetic inferences in Prunus (Rosaceae) using chloroplast ndhF and nuclear ribosomal ITS sequences.