THE UTILITY OF *RBCL* AND *MATK* REGIONS FOR DNA BARCODING ANALYSIS OF THE GENUS *SUAEDA* (AMARANTHACEAE) SPECIES

UZMA MUNIR[1], ANJUM PERVEEN[2] AND SYEDA QAMARUNNISA[3*]

[1]*Department of Botany, University of Karachi, Pakistan*
[2]*Centre for Plant Conservation, University of Karachi, Pakistan*
[3]*The Karachi Institute of Biotechnology and Genetic Engineering (KIBGE), University of Karachi, Pakistan.*
Corresponding author e-mail: sqnisa@uok.edu.pk

Abstract

The genus *Suaeda* (Forssk.) belongs to the family Chenopodiaceae. Identification of *Suaeda* species based on morphological data is quite difficult due to high phenotypic plasticity, few distinguishable and many overlapping characters. In current research, the efficiency of *rbcL* and *matK* (plants core barcode regions) for species identification of the genus *Suaeda* was assessed. The determination of intraspecific and interspecific divergence, assessment of barcoding gap, reconstruction of phylogenetic trees and evaluation of barcode regions for species identification (based on best match and best close match) were carried out.. The results revealed that *rbcL* showed comparatively less overlapping for the distribution of interspecific and intraspecific divergence. In addition, the highest discriminating ability for correct species identification was also observed in this region. Therefore, *rbcL* was found to be a significant barcode region for the identification of *Suaeda* species.

Key words: DNA barcoding, Species identification, *rbcL*, *matK*, Intraspecific, Interspecific, Monophyly.

Introduction

Suaeda (Forssk.) is an important halophytic genus which belongs to the family Chenopodiaceae and it comprises of c. 100 species which are cosmopolitan in distribution (Schenk & Ferren, 2001). Species of the genus *Suaeda* are difficult to identify because of the variability of phenotypic characters such as leaf shape, size, color and branching pattern of the stem.

The genus *Suaeda* has high rate of speciation and possesses very few distinguishing characters (Freitag *et al.,* 2001; Schutze *et al.,* 2003). A molecular based technique, DNA barcoding was introduced by Paul Hebert and his co-workers (2003) for the rapid species identification by comparing sequences of short standardized DNA marker with sequences of reference database. Selection of a barcode region involves standardization, which includes a number of criteria such as a barcode region should present in a group of interest, a barcode region should possess invariability of sequence in all individuals of the same species and enough variety of sequences between closely related species. A barcode region should also be short enough in size for ease of DNA extraction and sequencing by using single universal primer set (CBOL Plant Working Group., 2009; Shneyer, 2009).

A mitochondrial gene *Cox1* (cytochrome c oxidase subunit 1) has been proved as universal barcode for animals (Hebert *et al.,* 2003). However, *Cox1* is not appropriate as plant barcode region due to low evolutionary and the high rearrangement rate of plant mitochondrial genome (Wolfe *et al.,* 1987; Palmer & Herbon, 1988). For land plants, searching for a core plant barcode region has proved to be more difficult. Many recommendations have been given by different plant DNA barcode working groups; such as ITS and *trnH-psbA* (Kress *et al.,* 2005), *rbcL* by (Chase *et al.,* 2005) and *matK* and *trnH-psbA* (Newmaster *et al.,* 2008). In 2009 the CBOL (Consortium for the Barcode Of Life) executive committee has assigned a combination of *rbcL* and *matK* as plant core barcode regions and the use of *trnH-psbA* and ITS as supplementary barcodes.

A comparative barcode analysis for a large data set (6,286 individuals of 1,757 species represented by 141 genera, distributed in 75 families and 42 orders) of Angiosperm, in which only a few *Suaeda* species were included (Li *et al.,* 2011). There is no detailed report available on the barcode of *Suaeda* species, therefore, in present study *rbcL* with *matK* regions were used to assess the efficiency and suitability for identification purpose. These two regions were also used for identification of medicinal plants of Pakistan (Shinwari *et al.,* 1994a; 1994b; Shinwari, 1995; Schori & Showalter, 2011, Shinwari *et al.,* 2014).

Materials and Methods

Plant material: In present research a total of 20 representatives of the genus *Suaeda* (Chenopodiaceae) were included. To obtain the sequences of *rbcL* and *matK* regions, *Suaeda fruticosa* (Forssk. ex J. F. Gmel.)*, Suaeda monoica* (Forssk. ex J. F. Gmel.) and *Suaeda acuminata* (C. A. Mey.) Moq.,were used and their herbarium sheets were deposited at the Karachi University Herbarium (Centre for Plant Conservation). The Sequences of *Suaeda vera* (J. F. Gmelin), *Suaeda maritime* (C.A. Mey.) Moq. (Fig. 3) and *Suaeda glauca* (Bung.) Bung., were retrieved form GenBank. The list of GenBank accession numbers is given in Table 1. For phylogenetic analysis *Bienertia cycloptera* (Bienerteae) was selected as an outgroup.

DNA extraction, amplification and sequencing: Fresh and herbarium leaf samples were used to extract the total genomic DNA by using a CTAB DNA extraction method with some modifications (Doyle & Doyle, 1987). The *rbcL* region was amplified by using primer pair (forward): 5'-ATGTCACCACAAACAGAGACTAAAGC-3' and (reverse): 5'-GTAAAATCAAGTCCACCRCG-3' (Kress & Erickson, 2007). The *matK* region was amplified by using primer pair (forward): 5'-CGTACAGTACTTTTGTGTTTACGAG-3' (reverse): ACCCAGTCCATCTGGAAATCTTGGTTC-3'

Amplification was performed in total 25 µL reaction volume containing 1 x PCR buffer, 2.5 mM (for *rbcL*) and 2.0 mM (for *matK*) of MgCL$_2$, 0.4 mM of dNTPs, 0.5 µM of each primer, 1.0 unit of *Taq* DNA polymerase (Bioneer, Korea), deionized H$_2$O and 50 ng of genomic DNA. For *rbcL* region thermo cycles were programmed as follows: Initial template denaturation at 94°C for 5 min was followed by 35 cycles of 30 sec at 94°C, 30 sec at 54°C and at 72°C for 1 min, with a final extension of 10 min at 72°C. For *matK* the annealing temperature was 52°C, other conditions were same as for *rbcL*. The PCR products were purified by using a PCR product purification kit (Bioneer, Korea) according to the manufacturer's protocol and were sent to the commercial lab (Bioneer, Korea) for sequencing.

Data analysis for DNA barcoding: Forward and reverse sequences were aligned by using the software Multalin (Corpet, 1988) and submitted to GenBank for their accession numbers. The sequences were aligned by using software ClustalW (Thompson *et al.,* 1994). For each Barcode region intraspecific and interspecific divergence was estimated by calculating K2P distances in MEGA v.5.0 (Tamura *et al.,* 2007). Barcoding gap (distance between intraspecific and interspecific variation) was represented graphically according to Meyer & Paulay (2005). The significance between intraspecific and interspecific variation were examined by Wilcoxon signed rank test in SPSS 16.0. In order to evaluate the success rate of each barcoding marker, "Best match" and "Best close match" criteria were employed by using the software Taxon DNA (Meier *et al.,* 2006). To test the monophyletic relationship between species, the Neighbor-Joining (NJ) method under the K2P distance model was used in MEGA 5.0 (Tamura *et al.,* 2007). Branch

statistical support was calculated by using 1000 bootstrap values (Felsenstain, 1985).

Results

Intraspecific and interspecific divergence analysis: In current research the higher intraspecific and interspecific divergence was recorded for *matK* (0.0153 and 0.0502) than *rbcL* region (0.0054 and 0.0217). The variability of intraspecific and interspecific divergence for both assessed markers is summarized in Table 2.

Assessment of the significance of barcoding markers: Wilcoxon signed rank tests of the two DNA regions showed that at interspecific level, significant *P*value (0.003) was recorded between the two investigated chloroplast DNA loci (Table 3). Whereas, at intraspecific level non-significant difference (*P*=0.463) was observed for *rbcL* and *matK* regions (Table 4).

Assessment of DNA barcoding gap: The presence of the DNA Barcoding gap was assessed by plotting the distribution of intraspecific and interspecific divergence with the interval of 0.004. Present data revealed that *rbcL* is a region where comparatively less overlapping was examined between the two parameters (Fig. 1). On the other hand, *matK* exhibited large overlapping between intraspecific and interspecific divergence (Fig. 2).

Efficiency of markers for species discrimination: The species discriminatory power was recorded 85.0% and 70.0% for *rbcL* and 65.0% and 60.0% for *matK* region by applying best match and best close match criteria respectively. However, *rbcL*+*mat*K being two marker combination with (75.0% and 72.5%) correct species values (Table 5).

Table 1. List of samples used in this study with their accession numbers.

S. No.	Name of species	Voucher numbers / sample Id	GenBank accession numbers	
			rbcL	*matK*
1.	*Suaeda acuminata*	G.H.No: 75622	JX985732.1	KF679793.1
2.	*Suaeda acuminata*	Z456	JF944511.1	JF956549.1
3.	*Suaeda acuminata*	Z457	JF944510.1	JF956548.1
4.	*Suaeda acuminata*	Z458	JF944509.1	JF956547.1
5.	*Suaeda fruticosa*	G.H.No: 86472	JX985731.1	JX985733.1
6.	*Suaeda fruticosa*	G.H.No: 86539	KF679785.1	KF679791.1
7.	*Suaeda fruticosa*	G.H.No: 86540	KF679786.1	KF679792.1
8.	*Suaeda monoica*	G.H.No: 86471	KF679782.1	KF860862.1
9.	*Suaeda monoica*	G.H.No: 86541	KF679788.1	KF860863.1
10.	*Suaeda monoica*	G.H.No: 86542	KF500487.1	KF860864.1
11.	*Suaeda glauca*	DI580	JF944517.1	JF956555.1
12.	*Suaeda glauca*	Z459	JF944516.1	JF956554.1
13.	*Suaeda glauca*	Z460	JF944515.1	JF956553.1
14.	*Suaeda glauca*	Z461	JF944514.1	JF956552.1
15.	*Suaeda maritima*	NMW2902	JN891221.1	JN894369.1
16.	*Suaeda maritima*	Halo1 & 2	JX997825.1	JX997824.1
17.	*Suaeda maritima*	NMW771	JN893488.1	JN896006.1
18.	*Suaeda maritima*	NMW2903	JN891222.1	DQ468647.1
19.	*Suaeda vera*	NMW770	JN893487.1	JN896005.1
20.	*Suaeda vera*	NMW2905	JN891228.1	AY042658.1

Table 2. Analysis of intraspecific and interspecific variation within and between species.

Genetic divergence	*rbcL*	*matK*
Mean intraspecific distance	0.0054 ± 0.0019	0.0153 ± 0.0028
Mean interspecific distance	0.0217 ± 0.0021	0.0502 ± 0.0057
The minimum intraspecific distance	0.0000 ± 0.0000	0.0000 ± 0.0000
The minimum interspecific distance	0.0069 ± 0.0015	0.0066 ± 0.0023

Table 3. Wilcoxon signed rank test for intraspecific divergence.

W+	W-	Relative rank		n	*P*-values	Result
		W+	W-			
rbcL	*matK*	7	14	6	P=0.463	*rbcL=matK*

Table 4. Wilcoxon signed rank t test of interspecific divergence among DNA barcoding loci.

W+	W-	Relative rank		n	*P*-values	Result
		W+	W-			
rbcL	*matK*	8	112	15	P=0.003	*rbcL<mat*K

Table 5. Success of barcode regions for species identification.

Barcode region	Best match			Best close match			
	correct %	incorrect %	ambiguous %	correct %	incorrect %	ambiguous %	no match %
rbcL	17 (85.0)	0 (0.0)	3 (15.0)	14 (70.0)	0 (0.0)	1 (5.0)	5 (25.0)
matK	13 (65.0)	1 (5.0)	6 (30.0)	12 (60.0)	0 (0.0)	0 (0.0)	8 (40.0)
rbcL+matK	30 (75.0)	0 (0.0)	10 (25.0)	29 (72.5)	0 (0.0)	2 (5.0)	9 (22.5)

Phylogenetic tree based analysis: The tree which was inferred by using the sequence data of *rbcL* region resolved the monophyletic relationship between the members of *Suaeda monoica* (Forssk. ex J. F. Gmel.)*, Suaeda fruticosa* (Forssk. ex J. F. Gmel.)*, Suaeda vera* (J. F. Gmelin)and *Suaeda maritime* (C. A. Mey.) Moq. (Fig. 3). However, *Suaeda glauca* (Bung.) Bung., is forming a polyphyletic relationship with 98% bootstrap support. *Suaeda acuminata* JF944510 is nested within the monophyletic clade formed by *Suaeda fruticosa* and *Suaeda monoica*. Moreover, paraphyletic relationship was not observed in any species. Tree topology (Fig. 4) of *matK* based tree is depicting the monophyletic relationship of *Suaeda monoica*, *Suaeda fruticosa, Suaeda vera* and *Suaeda maritima* with high support of bootstrap 98%, 81%, 100% and 83% respectively. Likewise, polyphyletic association between the members of *Suaeda acuminata* and *Suaeda glauca* received strong statistical support (100%). Combined (*rbcL+matK*) data analysis revealed almost the same tree topology as obtained by a single data analysis of *matK* region (Fig. 5).
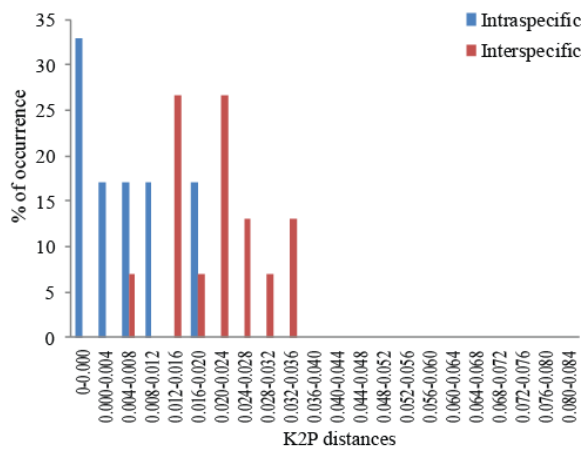

Fig. 1. Distribution of intraspecific and interspecific genetic variability for the *rbc*L region. The X-axes correspond to the K2P pairwise distances and the Y-axes relate to the percentage of occurrence.
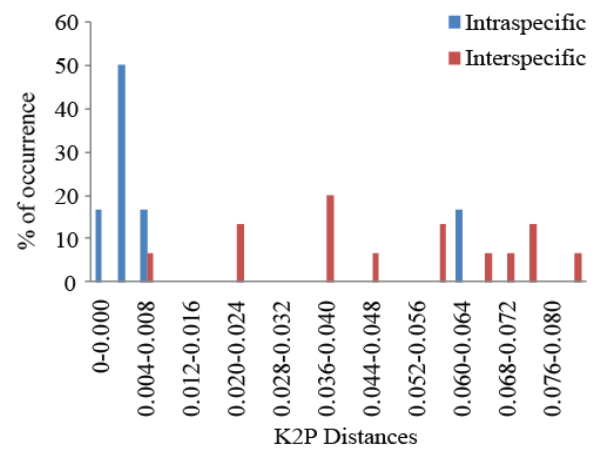

Fig. 2. Distribution of intraspecific and interspecific genetic variability for the *mat*K region. The X-axes correspond to the K2P pairwise distances and the Y-axes relate to the percentage of occurrence.
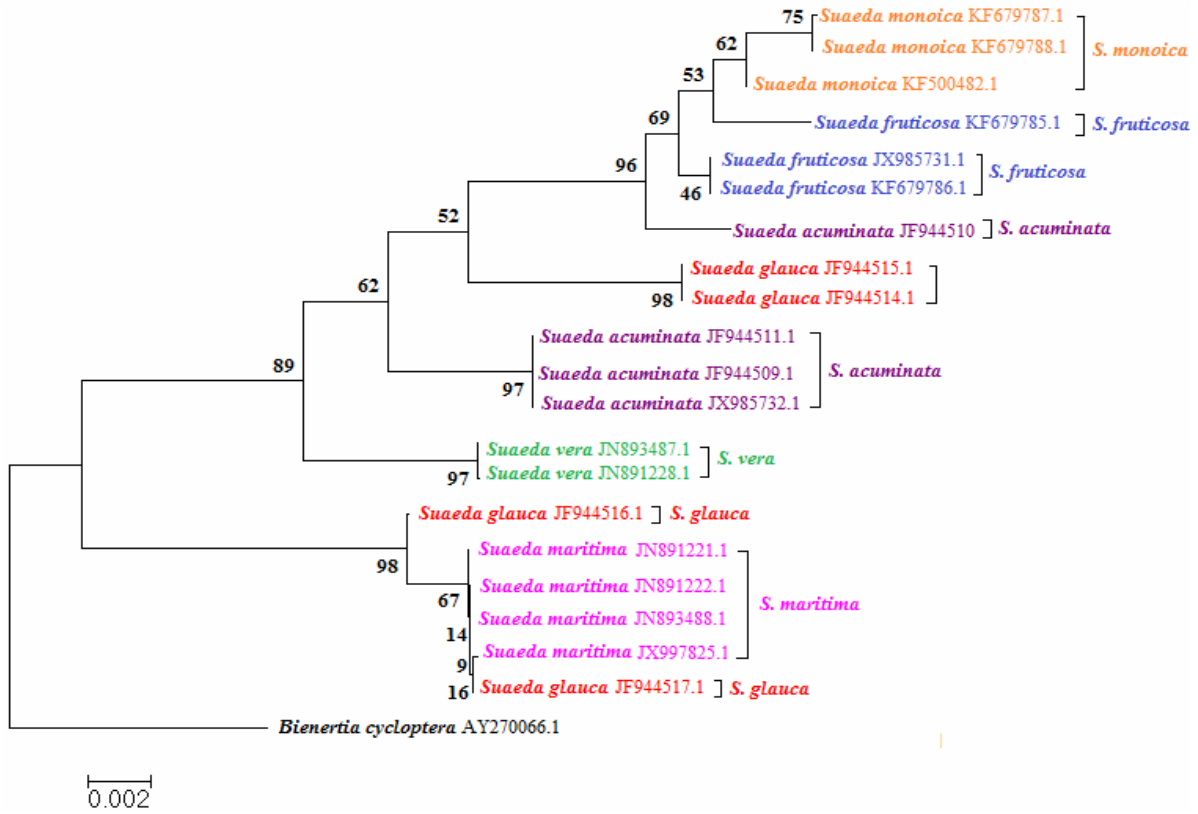
Fig. 3. NJ tree based on *rbcL* sequences, bootstrap values are shown at the relevant branches.
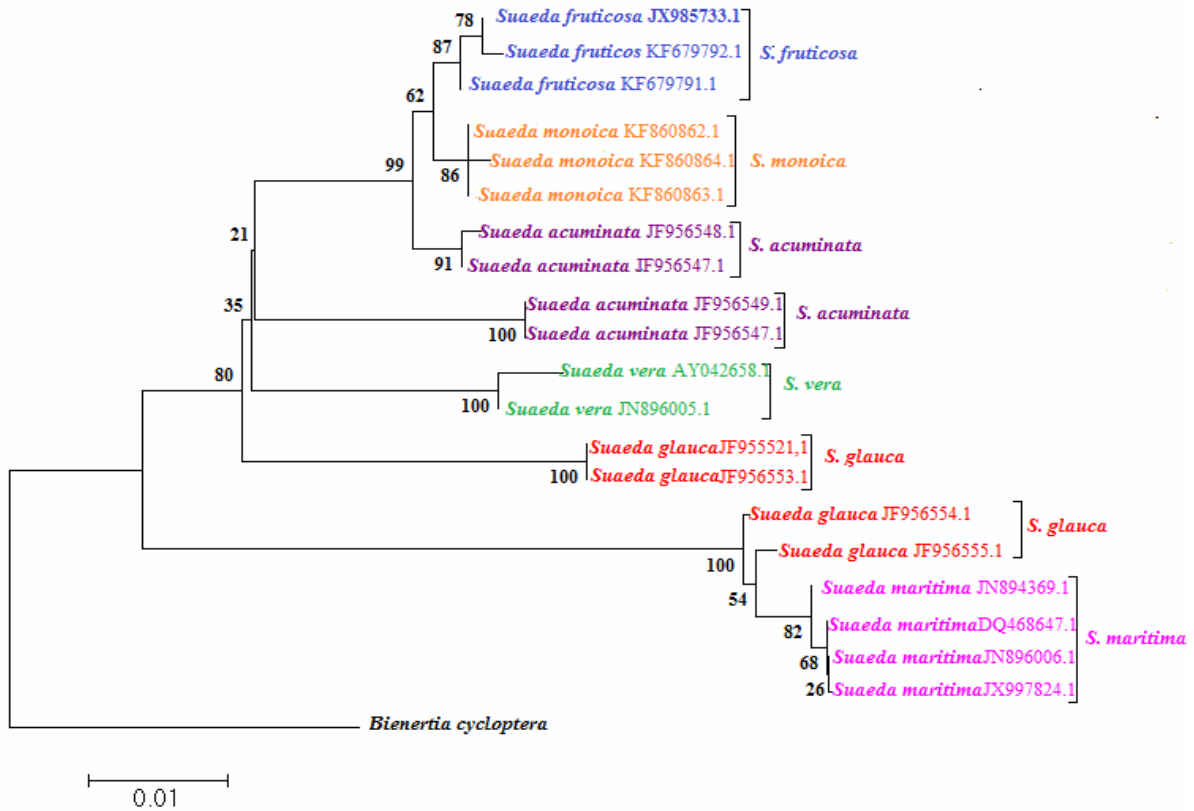
Fig. 4. NJ tree based on mat*K* sequences, bootstrap values are shown at the relevant branches.
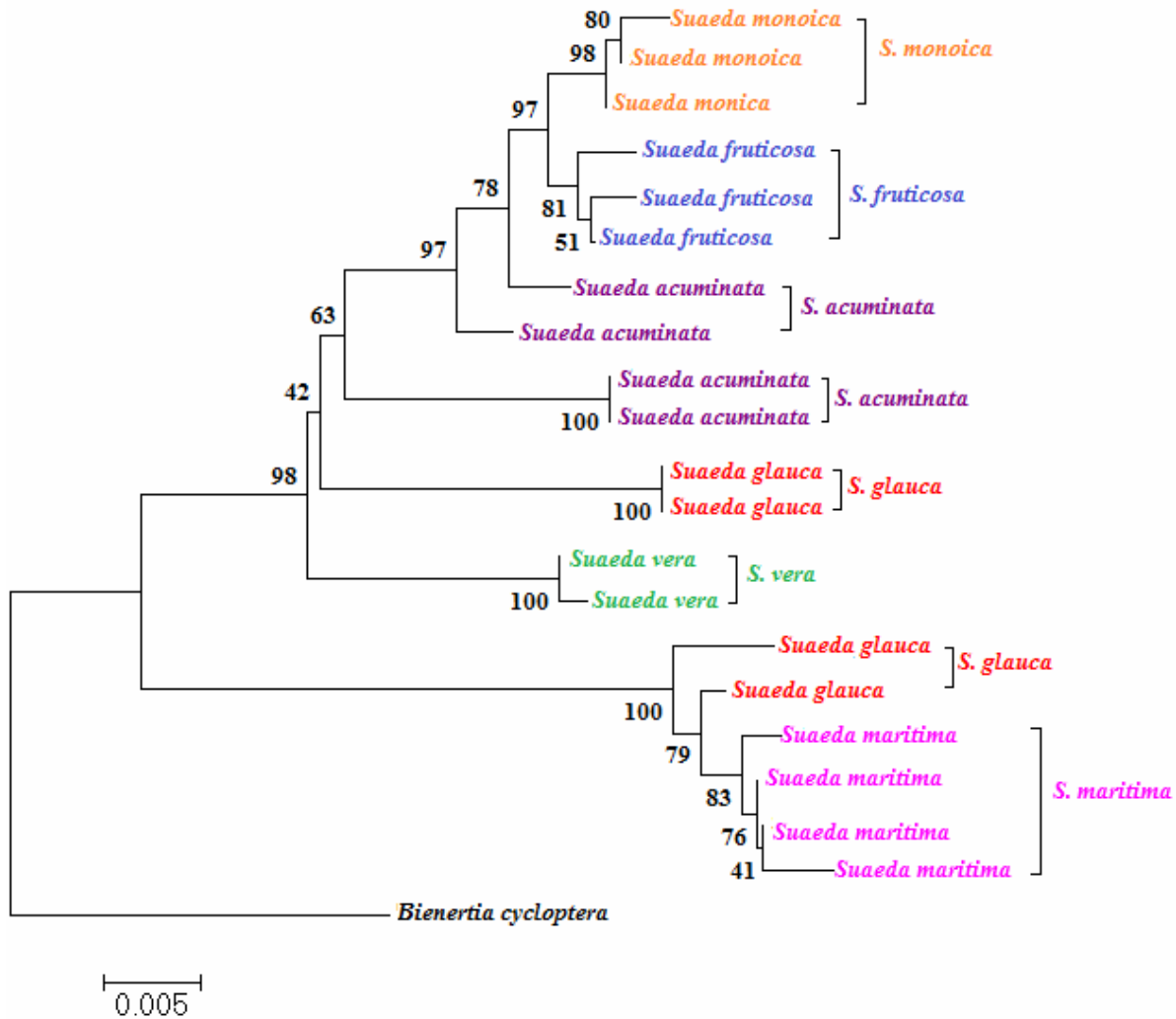
Fig. 5. NJ tree based on *rbcL+matK* sequences; bootstrap values are shown above the relevant branches.

Discussion

The utility of DNA Barcoding has been successfully assessed in most of the animal groups, however, great deal of efforts are still needed to establish core barcode region(s) in plants. In current barcoding analysis, comparison of two barcode markers showed that *rbcL* could serve as a potential barcode region for identification of genus *Suaeda* , because this region is having less divergence between the members of the same species and exhibiting enough genetic distance between the species, therefore showing the better performance than the *matK* region. Accuracy of barcoding marker depends on (barcoding gap) the extent of separation between intraspecific and interspecific divergence. Meyer & Paulay in 2005 suggested that barcoding technique becomes less effective by the increase of overlapping between intraspecific and interspecific variation, in such case selected marker do not reliably distinguishes between species. Although, *rbcL* region is not providing a well-defined barcoding gap, butcomparatively less overlapping was observed for *rbcL* region between the distribution of intraspecific and interspecific distances. After the recommendation of *rbcL+matK* regions as core barcodes,

many workers supported the recommendation of CBOL (Kress *et al.,* 2009; Burgess *et al.,* 2011). The discriminatory power of *rbcL* marker alone is higher (85.0%) than *rbcL+matK* combination, which can identify 75.0% species correctly by applying the best match criteria. Thus, our finding is not supporting the use of *rbcL+matK* combination as the barcode marker for the species identification of the targeted species of *Suaeda*. The least efficiency of *rbcL+matK* was observed in the other barcoding studies as well (Fu *et al.,* 2011; Jeanson *et al.,* 2011).

In tree based analysis NJ method was used to test the monophyletic relationship between the species because the NJ method has proven highly useful for estimating relatedness among species (Erickson & Driskell, 2012). NJ tree reconstruction method was also used in a DNA barcoding study for the identification of endemic plant species (Khan *et al.,* 2013). The polyphyletic relationship within *Suaeda glauca* species was examined constantly in *rbcL* and *matK* sequence based phylogenetic trees as well as in the combined data analysis. Inaccurate taxonomy and high level of divergence between the individuals, both may lead to the non-monophyletic (paraphyly or polyphyly) relationships (Fazekas *et al,.* 2009). Therefore,

in the present research polyphyletic association revealed that there might be some confusion in taxonomic assignments of *Suaeda glauca.* The unclear relationship between the replicates of *Suaeda glauca* species could be clarified in further investigation by increasing the sample size and using more molecular marker data.

Conclusions

DNA barcoding is found to be a useful and effective means for identification of *Suaeda* species. The current results revealed that the higher discriminatory resolution for *Suaeda* species identification is provided by a single marker (*rbcL*) than using the combination of markers. Comparison of two barcode markers showed that *rbcL* is a better candidate for the identification of genus *Suaeda* species.

References

Burgess, K.S., A.J. Fazekas, P.R. Kesanakurti, S.W. Graham, B.C. Husband, S.G. Newmaster, D.M. Percy, M. Hajibabaei and S.C.H. Barrett. 2011. Discriminating plant species in a local temperate flora using the *rbc*L+*matK* DNA barcode. *Methods. Ecol. Evol.*, 2: 333-340.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA.*, 106: 12794-12797.

Chase, M.W., N. Salamin, M. Wilkinson, J.M. Dunwell, R.P. Kesanakurthi, N. Haidar and V. Savolainen. 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. London.*, 360: 1889-1895.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic. Acids. Res.*, 16: 10881-10890.

Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bulletin*, 19: 11-15.

Erickson, D.L. and A.C. Driskell. 2012. Construction and analysis of phylogenetic trees using dna barcode data. *Methods. Mol. Biol.*, 858: 395-408.

Fazekas, A.J., P.R. Kesanakurti, K.S. Burgess, D.M. Percy, S.W. Graham, S.C.H. Barrett and S.G. Newmaster. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol. Ecol. Resource.*, 9: 130-139.

Felsenstain, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution.*, 39: 783-791.

Freitag, H., I.C. Hedge, S.M.H. Jafri, G. Heinririch, S. Omer and P. Uotila. 2001. *Flora of West. Pak, Chenopodiaceae,* 204: 01-213.

Fu, M.Y., W.M. Jiang and C.X. Fu. 2011. Identification of species within *Tetrastigma* (Miq.) Planch. (Vitaceae) based on DNA barcoding techniques. *J. Syst. Evol.*, 49: 237-245.

Hebert, P.D.N., A. Cywinska, S.L. Ball and J.R. DeWaard. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. London.*, 270: 313-322.

Jeanson, M.L., J.N. Labat and D.P. Little. 2011. DNA barcoding: a new tool for palm taxonomists? *Ann. Bot.*, 108: 1445-1451.

Khan, S., F. Al-Qurainy, M. Nadeem and M. Tarroum. 2013. Selection of chloroplast DNA markers for the development of DNA barcode and reconstruction of phylogeny of *Senecio asirensis* Boulous and Jri wood. *Pak. J. Bot.*, 45: 703-710.

Kress, W.J., K.J. Wurdack, E.A. Zimmer, L.A. Weigt and D.H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA.*, 102: 8369-8374.

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: 1-10.

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: 18621-18626.

Li, D.Z., L.M. Gao, H.T. Li, Hong, W. Ge, X.J. Liu, J.Q. Chen, Z.D. Zhou, S.L. Chen, S.L. Yang, J.B. Fu, C.X. Zeng, C.X. Yan, H.F. Zhua, Y.J. Suna, Y.S. Chena, S.Y. Zhaoa, L. Wanga, K. Yanga and T.G.W. Duana. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc Natl. Acad. Sci. USA.*, 108: 19641-19646.

Meier, R., S. Kwong, G. Vaidya and P.K.L. Ng. 2006. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Syst. Biol.*, 55: 715-728.

Meyer, C.P. and G. Paulay. 2005. DNA barcoding: Error rates based on comprehensive sampling. *PLoS. Biol.,* 3: 2229-2238.

Newmaster, S.G., A.J. Fazekas, R.A.D Steeves and J. Janovec. 2008. Testing candidate plant barcode regions with species of recent origin in the Myristicaceae. *Mol. Ecol. Notes.,* 8: 480-490.

Palmer, J.D and L.A. Herbon. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.,* 28: 87-97.

Schenk, H.J. and W.R. Ferren. 2001. On the sectional nomenclature of *Suaeda* (Chenopodiaceae). *Taxon,* 50: 857-873.

Schori, M. and A.M. Showalter. 2011. DNA barcoding as a means for identifying medicinal plants of Pakistan. *Pak. J. Bot.,* 43: 1-4.

Schutze, P., H. Freitag and K. Weising. 2003. An integrated molecular and morphological study of the subfamily Suaedoideae Ulbr. (Chenopodiaceae). *Plant. Syst. Evol.,* 239: 257-286.

Shinwari, Z.K., R. Terauchi, F.H. Utech and S. Kawano. 1994a. Recognition of the New World Disporum Section Prosartes as Prosartes (Liliaceae) based on the sequence data of the rbcL gene. *Taxon.,* 43: 353-366.

Shinwari, Z.K., R. Terauchi and S. Kawano. 1994b. Phylogenetic relationships among genera in the Liliaceae-Asparagoideae-Polygonatae sensu lato inferred from rbcL gene sequence data. *PI. Syst. Evol.*, 192: 263-277.

Shinwari, Z.K. 1995 Congruence between morphology and molecular phylogeneties in Prosartes (Liliaceae). *Pak. J. Bot.,* 27: 361-369.

Shinwari, Z.K., K. Jamil and N.B. Zahra. 2014. Molecular systematics of selected genera of family fabaceae. *Pak. J. Bot.,* 46(2): 591-598.

Shneyer, V.S. 2009. DNA barcoding is a new approach in comparative genomics of plants. *Russ. J. Genetics.,* 45: 1267-1278.

Tamura, K., J. Dudley, M. Nei and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol. Evol.,* 24: 1596-1599.

Thompson, J.D., D.G Higgins and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic. Acids. Res.,* 22: 4673-4680.

Wolfe, K.H., W.H. Li and P.M. Sharp. 1987. Rates of Nucleotide Substitution Vary Greatly among Plant Mitochondrial Chloroplast, and Nuclear DNAs. *Proc. Natl. Acad. Sci. USA.,* 24: 9054-9058.