International Journal Of Biology and Biological Sciences Vol. 4(1), pp. 019-026, April 2015
Available online at http://academeresearchjournals.org/journal/ijbbs
ISSN 2327-3062 ©2015 Academe Research Journals

Full Length Research Paper

A DNA barcode analysis of the genus *Suaeda* (Chenopodiaceae) species based on *rbcL* and *matK* regions of the chloroplast DNA

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.

Accepted 21 November, 2014

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 the current research, the efficiency of *rbcL* and *matK* (plants core barcode regions) for species identification of the genus *Suaeda* efficiency was assessed. For DNA barcode analysis, 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 out of the two investigated barcode regions, *rbcL* showed comparatively less overlapping for the distribution of interspecific and intraspecific divergence. In addition, the highest discriminatory 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, it belongs to the family Chenopodiaceae and it comprises c. 100 species which are cosmopolitan in distribution (Schenk and Ferren, 2001). Species of the genus *Suaeda* are very 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 in 2003 for the rapid species identification by comparing sequences of short standardized DNA marker with sequences of reference database (Hebert et al., 2003). 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 proven 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 and 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-*

*Corresponding author. E-mail: uzmalink@gmail.com.

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

S/N	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

Accession numbers which were highlighted with different colors were submitted in GenBank by the authors, while other accession numbers were retrieved by Genbank.

psbA (Kress et al., 2005), *rbcL* (Chase et al., 2005) and *matK* and *trnH-psbA* (Newmaster et al., 2008). In 2009, the CBOL executive committee 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 present were included (Li et al., 2011). There is no detailed report available on the barcode of *Suaeda* species, therefore, in this study *rbcL* with *matK* regions were used to assess the efficiency and suitability for identification purpose.

MATERIALS AND METHODS

Plant material

In the 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 acuminate* (C. A. Mey.) Moq., were examined and their herbarium sheets were deposited at the Karachi University Herbarium (Centre for plant conservation). 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 samples were used to extract the total genomic DNA by using a cetyltrimethyl ammonium bromide (CTAB) DNA extraction method with some modifications (Doyle and Doyle 1987). The *rbcL* region was amplified by using primer pair (forward): 5'-ATGTCACCACAAACAGAGACTAAAGC-3' and (reverse): 5'-GTAAAATCAAGT CCACCRCG-3' (Kress and Erickson, 2007). The *matK* region was amplified by using primer pair (forward): 5'-CGTACAGTACTTTTGTGTTTACGAG-3' (reverse): ACCCAGTCCATCTGGAAATCTTGGTTC-3' (Ki-Joong Kim unpublished). Amplification was performed in total 20 µL reaction volume containing 1 × 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 primers, 1.0 unit of *Taq* DNA polymerase (Bioneer, Korea), deionized H_2O 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 s at 94°C, 30 s at 54°C and at 72°C for 1 min, with a step of final extension of 10 min at 72°C. For *matK*, the annealing

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<matK*

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.

Sequence analysis

Forward and reverse sequences were aligned by using the software Multalin (Corpet 1988) and submitted to GenBank for their accession numbers.

Data analysis for DNA barcoding

All 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 and Paulay (2005). The significance between intraspecific and interspecific variation was examined by Wilcoxon signed rank test in SPSS 16.0 (Levesque, 2007). In order to evaluate the success rate of each barcoding marker, "Best match" and "Best close match" criteria were employed by using the software TaxonDNA (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 the current research, the highest intraspecific and interspecific divergence was recorded for *matK* (0.0153 and 0.0502) than that of the *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 (Figure 1). On the other hand, *matK* exhibited large overlapping between intraspecific and interspecific

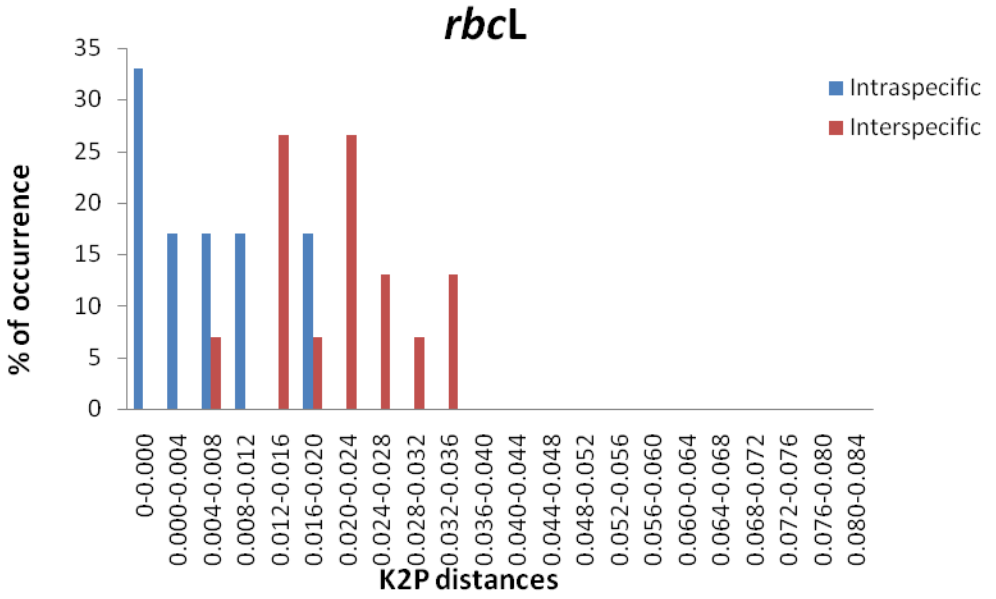

Figure 1. Distribution of intraspecific and interspecific genetic variability for the *rbc*L region. The X-axes correspond to the K2P pairwise distances and the Y-axes relate to the percentage of occurrence.

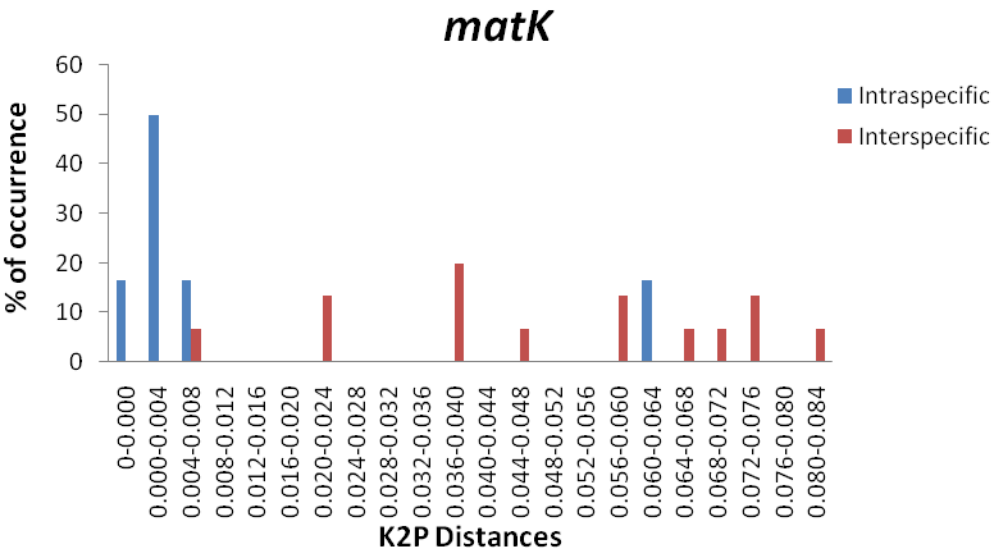

Figure 2. Distribution of intraspecific and interspecific genetic variability for the *mat*K gene. The X-axes correspond to the K2P pairwise distances and the Y-axes relate to the percentage of occurrence.

divergence (Figure 2).

Efficiency of markers for species discrimination

The species discriminatory power was recorded as 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 makes two marker combination with (75.0 and 72.5%) correct species values (Table 5).

Phylogenetic tree based analysis

The tree which was inferred by using the sequence data

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)

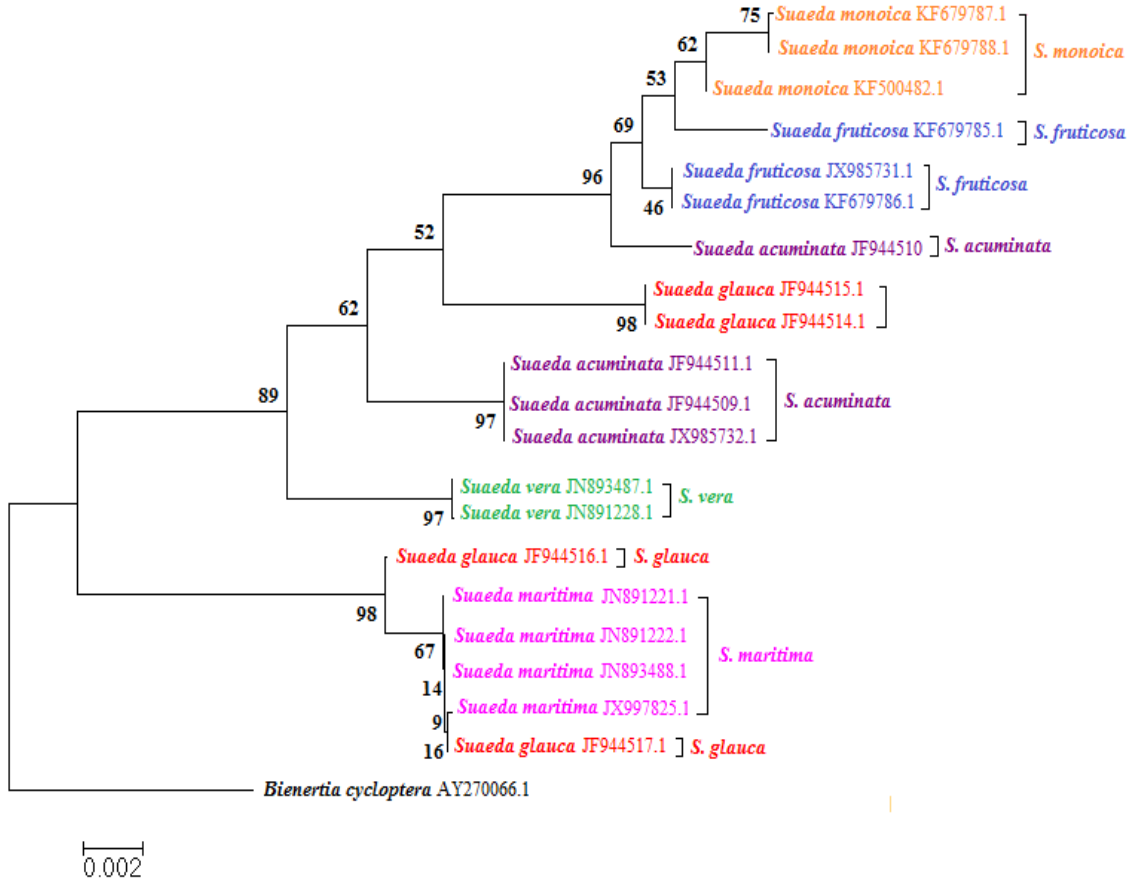

Figure 3. NJ tree based on *rbcL* sequences; bootstrap values are shown at the relevant branches.

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 maritima* (C. A. Mey.) Moq. (Figure 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 (Figure 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 as 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 (Figure 5).

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

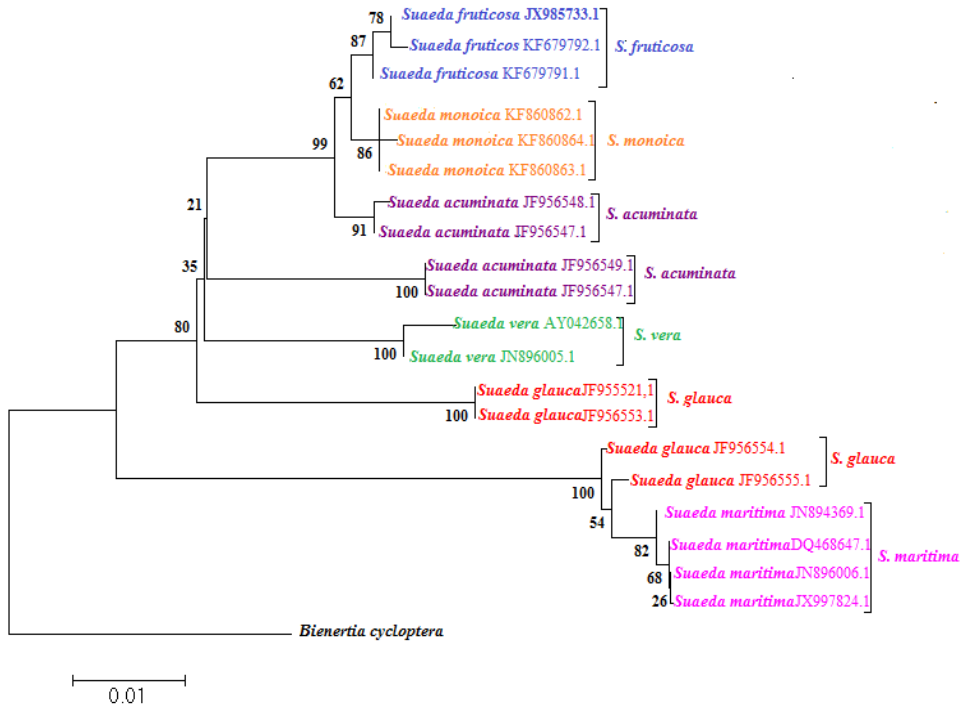

Figure 4. NJ tree based on mat*K* sequences; bootstrap values are shown at the relevant branches.

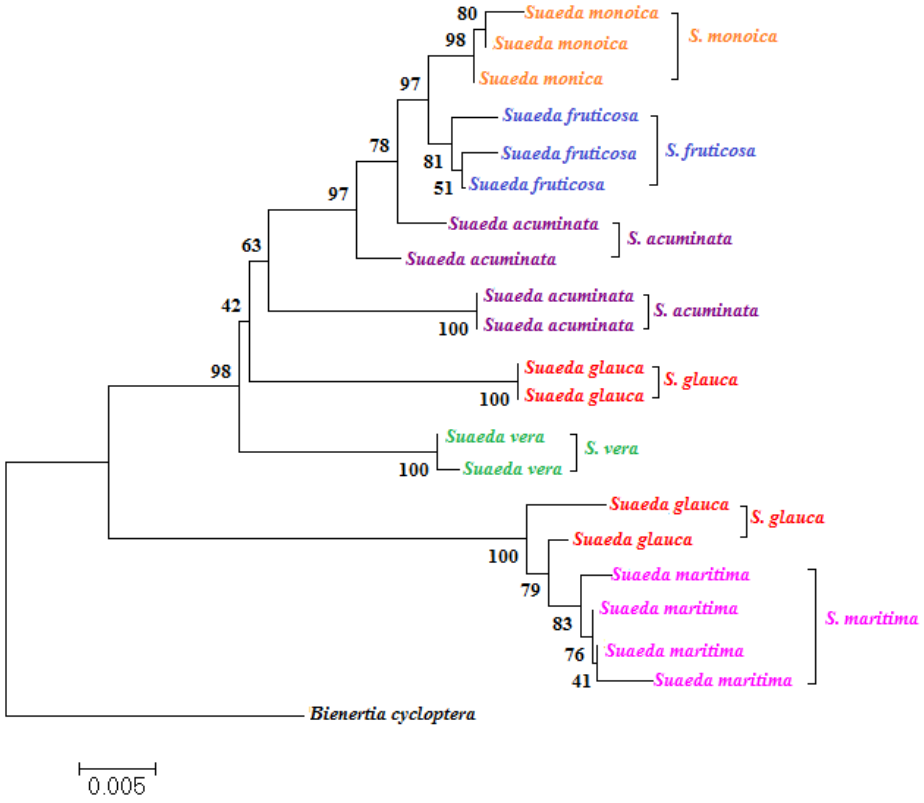

Figure 5. NJ tree based on *rbcL+matK* sequences; bootstrap values are shown above the relevant branches.

region(s) in plants. In the current barcoding analysis, comparison of two barcode markers showed that *rbcL* could be served as a potential barcode region for identification of genus *Suaeda* species, 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 and Paulay in 2005 suggested that barcoding technique becomes less effective by the increase of overlapping between intraspecific and interspecific variation; in such a case, the selected marker does not reliably distinguishes between species. Although, *rbcL* region is not providing a well-defined barcoding gap, but out of the two regions comparatively, less overlapping was observed for *rbcL* region. After the recommendation of *rbcL+matK* regions as core barcode, many workers have 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 and Driskell, 2012). 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 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 was found to be a useful and effective mean 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.

Results which are obtained from the current analysis may improve for high rate of species identified by the use of other DNA barcoding marker and by comprehensive species sampling. The present investigation contributes towards the establishment of DNA barcode for flowering plants.

REFERENCES

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy D M, Hajibabaei M, Barrett SCH (2011). Discriminating plant species in a local temperate florausing 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 MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V (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 JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull., 19: 11–15.

Erickson DL, Driskell AC (2012). Construction and Analysis of Phylogenetic Trees using DNA Barcode Data. Methods Mol. Biol., 858: 395-408.

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SCH, Newmaster S G (2009). Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Mol. Ecol. Resour., 9: 130-139.

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

Freitag H, Hedge IC, Jafri SMH, Heinririch G, Omer S, Uotila, P (2001). Fl. West. Pakistan. Chenopodiaceae, 204: 01-213.

Fu MY, Jiang WM, Fu CX (2011). Identification of species within *Tetrastigma* (Miq.) Planch.(Vitaceae) based on DNA barcoding techniques. J. Syst. Evol., 49: 237-245.

Hebert PDN, Cywinska A, Ball S L, DeWaard J R (2003). Biological identifications through DNA barcodes. Proc. R. Soc. London., 270: 313-322.

Jeanson ML, Labat JN, Little DP (2011). DNA barcoding: a new tool for palm taxonomists? Ann. Bot., 108: 1445-1451.

Kress WJ, Wurdack KJ, Zimmer EA (2005). Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. USA., 102: 8369-8374.

Kress JW, Erickson DL (2007). A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region.

PLoS. ONE, 2: 1-10.
Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham, E (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proc. Natl. Acad. Sci. USA., 106: 18621-18626.
Levesque R (2007). SPSS Programming and Data Management: A Guide for SPSS and SASUsers, Fourth Edition SPSS Inc., Chicago Ill., pp. 327-361.
Li DZ, Gao LM, Li HT, Hong W, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhua YJ, Suna YS, Chena SY, Zhaoa L, Wanga K, Yanga T, Duana GW (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, Kwong S, Vaidya G, Ng PKL (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Syst. Biol., 55: 715-728.
Meyer CP, Paulay G (2005). DNA barcoding: Error rates based on comprehensive sampling. PLoS. Biol., 3: 2229–2238.
Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J (2008). Testing candidate plant barcode regions with species of recent origin in the Myristicaceae. Mol. Ecol. Notes., 8: 480-490.
Palmer JD, Herbon LA (1988). Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J. Mol. Evol. 28: 87-97.
Schutze P, Freitag H, Weising K (2003). An integrated molecular and morphological study of the subfamily Suaedoideae Ulbr. (Chenopodiaceae). Plant. Syst. Evol., 239: 257–286.
Shneyer VS (2009). DNA barcoding is a new approach in comparative genomics of plants. Russ. J. Genet., 45: 1267-1278.
Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol., 24: 1596-1599.
Thompson J D, Higgins D G, Gibson T J (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 KH, Li WH, Sharp PM (1987). Rates of Nucleotide Substitution Vary Greatly among Plant Mitochondrial Chloroplast, and Nuclear DNAs. Proc. Natl. Acad. Sci. USA., 24: 9054–9058.