Evolutionary Analysis based on DNA Barcoding of Certain Aquatic Plants using *rbcL* Gene Sequences

Punita Parikh, Krupa Unadkat and Padamnabhi Nagar
Department of Botany, Faculty of Science, The Maharaja Sayajirao University of Baroda, Vadodara

ABSTRACT : DNA barcoding is currently gaining popularity due to its simplicity and high accuracy as compared to the complexity and subjective biases associated with morphology-based identification of taxa. The standard chloroplast DNA barcode for land plants is normally recommended by the Consortium for the Barcode of Life (CBOL) plant working group that needs to be evaluated for a wide range of plant species. In Vadodara district of Gujarat, there are many natural and manmade seasonal wetlands that serve as important gateways for the migratory waterfowl. Hence these wetlands contribute significantly towards balancing of ecosystem. Many researchers had identified plant species of these wetlands by traditional methods but the aspect of identification of plants through barcoding has largely been ignored. We therefore, tested the potential of the rbcL marker for the identification of aquatic plants belonging to diverse families of some wetlands of Vadodara. This had been carried out by 2-locus combination of rbcL and matK as the standard plant barcode Maximum likelihood tree analysis was also performed to evaluate the discriminatory power of the rbcL gene. The classical taxonomic classification was then compared with the classification obtained through DNA Barcode tree. For evolutionary analysis, Tajima's D Test, Substitutional Matrix and Nucleotide substitution model analysis were performed. Our findings showed that using rbcL gene sequences, majority of the samples, i.e. (90%) were identified at genus level but at species level only 10% identification was possible.

Key words: DNA barcoding; rbcL; aquatic plants; Identification; Phylogenetics, DNA barcoding, Evolutionary Analysis

INTRODUCTION

DNA barcoding is defined as methods for identifying species by using short orthologous DNA sequences, known as ''DNA barcodes'', that have been proposed and initiated to facilitate biodiversity studies, identify juveniles, associate sexes, and enhance forensic analyses (Von,2011 and Kress 2011). The criteria for the development of reliable barcode data was defined by the Consortium for the Barcoding of Life (CBOL) are : Candidate loci should be suitable for a wide range of taxa, show high variation between species, but it should be conserved within species, so that the intraspecific variation would be insignificant (Hollingsworth, 2009). Ideal barcodes should be routinely retrievable with a single primer pair, be amenable to bidirectional sequencing with little requirement for manual editing of sequence traces and be short enough to ease PCR amplification. Alignment and analysis of the resulting sequences ought to be straightforward, allowing a fast identification without profound prior knowledge about bioinformatics. Based on assessments of recoverability, sequence quality, and levels of species discrimination, the Consortium for the Barcode of Life (CBOL) plant working group has recommended a standard barcode comprising of ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) and/or maturase K (*matK*) for the barcoding of all land plants (CBOL Plant Working Group, 2009). However, whether *rbcL* shows sufficient variation to allow general identification below genus level remains uncertain. Moreover, nowadays it is widely accepted that any valid plant barcode will be multi-locus, preferably existing of a conservative coding region like *rbcL*, in

combination with a more rapidly evolving region, which is most likely non-coding (Kress et al, 2009). Nucleotide polymorphism was observed in alignments that included different species, and these polymorphisms were enough to discriminate these species. The power of a barcode to discriminate species is an essential criterion that determines its utility The mat K sequence had a lower calculated discrimination power of 66.66% (discriminated 6 of 10 species) than rbcL, which was about 90% (discriminated 9 of 10 species).(Neem et al, 2014)

However, the universality of barcode markers is hampered due to morphological/geographical variation and reticulate evolution in plant species (Roy et al., 2010). The ongoing research on plant barcoding suggests that the development of universal DNA barcoding markers for land plants is quite challenging; even the choice of the correct loci has been debated (Chase et al., 2005; Kress et al., 2005; Fazekas et al., 2008; de Groot et al., 2011). Arguments about the selected core loci for plant barcoding also related to the lack of discriminatory power and/or primer universality (Roy et al., 2010).

Traditional methods based on morphological criteria are difficult to apply accurately due to subjective biases. Particularly, in the case of medicinal plants, the use of chromatographic profiles of marker compounds to standardize botanical preparations has limited value because medicines are from varied sources and posess chemical complexity, which is affected by growth, storage conditions and harvest times (Joshi et al., 2004; Zhang et al., 2007).

DNA-based identification (barcoding) is a simple method, does not require taxonomic expertise and is free from subjective errors, which is not the case in morphological identification. Valid identification of unknown samples is the main goal of barcoding (Hebert and Gregory, 2005), despite ongoing criticism of the feasibility or even necessity of DNA barcoding for general taxonomic purposes (Will et al., 2005; Spooner, 2009). Nowadays, it is widely accepted that any valid plant barcode should be multi-locus, preferably comprising a conserved coding region such as *rbcL* and a more rapidly evolving region that is most likely non-coding (Kress et al., 2009). Sequences of the *rbcL* and *trnL-F* genes as two-locus DNA barcode have recently been used successfully to identify NW-European ferns, whereas the selected *matK* locus was unsuccessful for barcoding of the same (de Groot et al., 2011). However, whether *rbcL* exhibits sufficient variation to allow general identification of wild plants grown in arid environments below genus level remains unexplored. In continuation of previous studies on the PCR success rate (Bafeel et al., 2011) and molecular characterization of desert medicinal plants (Arif et al., 2010), we evaluated the bar coding performance of *rbcL* for the identification of aquatic plants and demonstrated genus- and species-level discriminations using this marker.

MATERIAL AND METHODS

Plant Collection

Plants were collected from two ponds namely Harni Pond and Gotri Pond and from Mahi River of District Vadodara, Gujarat, India.

Free Floating Plants: Some free floating plants like duckweeds, neither make very satisfactory pressed of dried specimens, nor is it convenient to collect them into bag. But Small 50 ml cap bottles are good collecting and preservation containers for this plant. To keep the collected plants in a moist condition, little water is poured into the bottle and then preserved them with a solution of 10% formalin which is a permanent preservative.

Emergent Hydrophytes Plant: Most submerged plants do not have any supporting tissue, they rely on water for support and they must be floated on to the card stock. Therefore, a paper which remains dimensionally stable under wetting is generally used for such type of plants. Therefore so, subsequent drying was done for the preservation of these plants.

DNA extraction

Plant leaf samples from 09 different species were used for DNA extraction. The specimens were macerated using a sterile mortar and pestle under liquid nitrogen. Genomic DNA was extracted using fresh leaves of plants weighing approximately 50 mg. Extraction was carried out using Macherey NAGEL kit. The concentration and quality of the extracted DNA were determined using gel electrophoresis and a NanoDrop 8000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). The isolated genomic DNA was stored at -20°C until used.

Quantification of genomic DNA

After extraction of genomic DNA, quantification was done according to Sambrook et al., 1982. 10 µL of extracted DNA was dissolved in 30 µl of Tris buffer (pH 8) and O.D. was taken at 260 and 280 nm (PowerWave HT Microplate Spectrophotometer, BioTek). Quantity of DNA was calculated by using following formula:

Amount of DNA (ng/µL) = O.D. at 260 X dilution factor X 50 (extension coefficient)

Quality was assessed by taking the O.D. at 260nm/ O.D. at 280nm. Samples which showed the O.D. between 1.6-1.8 were taken for further analysis.

PCR and gene sequencing

rbcL gene in plants was amplified in a volume of 20 µl containing 10 µl Taq PCR reaction mix, 10 pmol forward primer (rbcLa F - ATGTCACCACAAACAGAGACTAAAGC), 10 pmol reverse primer (rbcLa R - GTAAAATCAAGTCCACCRCG), template DNA (50 mg/µl) and sterile ion-free water (to make up the final desired volume). Amplification was carried out in Thermal cycler (Applied Biosystems VeritiR). Reactions were amplified through 35 cycles with the following conditions (Levin, 2003 and Kress & Erickson, 2007):Denturation: 30 seconds at 95°C. Annealing: 40 seconds 55°C, Extension: Two minutes at 72°C. This was followed by a final extension step at 72°C for seven minutes. Initial denturation was carried out at 95°C for five minutes. Electrophoresis was carried out at 150V. The gel images were recorded in JPEG or TIF formats using gel documentation system (Biorad, USA). The gels were analyzed by using the software Image lab version 3.0 (Biorad, USA). Purification of rbcl gene amplified products were done using GenElute™ PCR Clean-up kit (cat no. NA 1020-1kt) Sequencing was carried out using BigDyeR Terminator v 3.1 Cycle sequencing kit. The BigDye Terminator v3.1 Cycle Sequencing Kit provides the required reagent components for the sequencing reaction in a ready reaction, pre-mixed format. Cycle sequencing was performed in 10 µl volume. Purification of cycle sequencing product was done using BigDye XTerminatorR Purification Kit. Kit contains SAM™ Solution and BigDyeR XTerminator™ Solution. Capillary electrophoresis of cycle sequenced products was Performed on 3500 XL platform (Applied biosystems).

Analysis and Sequence alignment

Sequence analysis was performed using sequencing analysis version 5.4 (Apllied Biosystems) and BioEdit, biological sequence alignment editor (Ibis Biosciences) (Hall, 1999).

Basic Local Alignment Search Tool (BLAST) searches were applied to all produced sequences using available online databases (DDBJ/EMBL/GenBank). BLAST was never intended to be used in this manner, but could provide valuable insights into how well we can expect the possibly more appropriate plastid *rbcL* short sequence regions to perform as barcodes (Chase et al., 2005). There are very few *rbcL* records on the current BOLD (Barcode of Life Data) identification system (v 2.5) (Ratnasingham and Hebert, 2007); thus, queries might not return an authentic match. Identification at genus level was considered successful when all hits with maximal percent identity scores >95% involved a single genus. Species identification was considered successful only when the highest maximal percent identity included a single species and scored >95% (de Groot et al., 2011). The *rbcL* sequences were matched with the query sequences and available *rbcL* sequences of the examined plant species; if not available, then genera were retrieved from the DDBJ/EMBL/GenBank databases.

Sequence match analysis using BLAST on NCBI. Consensus sequences which showed significant match with the earlier identified data on NCBI were submitted to BOLDSYSTEMS according to the guidelines provided onto BOLD website (http://www.boldsystems.org/). For few species where NCBI data was not available were subjected to detailed and thorough morphological analysis and submitted to BOLD.

Tree-based analysis.

Phylogenetic analyses were conducted in MEGA5 (Tamura et al., 2007), and the phylogenetic trees were inferred with the maximum likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993). In phylogenetic analyses, genus identification was considered successful when the unknown sample formed a monophyletic group together with all members of a single genus, with a bootstrap support of >70%. An equal strategy was applied for species-level identification (de Groot et al., 2011).

Tests based on polymorphisms within species

Tajima's D:

This statistic measures the difference between two estimators of the population mutation rate, Θ_w and π (Tajima, F. 1989). Under neutrality, the means of Θ_w and π should be approximately equal to one another. Therefore, the expected value of Tajima's D for populations conforming to a standard neutral model is zero. Significant deviations from zero indicate a skew in the allele frequency distribution relative to neutral expectations. Positive values of Tajima's D arise from an excess of intermediate frequency alleles and can result from population bottlenecks, structure and/or balancing selection. Negative values of Tajima's D indicate an excess of low frequency alleles and can result from population expansions or positive selection.

Estimate of Substitution Matrix

Substitution pattern and rates were estimated under the Tamura-Nei (1993) model. In this method the nucleotide sequences of ancestral DNA are inferred by using the principle of maximum likelyhood, and the directional changes of nucleotides are determined by comparing a sequence with its immediate ancestral sequence. When the nucleotide at a site of an ancestral sequence was ambiguous and two nucleotides were possible at the site, each of the nucleotides was considered as the ancestral nucleotide with a probability of ½ .

Nucleotide Substitution Analysis

The rate of nucleotide substitution (r) was allowed to vary from branch to branch, so that it would be convenient to measure evolutionary time in terms of the expected number of substitutions (v=r*t). The pattern of nucleotide substitution was computed by using Tamura Nei and Tamura 3 parameter model.

The nucleotide substitutions matrices obtained from different branches were averaged by weighting each matrix and the number of inferred substitutions for the branch. Further averages of matrices for two different genes were also calculated by the same weighting method. The overall transition/transversions bias for these 3 sequence types have been calculated and analyzed by using the formula R = [A*G*k1 + T*C* k2] / [(A+G)*(T+C)]. Where, k1 is the transition/transversions rate ratio for purine and k2 is the transition/transversions rate ratio for pyrimidine. All positions containing gaps and missing data were eliminated from the dataset (complete-deletion option) using MEGA.

RESULTS AND DISCUSSION

When overall outputs of BLAST matching and tree analysis were compared, the latter strategy resulted in better taxonomic assignment. The use of *rbcL* sequences with BLAST searching yielded 90 % genus and 10 % species-level identifications, respectively (Table 1). % value for genus and species level was calculated using BLAST tool (Bafeel et al, 2011)

Sr. No.	Morphological Identification	BLAST Search Match	BLAST Similarity (%)	Phylogenetic Affinity	BOLD Submission Accession No.
1	*Potamogeton natans*	*Potamogeton natans*	100	*Potamogeton Sp.*	GENG462-14
2	*Lemna triscula*	*Lemna Sp.*	100	*Lemna Sp.*	GENG463-14
3	*Azolla pinnata*	*Azolla pinnata*	100	*Azolla pinnata*	GENG373-14
4	*Spirodela polyrhiza*	*Spirodela polyrhiza*	100	*Spirodela polyrhiza*	GENG372-14
5	*Utricularia vulgaris*	*Utricularia vulgaris*	99	*Utricularia Sp.*	GENG382-14
6	*Marsilea quadrifolia*	*Marsilea*	98	*Marselia Sp.*	GENG381-14
7	*Hygrophila auriculata*	*Hygrophila*	100	*Hygrophila*	GENG395-14
8	*Hygrorhyza aristata*	*Hygrorhyza aristata*	100	*Hygrorhiza aristata*	GENG371-14
9	*Lymnophyton Obtusifolium*	*Lymnophyton Obtusifolium*	97	*Lymnophyton Sp.*	GENG37414

BLAST = Basic local alignment search tool

Table 1. Database search match for similarities and phylogenetic relationship using *rbcL* gene sequences and BOLD submitted DNA Barcode with their Accession number.

Tree analyses with the *rbcL* gene sequences assigned the majority of samples (90%) up to genus level and 10% up to species level (Table 1; Figure 1). In phylogenetic analyses, we considered genus identification successful when the unknown sample formed a monophyletic group together with all members of a single genus, with a bootstrap support of >70%. An equal strategy was applied for species-level identification (de Groot et al., 2011). Tree analyses using *rbcL* sequences assigned 17% of the tested plant samples to known species. Our findings, not withstanding *rbcL* is considered to possess less species-discriminating power than *matK*, were possibly due to its minimal sequence variation (Asahina et al., 2010). The estimated range of the total number of plant species worldwide is believed to be approximately 310,000-422,000 (Graham, 2002). When the data analyses of this experiment were carried out, the DDBJ/EMBL/GenBank databases contained only 8289 nucleotide sequences of the

matK gene and 12,909 nucleotide sequences of the *rbcL* gene of plant species. The availability of the sequences of barcoding genes in the databases is expected to increase rapidly, and subsequently, their utilization in the identification of plant species.

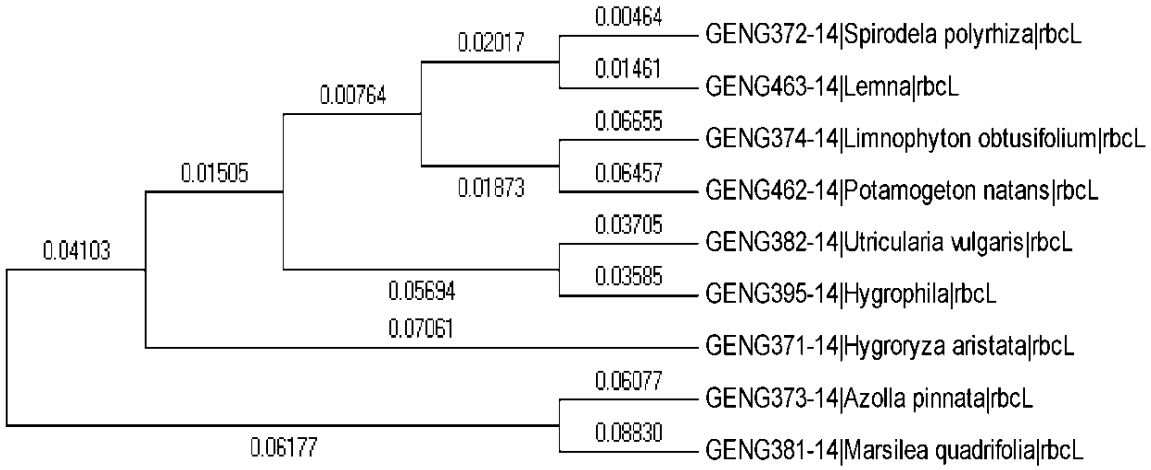

Figure 1 : Phylogenetic affinities of *rbcL* gene sequences of the plant samples. The evolutionary history was inferred by using the maximum likelihood method based on the Tamura-Nei model (1993).

The clades in the tree constructed with *rbcL* gene sequences were supported by 89% (25/28) of >50% bootstrap values (Figure 1). In a recent study, the *rbcL* marker exhibited intermediate-level (80%) resolution among the evaluated regions (*matK* > *atpF-atpH* > *rbcL* > *trnH-psbA* > *rpoC1*) (Burgess et al., 2011). Phylogenetic methods were applied in a recently conducted study of barcoding species using each barcode locus taken alone as well as in combinations to evaluate species recovery (Roy et al., 2010). When all sequences for a given locus were considered, ITS, *matK*, and *trnH-psbA* were able to form a species-specific clade for only *Berberis pachyacantha*. Not a single species was recovered with *rbcL* using any of the three methods. The clades formed in the trees were mostly mixtures of several species. Therefore, establishing a local barcode database will be valuable for a broad range of potential ecological applications, including the building of community phylogenies (Kress et al. 2009).

Sr. No.	Morphological Identification	1	2	3	4	5	6	7	8	9
1	*Hygrorhyza aristata*									
2	*Spirodela polyrhiza*	0.115								
3	*Azolla pinnata*	0.216	0.192							
4	*Limnophyton obstulifolium*	0.166	0.098	0.237						
5	*Marsilea quadrifolia*	0.245	0.208	0.149	0.234					
6	*Utricularia vulgaris*	0.136	0.113	0.242	0.161	0.256				
7	*Hygrophila Sp.*	0.156	0.116	0.249	0.147	0.263	0.072			
8	*Potemogeton natans*	0.141	0.107	0.229	0.130	0.245	0.151	0.154		
9	*Lemna Sp.*	0.115	0.019	0.186	0.115	0.203	0.122	0.128	0.119	

Table 2: Maximum Likelihood Correlation Chart of different plant group using rbcL gene sequences

Phylogenetic tree analysis using rbcL sequences assigned the tested plant samples to known species. The plants that were taken into consideration grouped *Spirodela polyrhiza, Limnophyton obtusifolium and Lemna sp.* together belongs to Alismatales. *Spirodela polyrhiza* and *Lemna sp.*which were considered under free floating hydrophytes while *Limnophyton obtusifolium* is rooted emergent hydrophytes. As per chart *Hygroryza aristata* shows lowest value with *Spirodela polyrhiza* and then *Limnophyton_obtusifolium.* Similarly *Spirodela polyrhiza* showed lowest value with *Limnophyton_obtusifolium* and then with *Lemna Sp.* This indicates that this plants are closely related with each other and are grouping together *(Table 2).*

Moreover, *Hygrophila Sp*. and *Limnophyton_obtusifolium* belongs to different families but they are grouped in to Rooted emergent Hydrophyte. In the phylogenic tree *Spirodela polyrhiza* and *Lemna* Sp. showed close relationship. So, they are grouped in free floating hydrophytes. Besides this as per correlation chart *Utricularia vulgaris* shows close relationship with *Lemna Sp. (Table 2)*

Contradictory results occur in *Hygrorhyza aristata* is correlating with lowest value at *Spirodela polyrhiza* but *Spirodela polyrhiza* includes in group of free floating hydrophyte while *Hygrorhyza aristata* is rooted emergent Hydrophyte. *(Table 2)*

Morphological identification is inapplicable while studying population biology. In such cases, barcoding is an efficient and valuable technique. Some ecologists have started using the barcoding approach to identify specific unknown plant samples for practical purposes (Li et al., 2009; Van de Wiel et al., 2009). Ongoing developments of new primers and improvements in sequencing techniques have facilitated the data-emergence process of plant barcoding (Soltis et al., 1996; Plunkett et al., 1997; Van de Wiel et al., 2009; Burgess et al., 2011). Recently, plant diversity belowground was determined using *rbcL* gene sequences as a core plant DNA barcoding marker (Kesanakurti et al., 2011).

Configuration	Count
Identical sites in all three sequences	46
Divergent sites in all three sequences	200
Unique differences in Sequence A	99
Unique differences in Sequence B	98
Unique differences in Sequence C	115

Table. 3 Results from the Tajima's test for 3 Sequences

The equality of evolutionary rate between sequences **A** (*GENG371-14 Hygroryza_aristata*|) and **B** (*GENG372-14 Spirodela_polyrhiza*), with sequence **C** (*GENG373-14 Azolla_pinnata*) used as an out group in Tajima's relative rate test. The χ^2 test statistic was 0.01 (*P* = 0.94320 with 1 degree[s] of freedom) . *P*-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages. The analysis involved 3 nucleotide sequences. Codon positions included were

1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 558 positions in the final dataset (Table 3). Evolutionary analyses were conducted in MEGA5.

	A	T/U	C	G
A	-	*4.66*	*3.74*	**11.66**
T/U	*4.33*	-	**18.48**	*3.52*
C	*4.33*	**23.00**	-	*3.52*
G	**14.36**	*4.66*	*3.74*	-

Table 4: Maximum Likelihood Estimate of Substitution Matrix

Each entry is the probability of substitution (*r*) from one base (row) to another base (column). Substitution pattern and rates were estimated under the Tamura-Nei (1993) model. Rates of different transitional substitutions are shown in **bold** and those of transversionsal substitutions are shown in *italics*. Relative values of instantaneous *r* should be considered when evaluating them. For simplicity, sum of *r* values is made equal to 100, The nucleotide frequencies are A = 26.65%, T/U = 28.67%, C = 23.03%, and G = 21.65%. For estimating ML values, a user-specified toplogy was used. The maximum Log likelihood for this computation was -1804.400. The analysis involved 9 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 424 positions in the final dataset (Table 4). Evolutionary analyses were conducted in MEGA5.

Table. 5 Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameter	BIC	AICc	lnL	Invariant	Gamma	R	Freq A	Freq T	Freq C	Freq G	A=>T	A=>C	A=>G	T=>A	T=>C	T=>G	C=>A	C=>T	C=>G	G=>A	G=>T	G=>C
K2+I	17	3660.6	3554.5	-1760.2	0.553064	n/a	2.4007	0.25	0.25	0.25	0.25	0.04	0.04	0.18	0.04	0.18	0.04	0.04	0.18	0.04	0.18	0.04	0.04
K2+G	17	3661	3555	-1760.4	n/a	0.37971	2.4912	0.25	0.25	0.25	0.25	0.04	0.04	0.18	0.04	0.18	0.04	0.04	0.18	0.04	0.18	0.04	0.04
T92+G	18	3665.5	3553.3	-1758.5	n/a	0.3646	2.5736	0.2766	0.2766	0.2234	0.2234	0.04	0.03	0.16	0.04	0.16	0.03	0.04	0.2	0.03	0.2	0.04	0.03
K2+G+I	18	3667.8	3555.6	-1759.7	0.498922	3.36206	2.4492	0.25	0.25	0.25	0.25	0.04	0.04	0.18	0.04	0.18	0.04	0.04	0.18	0.04	0.18	0.04	0.04
T92+G+I	19	3672.1	3553.6	-1757.7	0.509036	3.48865	2.5266	0.2766	0.2766	0.2234	0.2234	0.04	0.03	0.16	0.04	0.16	0.03	0.04	0.2	0.03	0.2	0.04	0.03
HKY+G	20	3679.6	3554.9	-1757.3	n/a	0.36566	2.557	0.2665	0.2867	0.2303	0.2165	0.04	0.03	0.16	0.04	0.17	0.03	0.04	0.21	0.03	0.19	0.04	0.03
TN93+G	21	3684.4	3553.5	-1755.6	n/a	0.37507	2.5697	0.2665	0.2867	0.2303	0.2165	0.04	0.03	0.12	0.04	0.2	0.03	0.04	0.25	0.03	0.15	0.04	0.03
HKY+G+I	21	3686.3	3555.4	-1756.6	0.505657	3.31966	2.5115	0.2665	0.2867	0.2303	0.2165	0.04	0.03	0.16	0.04	0.17	0.03	0.04	0.21	0.03	0.19	0.04	0.03
TN93+G+I	22	3691.3	3554.1	-1754.9	0.489497	2.8641	2.5266	0.2665	0.2867	0.2303	0.2165	0.04	0.03	0.12	0.04	0.2	0.03	0.04	0.24	0.03	0.15	0.04	0.03
GTR+G	24	3706.4	3556.8	-1754.2	n/a	0.36294	2.58	0.2665	0.2867	0.2303	0.2165	0.03	0.04	0.12	0.02	0.2	0.03	0.05	0.25	0.03	0.15	0.05	0.03
GTR+I	24	3707.2	3557.5	-1754.6	0.555544	n/a	2.4415	0.2665	0.2867	0.2303	0.2165	0.03	0.04	0.13	0.03	0.19	0.04	0.04	0.24	0.03	0.16	0.05	0.03
GTR+G+I	25	3714.6	3558.8	-1754.2	0	0.36294	2.58	0.2665	0.2867	0.2303	0.2165	0.03	0.04	0.12	0.02	0.2	0.03	0.05	0.25	0.03	0.15	0.05	0.03
K2	16	3753.4	3653.6	-1810.7	n/a	n/a	2.0563	0.25	0.25	0.25	0.25	0.04	0.04	0.17	0.04	0.17	0.04	0.04	0.17	0.04	0.17	0.04	0.04
T92	17	3763.5	3657.4	-1811.6	n/a	n/a	2.062	0.2766	0.2766	0.2234	0.2234	0.04	0.04	0.15	0.04	0.15	0.04	0.04	0.19	0.04	0.19	0.04	0.04
T92+I	18	3771.7	3659.5	-1811.6	0.00001	n/a	2.062	0.2766	0.2766	0.2234	0.2234	0.04	0.04	0.15	0.04	0.15	0.04	0.04	0.19	0.04	0.19	0.04	0.04
HKY	19	3777.2	3658.7	-1810.2	n/a	n/a	2.0621	0.2665	0.2867	0.2303	0.2165	0.05	0.04	0.15	0.04	0.16	0.04	0.04	0.19	0.04	0.18	0.05	0.04
TN93	20	3778.9	3654.2	-1807	n/a	n/a	2.0721	0.2665	0.2867	0.2303	0.2165	0.05	0.04	0.12	0.04	0.19	0.04	0.04	0.23	0.04	0.14	0.05	0.04
JC+I	16	3778.9	3679.1	-1823.5	0.539432	n/a	0.5	0.25	0.25	0.25	0.25	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
HKY+I	20	3785.4	3660.7	-1810.2	0.00001	n/a	2.0621	0.2665	0.2867	0.2303	0.2165	0.05	0.04	0.15	0.04	0.16	0.04	0.04	0.19	0.04	0.18	0.05	0.04
JC+G+I	17	3786.8	3680.7	-1823.3	0.504102	5.89873	0.5	0.25	0.25	0.25	0.25	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
TN93+I	21	3787.1	3656.2	-1807	0.00001	n/a	2.0721	0.2665	0.2867	0.2303	0.2165	0.05	0.04	0.12	0.04	0.19	0.04	0.04	0.23	0.04	0.14	0.05	0.04
JC+G	16	3789.3	3689.5	-1828.7	n/a	0.76959	0.5	0.25	0.25	0.25	0.25	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
GTR	23	3803.2	3659.8	-1806.8	n/a	n/a	2.0729	0.2665	0.2867	0.2303	0.2165	0.04	0.04	0.12	0.04	0.19	0.04	0.05	0.23	0.03	0.14	0.05	0.03
JC	15	3861.3	3767.7	-1868.8	n/a	n/a	0.5	0.25	0.25	0.25	0.25	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08

Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the **substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected),** Maximum Likelihood value (*lnL*), and the number of parameters (including branch lengths) are also presented (Tamura K. et. al, 2011). Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+*G*) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+*I*). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/transversion bias (*R*) are shown for each model, as well. They are followed by nucleotide frequencies (*f*) and rates of base substitutions (*r*) for each nucleotide pair. Relative values of instantaneous *r* should be considered when evaluating them. For simplicity, sum of *r* values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. The analysis involved 9 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 424 positions in the final dataset (Table 5). Evolutionary analyses were conducted in MEGA5.

Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

CONCLUSION:

In conclusion, this study provides preliminary assessment data that will be useful for wider application of DNA barcoding in ecological studies of aquatic plants. With the current development of primers, we found that *rbcL* is very useful for the barcoding of plant species. However, further protocol development to enhance clean DNA extraction, PCR amplification strategies, including the development of new primers and local authenticated databases would play important roles in efficient utilization of plant barcoding.

ACKNOWLEDGEMENT

The authors are thankful to Biodiversity Genebank, Gujarat State Biotechnology Mission, Gandhinagar for providing laboratory instruments. Biodiversity Genebank offers support to scientists dedicated to the discovery and will also for providing laboratory instrument as well as help me gain more knowledge. The authors thank Shri Akshaykumar Saxena, IFS- Director, Dr. Snehal Bagatharia, Sector Specialist-Agriculture Biotechnology, GSBTM, Gandhinagar, Dr. Rohan Pandya – Scientist, Mr Arpan Bhatt-Research Associate, Mr. Nitin Savaliya - SRF and an anonymous referee for constructive comments on the manuscript.

REFERENCES:

1. Asahina H, Shinozaki J, Masuda K, Morimitsu Y, et al. (2010). Identification of medicinal *Dendrobium* species by phylogenetic analyses using matK and rbcL sequences. *J. Nat. Med.* 64: 133-138.
2. Bafeel SO, Arif IA, Bakir MA, Khan HA, et al. (2011). Comparative evaluation of PCR success with universal primers of maturase K (matK) and ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit (rbcL) for barcoding of some arid plants. *Plant Omics J.* 4: 195-198.
3. Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, et al. (2011). Discriminating plant species in a local temperate flora using the rbcL+matK DNA barcode. *Method Ecol. Evol.* 2: 333-340.
4. CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U. S. A.* 106: 12794-12797.
5. Chase MW, Salamin N, Wilkinson M, Dunwell JM, et al. (2005). Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360: 1889-1895.
6. de Groot GA, During HJ, Maas JW, Schneider H, et al. (2011). Use of rbcL and trnL-F as a two-locus DNA barcode for identification of NW-European ferns: an ecological perspective. *PLoS One* 6: e16371.
7. Duffy AM, Kelchner SA and Wolf PG (2009). Conservation of selection on matK following an ancient loss of its flanking intron. *Gene* 438: 17-25.
8. Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.
9. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, et al. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* 3: e2802.
10. Graham S (2002). Global estimate of endangered plant species triples. Scientific American. Available at [http://www. scientificamerican.com/article.cfm?id=global-estimate-of-endang]. Accessed January 10, 2012.
11. Hebert PD and Gregory TR (2005). The promise of DNA barcoding for taxonomy. *Syst. Biol.* 54: 852-859.
12. Hollingsworth ML, Andra CA, Forrest LL, Richardson J, et al. (2009). Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol. Ecol. Resour.* 9: 439-457.
13. Joshi K, Chavan P, Warude D and atwardhan B (2004). Molecular markers in herbal drug technology. *Curr. Sci.* 87: 159-165.
14. Kesanakurti PR, Fazekas AJ, Burgess KS, Percy DM, et al. (2011). Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. *Mol. Ecol.* 20: 1289-1302.
15. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, et al. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102: 8369-8374.
16. Kress WJ, Erickson DL, Jones FA, Swenson NG, et al. (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci. U. S. A.* 106: 18621-18626.
17. Li FW, Tan BC, Buchbender V, Moran RC, et al. (2009). Identifying a mysterious aquatic fern gametophyte. *Plant Syst. Evol.* 281: 77-86.

18. Naeem A. Khan A , Cheema H , Khan I and Buerkert A (2010). DNA barcoding for species identification in the Palmae family. Genetics and Molecular Research 13 (4): 10341-10348.

19. Nei M. and Kumar S. (**2000**). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
20. Roy S, Tyagi A, Shukla V, Kumar A, et al. (2010). Universal plant DNA barcode loci may not work in complex groups: a case study with Indian berberis species. *PLoS One* 5: e13674.
21. Soltis DE, Kuzoff RK, Conti E, Gornall R, et al. (1996). matK and rbcL gene sequence data indicate that *Saxifraga* (Saxifragaceae) is polyphyletic. *Am. J. Bot.* 83: 371-382.
22. Spooner DM (2009). DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *Am. J. Bot.* 96: 1177-1189.
23. Tajima F. (**1993**). Simple methods for testing molecular clock hypothesis. *Genetics* **135**:599-607.
24. Tamura K and Nei M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10: 512-526.
25. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S. (**2011**). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* (In Press).
26. Vipan Kumar Sohpal1, Apurba Dey2, Amarpal Singh (2011) Substitution model analysis of human herpes simplex virus using molecular evolutionary genetic analysis Biotechnol. Bioinf. Bioeng. 1(4):451-458.
27. Von Crautlein M, Korpelainen H, Pietilainen M, Rikkinen J (2011) DNA barcoding: a tool for improved taxon identification and detection of species diversity. Biodiversity and Conservation 20: 373–389.