Analysis of Phylogeny and Evolutionary Divergence of RbcL Sequence of *Barleria longiflora* L.f

Baskaran[1], Karthikeyan[2]

P.G and Research Department of Botany, Bishop Heber College, Tiruchirappalli-620017, Tamil Nadu, India

Abstract: *Barleria longiflora L.f belongs to the family Acanthaceae which is endemic distribution. Presence of this plant in Erattamalai hill, located in Tiruchirappalli District of Tamil Nadu, India. The rbcl gene is widely used in systematic studies to resolve divergences at many taxonomic levels. The collected B. longiflora were subjected to sequence analysis of rbcL gene and the same was deposited in Gene Bank and an accession number has been assigned for the same (Accession No. KR861703.1). The data was further analyzed for the construction of neighbor joining tree and to infer the evolutionary divergence among the maximum identical sequences retrieved from NCBI Gene Bank through BLAST search. The results of the search revealed that the rbcL sequence of B. longiflora has not been sequenced earlier from any part of this world. It is obvious that the rbcL sequence o fB. longiflora may be used for the identification of this species reported from any part of the world through BLAST analysis if the identical sequences are submitted to Gene Bank in future.*

Keywords: ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene, *Barleria longiflora*, Erattaimalai hill,

1. Introduction

The large and primarily tropical plant family Acanthaceae contains at least 4000 species, placing it among the top 12 or so most diverse families of flowering plants. The group is part of the order Lamiales, which includes familiar plants such as snapdragons (*Antirrhinum*, Plantaginaceae), mints (Lamiaceae), African violets (Gesneriaceae), and monkey flowers (*Mimulus*, Phyrmaceae). Notably, as a result of the largely tropical distribution of Acanthaceae, species-level diversity (as well as some generic diversity) remains poorly understood, and there is little doubt that many new species remain to be discovered. A wide variety of essential and medicinally active ingredients have been identified and extracted from some members of the family. Many scientists have worked on some members of the family and have discovered some very essential active ingredients present in them [1]. Traditional methods of plant identification 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 is also of limited value because the medicines have varied sources and chemical complexity, which is affected by growth, storage conditions, and harvest times[2].

DNA-based identification (barcoding) is simple, 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)[3], despite ongoing criticism of the feasibility or even necessity of DNA barcoding for general taxonomic purposes (Will et al., 2005; Spooner, 2009)[4,5]. The most common gene used to provide sequence data for plant phylogenetic analyses is the plastid-encoded *rbc*L gene[6,7]. This single copy gene is approximately 1430 base pairs in length, is free from length mutations except at the far 3' end, and has a fairly conservative rate of evolution. The function of the *rbc*L gene is to code for the large subunit of ribulose 1, 5 bisphosphate carboxylase/oxygenase (RUBISCO or RuBPCase). The sequence data of the *rbc*L gene are widely used in the reconstruction of phylogenies throughout the seed plants. However, it is apparent that the ability of *rbc*L to resolve phylogenetic relationships below the family level is often poor[8]. Thus, interest exists in finding other useful DNA regions that evolve faster than does *rbc*L to facilitate lower-level phylogenetic reconstruction. The Consortium for the Barcode of Life (CBOL) plant working group has recommended a standard barcode comprising ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) for the barcoding of all land plants [9].

2. Materials and Methods

Young leaves of the respective species were collected in sterile Ziploc bags and stored at -20 °C until further use.

DNA isolation using NucleoSpin® Plant II Kit (Macherey-Nagel**)** About 100 mg of the tissue is homogenized using liquid nitrogen and the powdered tissue is transferred to a microcentrifuge tube. Four hundred microlitres of buffer PL1 is added and vortexed for 1 minute. Ten microlitres of RNase A solution is added and inverted to mix. The homogenate is incubated at 65°C for 10 minutes. The lysate is transferred to a Nucleospin filter and centrifuged at 11000 x g for 2 minutes. The flow through liquid is collected and the filter is discarded. Four hundred and fifty microlitres of buffer PC is added and mixed well. The solution is transferred to a Nucleospin Plant II column, centrifuged for 1 minute and the flow through liquid is discarded. Four hundred microlitre buffer PW1 is added to the column, centrifuged at 11000 x g for 1 minute and flow though liquid is discarded. Then 700 µl PW2 is added, centrifuged at 11000 x g and flow through liquid is discarded. Finally 200 µl of PW2 is added and centrifuged at 11000 x g for 2 minutes to dry the silica membrane. The column is transferred to a new 1.7 ml tube and 50 µl of buffer PE is

added and incubated at 65°C for 5 minutes. The column is then centrifuged at 11000 x g for 1 minute to elute the DNA. The eluted DNA was stored at 4°C.

Agarose Gel Electrophoresis for DNA Quality check. The quality of the DNA isolated was checked using agarose gel electrophoresis. 1µl of 6X gel-loading buffer (0.25% bromophenol blue, 30% sucrose in TE buffer pH-8.0) was added to 5µl of DNA. The samples were loaded to 0.8% agarose gel prepared in 0.5X TBE (Tris-Borate-EDTA) buffer containing 0.5 µg/ml ethidium bromide. Electrophoresis was performed with 0.5X TBE as electrophoresis buffer at 75 V until bromophenol dye front has migrated to the bottom of the gel. The gels were visualized in a UV transilluminator (Genei) and the image was captured under UV light using Gel documentation system (Bio-Rad) (Figure 1). PCR amplification reactions were carried out in a 20 µl reaction volume which contained 1X Phire PCR buffer (contains 1.5 mM MgCl$_2$), 0.2mM each dNTPs (dATP, dGTP, dCTP and dTTP), 1 µl DNA, 0.2 µl Phire Hotstart II DNA polymerase enzyme, 0.1 mg/ml BSA and 3% DMSO, 0.5M Betaine, 5pM of forward and reverse primers. The PCR amplification was carried out in a PCR thermal cycler (GeneAmp PCR System 9700, Applied Biosystems). Forward primer ATGTCACCACAAACAGAGACTAAAGC and Reverse primer GTAAAATCAAGTCCACCRCG. PCR amplification profile for rbcL 98°C -30 sec, 98 ^{o}C -5 sec, 60 ^{o}C -10 sec and 72 ^{o}C -15 sec for 30 cycles, 72 ^{o}C - 60 sec and stored at 4 ^{o}C.

Agarose Gel electrophoresis of PCR products were checked in 1.2% agarose gels prepared in 0.5X TBE buffer containing 0.5 µg/ml ethidium bromide. 1 µl of 6X loading dye was mixed with 5 µl of PCR products and was loaded and electrophoresis was performed at 75V power supply with 0.5X TBE as electrophoresis buffer for about 1-2 hours, until the bromophenol blue front had migrated to almost the bottom of the gel. The molecular standard used was 2-log DNA ladder (NEB). The gels were visualized in a UV transilluminator (Genei) and the image was captured under UV light using Gel documentation system (Bio-Rad). Consists of two hydrolytic enzymes, Exonuclease I and Shrimp Alkaline Phosphatase (SAP), in a specially formulated buffer for the removal of unwanted primers and dNTPs from a PCR product mixture with no interference in downstream applications. Five micro litres of PCR product is mixed with 2 µl of ExoSAP-IT and incubated at 37°C for 15 minutes followed by enzyme inactivation at 80°C for 15 minutes [10].

Sequencing using BigDye Terminator v3.1 sequencing reaction was done in a PCR thermal cycler (GeneAmp PCR System 9700, Applied Biosystems) [11]. The PCR mix consisted of the following components: PCR Product (ExoSAP treated) - 10-20 ng, Primer - 3.2 pM (either Forward or Reverse) Sequencing Mix - 0.28 µl , 5x Reaction buffer -1.86 µl, Sterile distilled water- make up to 10µl. The sequencing PCR temperature profile consisted of a 1st cycle at 96°C for 2 minutes followed by 30 cycles at 96°C for 30 sec, 50°C for 40 sec and 60°C for 4 minutes for all the primers.

Post Sequencing PCR Clean up Make master mix I of 10µl milli Q and 2 µl 125mM EDTA per reaction. Add 12µl of master mix I to each reaction containing 10µl of reaction contents and are properly mixed. Make master mix II of 2 µl of 3M sodium acetate pH 4.6 and 50 µl of ethanol per reaction. Add 52 µl of master mix II to each reaction. Contents are mixed by inverting. Incubate at room temperature for 30 minutes Spin at 14,000 rpm for 30 minutes decant the supernatant and add 100 µl of 70% ethanol Spin at 14,000 rpm for 20 minutes. Decant the supernatant and repeat 70% ethanol wash Decant the supernatant and air dry the pellet. The cleaned up air dried product was sequenced in ABI 3500 DNA Analyzer (Applied Biosystems). The sequence quality was checked using Sequence Scanner Software v1 (Applied Biosystems). Sequence alignment and required editing of the obtained sequences were carried out using Geneious Pro v5.1. [12].

The sequenced rbcLof *B. longiflora* was deposited in Gene Bank (Accession Number: KR861703.1). Identical sequences of rbcLwere obtained from Gene Bank using BLAST algorithm and of them the sequences with maximum identity (whose Accession Numbers are: KR337263.1, KF890170.1, KF890169.1, KJ667661.1, L01886.1, L12594.1, AB586151.1, AB586149.1, AB586150.1, AM234922.1, KT626791.1, AB586155.1, DQ017800.1, DQ017776.1, DQ017774.1, AF245020.1, DQ017786.1, L02434.2, L01935.2) Evolutionary relationships of taxa .The evolutionary history was inferred using the Neighbor-Joining method [13]. The optimal tree with the sum of branch length = 20.33665948 is shown. (Next to the branches). The evolutionary distances were computed using the Maximum Composite Likelihood method [14] and are in the units of the number of base substitutions per site. The analysis involved 20 nucleotide sequences. There were a total of 697 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [15].

3. Results and Discussion

The plant *Barleria longiflora* L.f belongs to the family Acanthaceae collected from Erattamalai hill. Approximately 700bp DNA was isolated during the quality check through agarose gel electrophoresis method. The gene amplification adopted in the present study yielded enough quantity of DNA for further sequence analysis of rbcL. The rbcLgene had 697bp and the same sequence had been deposited in the Gene Bank with the Accession Number KR861703.1. This study provided an opportunity to utilize rbcL sequence for identification of this species in future. During the BLAST search no sequence matches for this gene could be identified from databases on plant. Hence it may be concluded that the rbcL sequence of *B. longiflora* was a first record for Gene Bank. The estimated value of the shape parameter for the discrete Gamma Distribution is 0.0500. Substitution pattern and rates were estimated under the Tamura-Nei (1993) model (+G) [16]. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+G]). Mean evolutionary rates in these categories were 0.00, 0.00, 0.00, 0.03, 4.97 substitutions per site. The nucleotide frequencies are A = 28.42%, T/U = 28.17%, C = 21.03%, and G = 22.38%. For estimating ML values, a tree topology was automatically computed. The

maximum Log likelihood for this computation was -1466.297. The analysis involved 20 nucleotide sequences. Clustered sequences showed 1794 conserved sites, 335 variable sites, 74 parsim-info sites, 203 singleton sites of 529bp. Further the results revealed that there were 1365 zero-fold sites, 412 two-fold sites and 175 four-fold sites. Of the 529bp sites, 663 sites showed 100 % coverage.

The results of NJ analysis of 697bp fragment of the rbcL gene belonged to *B. longiflora* with the twenty sequences obtained through BLAST showed different branch lengths in the Phenogram. Maximum identical sequences were not available for *B. longiflora* in this NJ analysis (Fig. 1). The evolutionary history was inferred using the NJ method. The optimal tree with the sum of branch length was 20.33. The tree was drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary

distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [17] and were in the units of the number of base substitutions per site.

A detailed estimate of evolutionary divergence of rbcL sequence of with their similar sequences through BLAST search is provided in Table 1. The results on the distance analysis indicated that the overall average for all species *B. longiflora* was 0.026. The maximum evolutionary distance observed between *Barleria prionitis* and *B. acuminata* was 3.865 (Table 1). Therefore, it is concluded that rbcL sequence of *B. longiflora* may be used for the identification of this species reported from any part of the world through BLAST analysis if the identical sequences are submitted to Gene Bank in future.
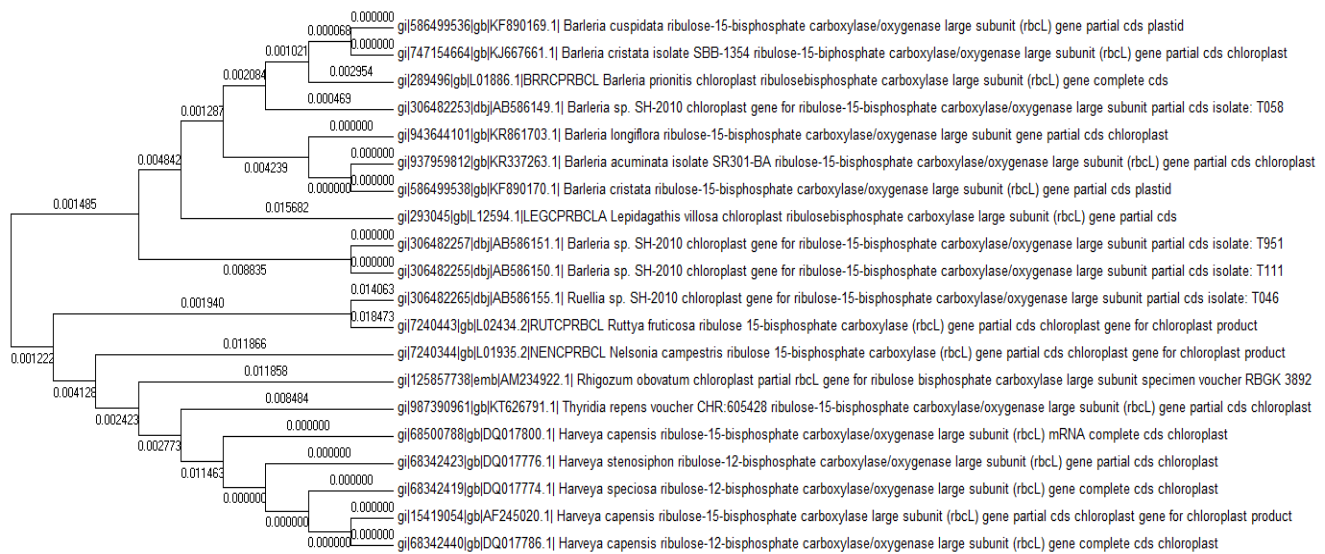
Figure 1: Neighbor-Joining (NJ) analysis of rbcL sequence of *B. longiflora* with their similar sequences through BLAST search

Table 1: Estimate of Evolutionary Divergence of rbcL sequence of *B. longiflora* with their similar sequences through BLAST search

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	*Barleria logiflora* (KR861703.1)																				
2	*Barleria acuminate* (KR337263.1)	0.000																			
3	*Barleria cristata* (KF890170.1)	0.000	0.000																		
4	*Barleria cuspidate* (KF890169.1)	0.008	0.008	0.008																	
5	*Barleria cristata* (KJ667661.1)	0.008	0.008	0.008	0.000																
6	*Barleria prionitis* (L01886.1)	0.011	0.011	0.011	0.003	0.003															
7	*Lepidagathis villosa* (L12594.1)	0.020	0.020	0.020	0.020	0.020	0.023														
8	*Barleria sp.* (AB586151.1)	0.022	0.022	0.022	0.017	0.017	0.020	0.031														

9	*Barleria sp* (AB586149.1)	0.006	0.006	0.006	0.002	0.002	0.005	0.021	0.015											
10	*Barleria sp* (AB586150.1)	0.022	0.022	0.022	0.017	0.017	0.020	0.031	0.000	0.015										
11	*Rhigozum obovatum* (AM234922.1)	0.026	0.026	0.026	0.034	0.034	0.034	0.042	0.033	0.033	0.033									
12	*Thyridia repens* (KT626791.1)	0.031	0.031	0.031	0.033	0.033	0.033	0.042	0.028	0.031	0.028	0.023								
13	*Ruellia sp* (AB586155.1)	0.031	0.031	0.031	0.029	0.029	0.028	0.037	0.028	0.028	0.028	0.037	0.033							
14	*Harveya capensis* (DQ017800.1)	0.031	0.031	0.031	0.036	0.036	0.036	0.045	0.031	0.034	0.031	0.026	0.020	0.040						
15	*Harveya stenosiphon* (DQ017776.1)	0.031	0.031	0.031	0.036	0.036	0.036	0.045	0.031	0.034	0.031	0.026	0.020	0.040	0.000					
16	*Harveya speciosa* (DQ017774.1)	0.031	0.031	0.031	0.036	0.036	0.036	0.045	0.031	0.034	0.031	0.026	0.020	0.040	0.000	0.000				
17	*Harveya capensis* (AF245020.1)	0.031	0.031	0.031	0.036	0.036	0.036	0.045	0.031	0.034	0.031	0.026	0.020	0.040	0.000	0.000	0.000			
18	*Harveya capensis* (DQ017786.1)	0.031	0.031	0.031	0.036	0.036	0.036	0.045	0.031	0.034	0.031	0.026	0.020	0.040	0.000	0.000	0.000	0.000		
19	*Ruttya fruticosa* (L02434.2)	0.033	0.033	0.033	0.034	0.034	0.037	0.041	0.033	0.033	0.033	0.040	0.039	0.033	0.037	0.037	0.037	0.037	0.037	
20	*Nelsonia campestris* (L01935.2)	0.032	0.032	0.032	0.028	0.028	0.028	0.037	0.026	0.029	0.026	0.026	0.022	0.028	0.032	0.032	0.032	0.032	0.032	0.039

4. Acknowledgements

References

[1] Caceres, D. D., Hancke, J. L., Burgos, R. A., Sandberg, F. and Wikman, G. K. 1999. Use of visual analogue scale measurements (VAS) to assess the effectiveL.f of standardized Andrographis paniculata extract SHA-10 in reducing the symptoms of common cold. A randomized double blind-placebo study. Phytomedicine 6: 217–223.

[2] Joshi K, Chavan P, Warude D and atwardhan B (2004). Molecular markers in herbal drug technology. *Curr. Sci.* 87: 159-165.

[3] Hebert PD and Gregory TR (2005). The promise of DNA barcoding for taxonomy. *Syst. Biol.* 54: 852-859.

[4] Will KW, Mishler BD and Wheeler QD (2005). The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54: 844-851.

[5] Spooner DM (2009). DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *Am. J. Bot.* 96: 1177-1189.

[6] Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B.D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y. -L., Kron, K. A., Rettig, J. H., Conti, E., Palmer, J. D., Manhart, J. R., Sytsma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedren, M., Gart, B. S., Jansen, R. K., Kim, K. -J., Wimpee, C. P., Smith, J. F., Furnier, G. R., Strauss, S. H., Xiang, Q. -Y., Plunkett, G. M., Soltis, P. S., Swensen, S. M., Williams, S. E., Gradek, P. A., Quinn, C. J., Eguiarte, L. E., Barrett, S. C. H., Dayanandan, S., And Albert, V. A. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanic Garden* 80: 528-580.

[7] DONOGHUE, M. J., OLMSTEAD, R. G., SMITH, J. F., AND PALMER, J. D. 1993.

[8] Phylogenetic relationships of Dipsacales based on *rbc*L sequences. *Annals of the Missouri Botanic Garden* 79: 333-345.

[9] DOEBLEY, J., M. DURBIN, E. M. GOLENBERG, M. T. CLEGG, AND D. P. MA. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbc*L) nucleotide sequence among the grasses (Gramineae). *Evolution* 44: 1097-1108.

[10] CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U. S. A.* 106: 12794-12797.

[11] ExoSAP-IT – User Guide, GE Healthcare.

[12] BigDye Terminator v3.1 Cycle sequencing Kit – User Manual, Applied Biosystems.

[13] Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T and Wilson A (2010) Geneious v5.1, Available from http://www.geneious.com.

[14] Saitou N. and Nei M. (**1987**). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.

[15] Tamura K., Nei M., and Kumar S. (**2004**). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences* (*USA*) **101**:11030-11035.

[16] 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* 28: 2731-2739.

[17] Tamura K. and Nei M. (**1993**). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**:512-526.

[18] Tamura K., Nei M., and Kumar S. (**2004**). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences* (*USA*) **101**:11030-11035.