MALAYSIAN JOURNAL OF BIOCHEMISTRY & MOLECULAR BIOLOGY

The Official Publication of The Malaysian Society For Biochemistry & Molecular Biology (MSBMB)
http://mjbmb.org

ANALYSIS OF PHYLOGENY AND EVOLUTIONARY DIVERGENCE OF *matK* and *rbcL* SEQUENCE OF *Barleria longiflora* L.f

Baskaran, A.*[1] and Karthikeyan, V.[2]

[1]*Department of Micro Biology, Urumu Dhanalakshmi College, Tiruchirapalli-620019, Tamil Nadu, India*
[2]*Department of Botany, Bishop Heber College, Tiruchirappalli-620017, Tamil Nadu, India*

*Corresponding Author: baskaranbhc@gmail.com

SHORT COMMUNICATION

History
Received: 18th July 2019
Accepted: 9th March 2020

Keywords:

Maturase K (matK), Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL), Barleria longiflora

Abstract

DNA barcodes are short parts of DNA used to identify an unknown specimen of a species. The matK and rbcL gene is widely used in systematic studies to resolve divergences at many taxonomic levels. *Barleria longiflora* L.f belongs to the family Acanthaceae which is endemic in distribution. The collected specimens were subjected to sequence analysis of matK and rbcL gene and the same was deposited in Gene Bank and an accession number has been assigned for the same (Accession No: KR861702.1 and KR861703.1). The data were 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 matK and rbcL sequence of *B. longiflora* has not been sequenced earlier from any part of the world. The matK and 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.

INTRODUCTION

Traditionally, taxonomic identification has relied upon morphological characters. In the last two decades, molecular tools based on DNA sequences of short standardized gene fragments, termed DNA barcodes [1]. DNA barcoding is a widely used molecular-based system, which can identify biological specimens, and is used for the identification of both raw materials and processed food [2]. The purpose of DNA barcoding is nucleotide sequence-based identification of multiple plant species with accuracy and is one of the widely accepted technology [3]. DNA barcoding is currently gaining popularity due to its simplicity and high accuracy as compared to the complexity and subjective biases associated with the morphology-based identification of taxa [4]. The use of DNA barcodes, which are short gene sequences taken from a standardized portion of the genome and used to identify species, is entering a new phase of application as more and more investigations employ these genetic markers to address questions relating to the ecology and evolution of natural systems [5].

Recently, molecular systematics in plants, as well as other organisms, has been widely used for species identification and the determination of phylogenetic relationships. The standard chloroplast DNA barcode for land plants recommended by the Consortium for the Barcode of Life (CBOL) plant working group needs to be evaluated for a wide range of plant species [4]. In plants, chloroplast genes, including the maturase-coding gene (*matK*), the large subunit of ribulose 1,5-bisphosphate carboxylase-coding gene (*rbcL*), the non-coding plastid trnH-psbA intergenic spacer region and encoding subunit B of light-independent protochlorophyllide reductase (chlb), are usually used for molecular phylogenetic analyses [6].

The Consortium for the Barcode of Life (CBOL), highly recommends the use of portions of two plastid coding regions, *rbcL* and *matK*, taken together, as a barcode for plants. One of themost widely used regions for plant barcoding is rbcL (Ribulose-1, 5–bisphosphate carboxylase/ oxygenase large subunit gene) which is responsible for the production of the large subunit of the enzyme RuBisCo (important for carbon fixation) and *matK* is one of the most rapidly evolving coding portions of the chloroplast genome. The high discrimination power at family, genus, and even species levels recommends *matK* as one of the most versatile candidates for barcoding [7].

MATERIALS AND METHODS

Tissue sampling and storage

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 was transferred to a microcentrifuge tube. 400 µl of buffer PL1 was added and vortexed for 1 minute. 10 µl of RNase A solution was added and inverted to mix. For ten minutes at 65ºC the homogenate was incubated. Then the lysate was transferred to a nucleospin filter and centrifuged at 11000 x g for 2 minutes. The flow-through liquid was collected, and the filter was discarded. Four hundred and fifty microlitres of buffer PC was added and mixed well. The solution was transferred to a nucleospin plant II column, centrifuged for 1 minute and the flow-through liquid was discarded. Four hundred microlitre buffer PW1 was added to the column, centrifuged at 11000 x g for 1 minute and flow though liquid was discarded. Then 700 µl PW2 was added, centrifuged at 11000 x g and flow-through liquid was discarded. To dry the silica membrane 200 µl of PW2 was added and centrifuged at 11000 x g for 2 minutes. The column was transferred to a new 1.7 ml tube and 50 µl of buffer PE was added and incubated at 65ºC for 5 minutes. The column was then centrifuged at 11000 x g for 1 minute to elute the DNA. The eluted DNA was stored at 4ºC [8].

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. At 75V electrophoresis was performed with 0.5X TBE as electrophoresis buffer until bromophenol dye front has migrated to the bottom of the gel. In a UV transilluminator (Genei) the gels were pictured, and the image was captured under UV light using Gel documentation system (Bio-Rad) (**Figure 1**).

PCR Analysis

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).

Target	Primer Name	Direction	Sequence (5' → 3')
matK	390f	Forward	CGATCTATTCATTCAATATTTC
	1326r	Reverse	TCTAGCACACGAAAGTCGAAGT
*rbc*L	rbcLa_f	Forward	ATGTCACCACAAACAGAGACTAAAGC
	rbcL724_rev	Reverse	GTAAAATCAAGTCCACCRCG

Agarose Gel electrophoresis of PCR products

In 1.2% agarose gels prepared in 0.5X TBE buffer containing 0.5 µg/ml ethidium bromide the PCR products were checked. 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).

ExoSAP-IT Treatment

ExoSAP-IT (GE Healthcare) consists of two hydrolytic enzymes namely, exonuclease I and shrimp alkaline phosphatase (SAP), in a specially formulated buffer for the removal of dNTPs and unwanted primers from a PCR product mixture with no interference in downstream applications. Five micro litres of PCR product were mixed with 2 µl of ExoSAP-IT and incubated at 37ºC for 15 minutes followed by enzyme inactivation at 80ºC for 15 minutes.

Sequencing using BigDye Terminator v3.1

The sequencing reaction was done in a PCR thermal cycler (GeneAmp PCR System 9700, Applied Biosystems) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) following manufactures protocol.

The PCR mix consisted of the following components:

PCR Product (ExoSAP treated)	: 10-20 ng
Primer (Forward or Reverse)	: 3.2 pM each
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 96ºC for 2 minutes followed by 30 cycles at 96ºC for 30 sec, 50ºC for 40 sec and 60ºC for 4 minutes for all the primers.

Post Sequencing PCR Clean up

The master mix I of 10µl Milli-Q and 2 µl 125mM EDTA per reaction was made and 12µl of master mix I was added to each reaction containing 10µl of reaction contents and are properly mixed. Master mix II of 2 µl of 3M sodium acetate pH 4.6 and 50 µl of ethanol per reaction was made. 52 µl of master mix II was added to each reaction. Contents are mixed by inverting. It was incubated at room temperature for 30 minutes. The content was centrifuged at 14,000 rpm for 30 minutes. The supernatant was decanted and added 100 µl of 70% ethanol. It was centrifuged at 14,000 rpm for 20 minutes. The supernatant was decanted and washed 70% ethanol. Finally, the supernatant was decanted, and it was air-dried. The cleaned-up air-dried product was sequenced in ABI 3500 DNA Analyzer (Applied Biosystems).

Sequence Analysis

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.

RESULTS AND DISCUSSION

The plant *Barleria longiflora* L.f belongs to the family Acanthaceae. Approximately 1000bp DNA was isolated during the quality check through the agarose gel electrophoresis method (**Figure 1**). The gene amplification adopted in the present study

yielded enough quantity of DNA for further sequence analysis of *matK* and *rbcL*. The *matK* gene had 898bp and the *rbcL* gene had 697bp and the same sequence had been deposited in the Gene Bank with the Accession Number KR861702.1 and KR861703.1. This study provided an opportunity to utilize *matK* and *rbcL* sequences for the identification of this species in the 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 *matK* and *rbcL* sequence of *B. longiflora* was a first record for Gene Bank.

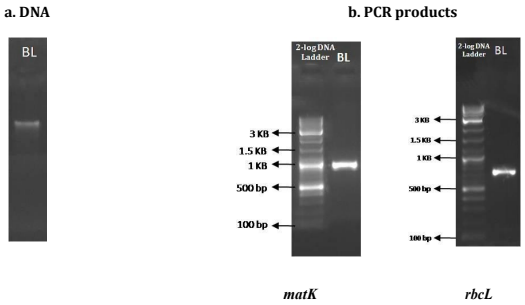

Figure 1. Agarose gel electrophoresis for DNA quality and PCR products

The results of Neighbor-Joining method (NJ) analysis of 898bp fragment of the *matK* and 697bp fragment of the *rbcL* gene belonged to *B. longiflora* with the nine sequences obtained through BLAST showed different branch lengths in the Phenogram. Maximum identical sequences were not available for *B. longiflora* in this NJ analysis in both matK and rbcL gene. (**Figures 2 & 3**).

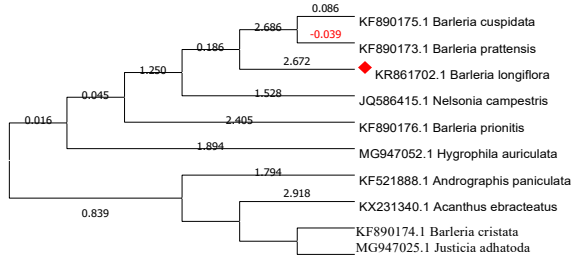

Figure 2. Neighbor-Joining (NJ) analysis of *matK* sequence of *B. longiflora* with their similar sequences through BLAST search.

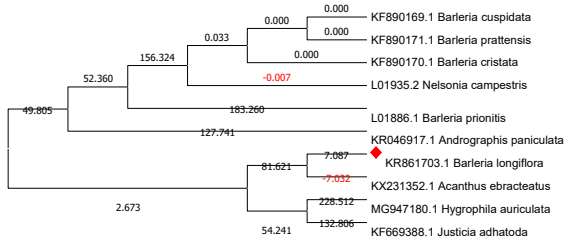

Figure 3. Neighbor-Joining (NJ) analysis of **rbcL** sequence of *B. longiflora* with their similar sequences through BLAST search

The evolutionary history was inferred using the Neighbor-Joining method [9]. The optimal tree with the sum of branch length of *matK* = 22.53425277 and *rbcL* = 1069.42380238 was shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [10]. The evolutionary distances were computed using the Maximum Composite Likelihood method [11] and are in the units of the number of base substitutions per site. The analysis involved 10 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There was a total of 70 positions in the final dataset in *matK* and 40 in *rbcL*. Evolutionary analyses were conducted in MEGA6 [12].

A detailed estimate of evolutionary divergence of *matK* and *rbcL* sequence of with their similar sequences through BLAST search is provided in Tables 1 and 2. The results on the distance analysis indicated that the overall average for all species *B. longiflora* in *matK* was 5.905 and *rbcL* was 323.245. The maximum evolutionary distance observed between *Barleria longiflora* L.f and *Acanthus ebracteatus* was 11.165 in matK and 543.525 in *rbcL* (**Table 1 and 2**). Therefore, it is concluded that *matK* and *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.

Table 1. Estimates of Evolutionary Divergence of matK sequence of *B.longiflora* with their similar sequences through BLAST search

1	Barleria longiflora									
2	Barleria cristata	5.563								
3	Barleria cuspidate	5.379	5.842							
4	Barleria prionitis	6.890	6.773	7.903						
5	Barleria prattensis	5.382	6.070	0.046	7.790					
6	Acanthus ebracteatus	8.028	5.184	11.377	7.204	11.165				
7	Hygrophila auriculata	6.992	3.663	5.567	4.302	5.858	5.198			
8	Justicia adhatoda	8.412	3.427	6.486	4.792	5.198	4.712	6.635		
9	Nelsonia campestris	4.524	7.152	4.198	4.346	4.373	7.165	4.511	7.907	
10	Andrographis paniculata	7.284	4.054	7.079	4.980	6.996	5.109	5.104	4.852	4.248

Table 2. Estimates of Evolutionary Divergence of rbcL sequence of *B.longiflora* with their similar sequences through BLAST search

1	*Barleria longiflora*									
2	*Barleria cristata*	291.133								
3	*Barleria cuspidata*	291.133	0.000							
4	*Barleria prionitis*	543.525	339.607	339.607						
5	*Barleria prattensis*	291.133	0.000	0.000	339.607					
6	*Acanthus ebracteatus*	0.055	287.407	287.407	543.525	287.407				
7	*Hygrophila auriculata*	446.405	539.157	539.157	538.314	539.157	346.428			
8	*Justicia adhatoda*	235.021	500.711	500.711	398.483	500.711	238.338	361.318		
9	*Nelsonia campestris*	287.407	0.026	0.026	339.607	0.026	238.718	539.157	507.485	
10	*Andrographis paniculata*	306.765	351.346	351.346	303.771	351.346	305.340	482.618	259.249	351.34

REFERENCE

1. Fišer Pečnikar, Ž., & Buzan, E. V. (2014). 20 Years Since The Introduction Of DNA Barcoding: From Theory To Application. *Journal of Applied Genetics*, *55*(1), 43–52. doi:10.1007/s13353-013-0180-y.

2. Galimberti, A., De Mattia, F., Losa, A., Bruni, I., Federici, S., Casiraghi, M., Labra, M. (2013). DNA Barcoding As A New Tool For Food Traceability. *Food Research International*, *50*(1), 55–63. doi:10.1016/j.foodres.2012.09.036.
3. Shinwar, Z. K., Jan1, S. A., Khalil, A. T., Khan, A., Ali, M., Qaiser, M., & Zahra, N. B. (2018). Identification And Phylogenetic Analysis Of Selected Medicinal Plant Species From Pakistan: DNA Barcoding Approach. *Pak. J. Bot., 50*(2), 553–560.
4. Bafeel, S. O., Arif, I. A., Bakir, M. A., Al Homaidan, A. A., Al Farhan, A. H., & Khan, H. (2012). DNA Barcoding Of Arid Wild Plants Using Rbcl Gene Sequences. *Genetics and Molecular Research : GMR*, *11*(3), 1934–1941. doi:10.4238/2012.July.19.12.
5. Kress, W. J., Robledo, C. G., Uriarte, M., L, D., & Erickson. (2014). DNA Barcodes for Ecology, Evolution, And Conservation. *Trends in Ecology & Evolution*, 1–11. doi:10.1016/j.tree.2014.10.008.
6. Zhao, Y. S., Xie, L. X., Mao, F. Y., Cao, Z., Wen Wang, P., Zhao, Q. P., & Zhang, X. H. (2016). Identification Of Ephedra Species By Phylogenetic Analyses Using Matk And ITS1 Sequences. *Acta Societatis Botanicorum Poloniae*, 1–9. doi:10.5586/asbp.3497.
7. Mohamed, A. (2016). DNA Barcoding of Five Medicinal Plants from Siwa Oasis, Egypt. *KMITL Sci. Tech. J.*, *16*(2), 89–96.
8. Osathanunkul, M., Suwannapoom, C., Osathanunkul, K., Madesis, P., & De Boer, H. (2016). Evaluation of DNA Barcoding Coupled High Resolution Melting for Discrimination of Closely Related Species in Phytopharmaceuticals. *Phytomedicine*, *23*(2), 156–165. doi: 10.1016/j.phymed.2015.11.018.
9. Saitou, N., & M, N. (1987). The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, *4*, 406–425.
10. Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*, *39*, 783–791.
11. Tamura, K., M., N., & S, K. (2004). Prospects for Inferring Very Large Phylogenies by Using The Neighbor-Joining Method. *Proceedings of the National Academy of Sciences*, (101), 11030–11035.
12. Tamura, K., G., S., D., P., A., F., & S, K. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, (30), 2725–2729.