



## Molecular identification of four medicinal plants using DNA barcoding approach from Chittagong, Bangladesh

Sabrina Amin<sup>1</sup>, Srabasti Ghosh<sup>1</sup>, Baishakhi Biswas<sup>1</sup>, Md. Arifuzzaman<sup>1</sup>, Md. Abul Kalam Azad<sup>1,\*</sup>, AMAM Zonaed Siddiki<sup>2,3</sup>

<sup>1</sup>Department of Biochemistry and Biotechnology, Faculty of Basic Medical and Pharmaceutical Sciences, University of Science and Technology Chittagong (USTC), Foy's Lake, Khulshi-4202, Chittagong, Bangladesh.

<sup>2</sup>Genomics Research Group, Chittagong Veterinary and Animal Science University, Khulshi -4225, Chittagong, Bangladesh.

<sup>3</sup>Nextgen Informatics Pvt Ltd.

\*Corresponding author: Md. Abul Kalam Azad, Department of Biochemistry and Biotechnology, Faculty of Basic Medical and Pharmaceutical Sciences, University of Science and Technology Chittagong (USTC), Foy's Lake, Khulshi-4202, Chittagong, Bangladesh. Phone: +8801836930655, Email: [akazadbtech@gmail.com](mailto:akazadbtech@gmail.com).

**Academic Editor:** Dr. Shahed Uddin Ahmed Shazib, Smith College & University, Massachusetts, USA.

Received: 12 June 2020; Accepted: 02 August 2020; Published: 11 August 2020.

**ABSTRACT:** Accurate identification of important plants is essential for their safety, efficacy and herbal remedies. The study was aimed to identify 4 locally available medicinal plants using DNA barcoding approach. Genomic DNA was extracted from plant samples followed by their amplification by the conventional PCR approach. Short sequence diversity of standardized specific coding gene regions of *matK* gene of plastid genome was used to compare and differentiate the plant species. Subsequently, all the samples were purified and sequenced successfully. A phylogenetic tree was constructed to assess their cross-species relationship. All the samples showed a high similarity rate with their homologs after blasting them in NCBI database. The phylogenetic study showed a distinguished relationship with each other. All the result indicates that DNA barcoding approach could be successfully used for reliable identification of medicinal plants and *matK* gene is a good candidate for this approach.

**KEYWORDS:** DNA barcode, medicinal plant, *matK* gene, species identification, phylogeny

### INTRODUCTION

Reliable identification of any species is crucial for monitoring large-scale bio-diversity and conservation [1]. DNA-barcode is a short DNA-sequence that identifies a species, by comparing the sequence of an unknown specimen to barcodes in a sequence database of known species [2]. The purpose of DNA barcoding is to authenticate known species by matching their sequences to the assembled reference libraries of barcode sequences as well as to facilitate the discovery of novel species. Since mid-Nineties, DNA technology has been widely used for the identification of medicinal plants [3]. Currently, four standard DNA barcodes (*rbcL*, *matK*, *trnH-psbA* and ITS) have been extensively used in the identification of plant species in DNA barcoding method. Similar to their morphological, ecological and behavioral differences show differences in their DNA sequences [4]. It is generally accepted that *rbcL* and *matK* gene fragment can be used as standard barcodes

with ITS gene fragment and *trnH-psbA* gene fragment [5, 6].

In recent years the *matK* coding region is one of the most advanced regions in chloroplasts that shows a high level of species discrimination among angiosperms [7, 8]. The *matK* region is commonly used to identify medicinal materials of different geographical origins for its high variability properties.

Drug discovery from medicinal plants is very promising and about 60% of the antitumor and anticancer drugs have been derived from natural products [5, 9]. The identification of medicinal plants by using molecular markers has recently been demonstrated on a large scale [7]. It is possible to derive adulterants from herbal medicinal materials from closely related plant species. The correct identification of medicinal plants is very crucial for their safe and proper use in novel drug discovery [10].

Thus, the study was designed to identify local medicinally important plants using a growingly popular DNA barcode method.

## MATERIALS AND METHODS

### Collection of plant

A total of four different plant specimens known as *Azadirachta indica*, *Justicia adhatoda*, *Calotropis procera* and *Mikania scandens* were collected during their reproductive growth stage from Panchlaish area of Chittagong in Bangladesh (22°22'N 91°49.5'E).

### Tissue sample preparation

Young and healthy leaves from the selected plant specimens were used for the preparation of samples. The samples were prepared with the proper sample ID after collection (Table 1). Leaves were properly cleaned and kept at room temperature in sealed plastic packs [11]. The specimens were later used for DNA extraction at molecular biology lab.

**Table 1.** Accession number of submitted sequences in NCBI/GeneBank

Sample No.	Species	Accession no
01	<i>Azadirachta indica</i>	MK204499
02	<i>Justicia adhatoda</i>	MK204500
03	<i>Calotropis procera</i>	MK204501
04	<i>Mikania scandens</i>	MK204502

### DNA extraction

DNA was extracted from a small amount of leaf tissue sample using the CTAB (cetyl trimethylammonium bromide) extraction protocol by incubating it at 65°C for 2 hours [12]. An equal volume of chloroform and isoamyl alcohol were used during extraction and it was centrifuged at 10,000 rpm for 10 minutes. The aqueous layer was transferred to another microcentrifuge tube and added 2 volumes of cold absolute ethanol. After keeping it at -20°C for 2-3 hours it was spun at 10,000 rpm for 10 minutes at 4°C. Finally, the pellet was washed with 80% ethanol followed by centrifugation at 10,000 rpm for 2 minutes. The pellet was then collected and dried and added with 20µl nuclease-free water.

### PCR and sequencing

The primer used for amplification of *matK* gene was as follows: 3F-KIM (5'-CGTACAGTACTTTTGTGTTTACGAG-3') and 1R-KIM (5' ACCCAGTCCATCTGGAAATCTTGGTTC-3')[13]. The *matK* gene fragment was amplified following the plant protocol of CCDB[7, 14]. Using thermal cycler (2720 Thermal cycler, Applied Biosystems, USA), PCR was carried out with the above primer. The PCR mixture contained 2 µl plant genomic DNA, 5 µl Green Master mix, 2X (Go Taq® G2 Hot Start Version, Cat No. M7424, USA), 1 µl each of forward and reverse primers (10 pmol) and 1µl nuclease-free water. PCR in a reaction mixture of 10 µl was prepared with the PCR thermal profile as 95°C for 2 min, 95 °C for 30 sec; 50 °C for 1min 30 sec; 72 °C for 40 sec for 45 cycles and a final extension at 72 °C for 5 min. [15]. According to CCDB(Canadian Centre for DNA Barcoding) protocol, diluted PCR replicons used directly for sequencing [16]. The purified PCR products were sequenced bi-directionally using automated DNA sequencer (ABI 3730 xl, Hitachi Ltd., USA) by Bioneer Sequencing Service (South Korea) where the same primers were used as like as PCR.

### Sequence editing, alignment and phylogeny analysis

Once the sequences were available, Bioedit software was used to align, compare and refine the sequences. Multiple sequence alignment was done by using MEGA[17]. A phylogenetic tree was constructed using MEGA7 tools to understand their evolutionary relationship (Figure 2). The sequences were submitted at the NCBI (National Center for Biotechnology Information) database using BankIt submission tool.

## RESULTS

### PCR amplification

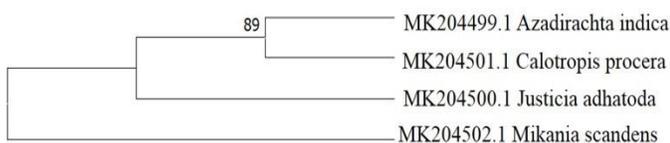
Amplification of the *matK* DNA barcode regions of 4 samples with universal primers showed 100% PCR performance. All the samples showed a clear band at 800 bp length (Figure 3). It indicates that *matK* region of all the samples was amplified.



**Figure 1.** Physical appearance of nine medicinal plants used in this study. a) *Azadirachta indica*, b) *Justicia adhatoda*, c) *Calotropis procera*, d) *Mikania scandens*

### Sequencing success

The *matK* sequences from reverse primer showed good quality sequences at the beginning and poor-quality sequence at the end. That means *matK* reverse sequence chromatograms from reverse primer had sharp peaks at the beginning and considered reverse sequences for barcoding analysis. In this study, reverse sequences of each sample were bi-directionally sequenced and annotated with Sequence Scanner v1.0 and Bioedit. The sequences originated from this study were submitted at the NCBI database with the accession numbers MK204499, MK204500, MK204501, and MK204502 (Table 1).



**Figure 2.** Evolutionary relationships of taxa.

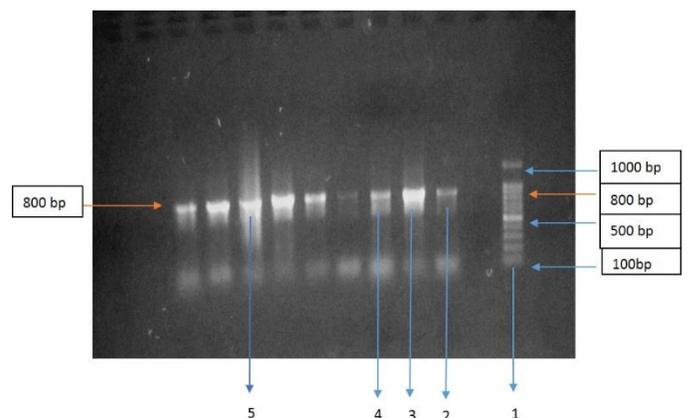
### Phylogenetic analysis and BLAST matching

The comparative study of the molecular sequence data is important for the reconstruction of plant evolutionary history. In this regard, a Neighbor-Joining (NJ) phylogenetic tree was constructed by MEGA7 (Figure 2). The analysis involved 4 nucleotide sequences. From the phylogenetic analysis, we have found that the NJ tree had 3 branches where *Azadirachta indica* and *Calotropis procera* clustering with *Justicia adhatoda*. In other

branches *Mikania scandens* and *Justicia adhatoda* did not show any clustering with other species. The cladding developed in the trees was mainly a multi-plant mixture. The interaction and differences between plants and the big and small divisions are clear between the plant species.

To find out sequence similarities and their taxonomic confirmation, BLASTn tool was used from the NCBI database. From the result, *matK* sequence of sample *Azadirachta indica* showed 99% similarity with *Azadirachta indica* (Accession: EF489115). *Justicia adhatoda* showed 94% identity with *Justicia campylostemon* database (NCBI) sequence from Southern Africa (Accession: JX518170). *Calotropis procera* showed 88% nearest similarity with *Calotropis procera* database (NCBI) sequence (Accession: KT344854) and *Mikania scandens* showed 98% nearest similarity with *Dyscritothamnus mirandae* database (NCBI) sequence (Accession: AY215786). This percentage of identity from the Blast result confirmed the taxonomic identity of all the 4 samples.

All the results of analyses strongly suggest that 4 different medicinal plant species from the same Kingdom (Plantae) have *matK* gene in their chloroplast genome and *matK* can be used as a common barcoding primer for these 4 different medicinal plants. It also indicates that all the 4 medicinal plants were successfully identified by DNA barcoding method.



**Figure 3.** Partially amplified *matK* sequences by one pair of barcode primer by PCR on 1% agarose gel. Lane 1: 100 bp DNA ladder; Lane 2: *matK* genome of *Azadirachta indica*; Lane 3: *matK* genome of *Mikania scandens*; Lane 4: *matK* genome of *Calotropis procera*;

### DISCUSSION

Traditional taxonomic methods to identify medicinal plants and their fruits or leaves to treat various diseases are quite useful, but the herbal industry suffers from the adulteration and reliable identification of medicinal

herbs. This is mostly due to the existence of closely related species and the lack of proper taxonomic identification. DNA barcoding is considered a genetic and bioinformatics method to classify and recognize plant species at the level of molecular taxonomy. For this study we selected 4 locally available plants. Though the selected plants are very familiar and widely used in

Bangladesh, but DNA barcoding method may help to distinguish local species from other globally available species. It's also may use to see their evolutionary relationship. Above all, this study may show a way for the identification of other locally important plants using DNA barcoding approach which are unknown.

**Table 2.** BLAST output for *matK* gene of selected medicinal plants

Query	Query coverage	E- value	Max. identity	Accession	Homology with	Score
<i>Azadirachta indica</i>	100%	0.0	99%	EF489115	matk gene of <i>A. indica</i>	1415
<i>Justicia adhatoda</i>	94%	1e-168	89%	JX518170	matk gene of <i>Justicia campylostemon</i>	603
<i>Calotropis procera</i>	99%	0.0	88%	KT344854	matk gene of <i>Calotropis procera</i>	645
<i>Mikania scandens</i>	100%	0.0	98%	AY215786	matk gene of <i>Dyscritothammus mirandae</i>	1480

The *matK* gene is one of the most quickly developing plastid coding regions and has been extensively studied [18, 19].

Chloroplast's *matK* gene is highly conserved in plants. In this study, *matK* gene was studied to identify these local medicinal plant species. In this regard, all the four extracted DNA from the samples were performed PCR for amplification. In Figure 3, all 4 nucleotides showing a clear band at 800 bp length that indicates that all the samples are successfully amplified. After successfully sequencing, to analyze their evolutionary relationship a neighbor-joining tree was constructed (Figure 2). The evolutionary history was derived by the Neighbor-Joining method [20]. The standard bootstrap tree derived from 1000 replicates [21]. The branch length of the tree was 0.67193986. Branches referring to partitions replicated in fewer than 50 percent of bootstrap replicates is collapsed. Next to the divisions, the number of duplicate trees is shown in which the associated taxa clustered together in the bootstrap analysis (1000) replicate [21]. Using the Maximum Composite Likelihood method, the evolutionary distances were estimated and are in the units of the number of base substitutions per site [22]. 4 nucleotide sequences were involved in the study. The inclusive Codon positions were 1st+2nd+3rd+Noncoding. All positions which contain gaps and missing data were removed. The final dataset had a total of 383 positions. This whole phylogenetic analysis was conducted in MEGA7 [23]. From the phylogenetic tree, it is clearly seen that only *Azadirachta indica* and *Calotropis procera* clustering with *Justicia adhatoda*. Interestingly, *Justicia adhatoda* and *Mikania scandens* did not clustered with others. So, the result showed that *matK* could efficiently distinguish all the plants.

Blastn tool was used to find the nearest similarity with the same or different species. From the Blastn result, it

showed a high rate of similarity with its homologous sequence (Table 2). The highest similarity was found in *Azadirachta indica* (99%) and the lowest one was *Calotropis procera* that showed 88% of similarity with its homologs. This Blastn result confirmed the taxonomical accuracy of all the samples.

## CONCLUSIONS

Based on sequence analysis and blast result, it indicates that the DNA barcoding approach is a good application for the identification of these locally available medicinal plant species. Moreover, species identification in this study will help to develop further experiments with other important medicinal plants inside Bangladesh.

## ACKNOWLEDGEMENT

We would like to show our gratitude to Dr. Nurul Absar, Head of the Department of Biochemistry and Biotechnology, University of Science and Technology Chittagong (USTC) for sharing his pearls of wisdom and support us during this research.

## AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Sabrina Amin, Srabasti Ghosh, Baishakhi Biswas, Md. Arifuzzaman, Md. Abul Kalam Azad, and AMAM Zonaed Siddiki. The first draft of the manuscript was written by [Md. Abul Kalam Azad] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## FUNDING

This research received no external funding.

## CONFLICTS OF INTEREST

Authors declared that they have no conflict of interest.

## REFERENCES

- [1] DeSalle R, Amato GJNRG. The expansion of conservation genetics. 2004;5:702-12.
- [2] Stoeckle MJB. Taxonomy, DNA, and the bar code of life. 2003;53:796-7.
- [3] Li M, Cao H, BUT PPH, SHAW PCJJoS, Evolution. Identification of herbal medicinal materials using DNA barcodes. 2011;49:271-83.
- [4] Wilson JJ, Sing KW, Lee PS, Wee AKJCB. Application of DNA barcodes in wildlife conservation in Tropical East Asia. 2016;30:982-9.
- [5] Group CPW, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, et al. A DNA barcode for land plants. 2009;106:12794-7.
- [6] Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DHJPotNAoS. Use of DNA barcodes to identify flowering plants. 2005;102:8369-74.
- [7] Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, et al. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. 2008;3.
- [8] Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, et al. DNA barcoding the floras of biodiversity hotspots. 2008;105:2923-8.
- [9] Cao H, Liu Y, Cai J, Wang Z, Xu LJZyxzz. Correlative analysis between geographical distribution and nucleotide sequence of chloroplast matK gene of *Cnidium monnieri* fruit in China. 2001;36:373-6.
- [10] Chen F, Chan H-YE, Wong K-L, Wang J, Yu M-T, But PP-H, et al. Authentication of *Saussurea lappa*, an endangered medicinal material, by ITS DNA and 5S rRNA sequencing. 2008;74:889-92.
- [11] Pathak MR, Mohamed AA, Farooq MJAJoPS. DNA Barcoding and Identification of Medicinal Plants in the Kingdom of Bahrain. 2018;9:2757-74.
- [12] Gurib-Fakim AJMaoM. Medicinal plants: traditions of yesterday and drugs of tomorrow. 2006;27:1-93.
- [13] Costion C, Ford A, Cross H, Crayn D, Harrington M, Lowe AJPo. Plant DNA barcodes can accurately estimate species richness in poorly known floras. 2011;6:e26841.
- [14] Kuzmina M, Ivanova N. PCR Amplification for Plants and Fungi. CCDB Protocols. 2016.
- [15] Ivanova NV, Kuzmina MLJMer. Protocols for dry DNA storage and shipment at room temperature. 2013;13:890-8.
- [16] Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, et al. Identification of Amazonian trees with DNA barcodes. 2009;4.
- [17] Edgar RCJNar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. 2004;32:1792-7.
- [18] Asahina H, Shinozaki J, Masuda K, Morimitsu Y, Satake MJJonm. Identification of medicinal *Dendrobium* species by phylogenetic analyses using matK and rbcL sequences. 2010;64:133-8.
- [19] Mahadani P, Sharma GD, Ghosh SKJPM. Identification of ethnomedicinal plants (Rauvolfioideae: Apocynaceae) through DNA barcoding from northeast India. 2013;9:255.
- [20] Saitou N, Nei MJMb, evolution. The neighbor-joining method: a new method for reconstructing phylogenetic trees. 1987;4:406-25.
- [21] Dopazo JJJome. Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. 1994;38:300-4.
- [22] Tamura K, Nei M, Kumar SJPotNAoS. Prospects for inferring very large phylogenies by using the neighbor-joining method. 2004;101:11030-5.
- [23] Kumar S, Stecher G, Tamura KJMb, evolution. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. 2016;33:1870-4.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.