Publications, Seminar and Workshop

LIST OF PUBLICATIONS:

1. **P Mahadani** and S K Ghosh (2013) DNA Barcoding: A tool for species identification from herbal juices, DNA Barcodes (DOI: 10.2478/dna-2013-0002). **ISSN**: 2299-1077.
2. B A Laskar; M J Bhattacharjee; B Dhar, **P Mahadani**, S Kundu, S K Ghosh (2013) The Species dilemma of Northeast Indian Mahseer (Actinopterygii: Cyprinidae): DNA barcoding in clarifying the riddle. PloS One . *e*ISSN: 1932-6203.
3. **P Mahadani,** G D Sharma, S K Ghosh (2013) Identification of Ethno-medicinal Rauvolfioideae (Apocynaceae) plants through DNA barocding from Northeast India. Pharmacognosy Magazine (http://www.phcog.com/aheadofprint.asp). ISSN: 0973-1296. (**In press**)
4. **P Mahadani** M M Das, S K Ghosh (2013) *matK* sequence based plant DNA barcoding failed to identify *Bambusa* (family: Poaceae) species from Northeast India. Journal of Environment & Sociobiology. ISSN: 0973-0834. (**In press**)
5. **P Mahadani**, A Chetry, and P R Ghosh S K Ghosh, (2013) DNA Barcode of Royal Bengal Tiger (*Panthera tigris*) and Domestic Cat (*Felis sylvestris catus*) using own designed PCR Primers Journal of Environment & Sociobiology. ISSN: 0973-0834 (**In press**).

Communicated Papers:

1. Utility of indels for species level identification in complex plant group: A study with an intergenic spacer in Citrus (2013) **P Mahadani** and S K Ghosh (Molecular Biology Report).

Book Chapters:

1. **Pradosh Mahadani,** Ksh Miranda Devi, Mridul M Das, Mohua Chakraborty, Fazlur Rahman , Jagadish Hansa , and Sankar Kumar Ghosh (2012) **BIOINFORMATICS IN DNA BARCODING.** Pp. 105-136. In: A TEXT BOOK ON DNA BARCODING (ed. Ghosh, S.K.), Books Space, Kolkata. IBSN: 81-922989-4-8.

2. Sankar Kumar Ghosh, M Joyraj Bhattacharjee, **Pradosh Mahadani** and Boni Amin Laskar (2012) **FUNDAMENTAL OF DNA BAROCDING.** Pp. 17-50. In: A TEXT BOOK ON DNA BARCODING (ed. Ghosh, S.K.), Books Space, Kolkata. IBSN: 81-922989-4-8**.**

3. M Joyraj Bhattacharjee, Shantanu Kundu, Ksh Miranda Devi, Bishal Dhar, Rosy Mondal, **Pradosh Mahadani,** Monika Anthem and Sankar Kumar Ghosh (2012) **MOLECULAR BIOLOGY IN DNA BAROCDING.** Pp. 71-104. In: A TEXT BOOK ON DNA BARCODING (ed. Ghosh, S.K.), Books Space, Kolkata. IBSN: 81-922989-4-8**.**

4. **Pradosh Mahadani** and Sankar K Ghosh (2012) **Concept of DNA Passport Data of Medicinal Plants**. Pp. 241-251. In: Research in Medicinal and Aromatic Plants (ed. M. D. Chowdhury, G.D. Sharma, A.D. Talukdar, S. Chowdhury), Swastic Publication, New Delhi. ISBN 13-9789381084953**.**

LIST OF TRAININGS/ WORKSHOPS PURSUED:

1. Attended **"National Workshop on DNA Barcode of Life"** Organized by Department of Biotechnology, Assam University, Silchar, Assam on April 07, 2009.

2. Attended Ten days National Workshop of hands on Training Course on **"Basic Tools in Molecular Biology and Genomics"** Organized by Department of Biotechnology, Assam University, Silchar during 7 -16 December, 2010.

3. Participated Workshop on **"Molecular Phylogenetics and Evolution"** Organized by Department of Biotechnology, Mizoram University, Aizwal during 22-24 November, 2010.

4. Participated Workshop in the workshop "**Data analysis using Microsoft Excel for young researcher**" organized by Department of Business Administration, Assam University, Silchar during April, 29-30, 2011.

5. Attended ***INSA-AUS Lecture Series*** **" Science for shaping the future of India"** organized by Indian National Science Academy and Assam University, Silchar on 10th December, 2012.

6. Completed Seven days Training course on "**Classical and Modern Methods in Plant Systematics**" organized by **CSIR-National Botanical Research Institute, Lucknow** during March 4-10, 2013

LIST OF ORAL/POSTER PRESENTATION:

1. Oral presentation on **"Evaluation of partial *matK* sequences in medicinal Rauvolfioideae (Apocynaceae) species as DNA passport from Northeast India"** at 22nd Pacific Science Congress held at ***Kuala Lumpur, Malaysia*** during 14-17 June 2011.

2. Oral presentation on" **Implication of plant DNA barcode in ethnobotany from North East India**" at **12th International Congress of Ethno-pharmacology** will be held on Jadavpur University, February 17-19, 2012.

3. Oral presentation on "**Molecular identification of medicinal plants based on *matK* sequences"** at the National seminar on Medicinal Plants and Microbe Diversity and their Pharmaceutical held from 19-21 December ,2010 on **Tezpur University**.

4. Presented a poster on **"Medicinal plant DNA barcoding from Northeast India"** at the International symposium on Current Status and Opportunities in

Aromatic and Medicinal Plants held at **CIMAP, Lucknow** during February 21-24, 2010.

5. Presented a paper on "**Development of species level DNA passport of Medicinal plants in Apocynaceae family** " at **98th Indian Science Congress** held at SRM University, Chennai during January 3-7, 2011.

6. Presented a poster on "**Application of DNA barcode in Pomela (C*itrus maxima*)**" at International Conference on Biodiversity Conservation and Environmental Health held at Assam University, Silchar during 16-17th March 2012

NCBI SEQUENCE SUBMISSION DETAILS:

No Sequence data submission in NCBI: 123

Gene Name	Accession No	No of sequences
matK of Chloroplast DNA	**JN228929-JN228942,JN416982, JN416981, JN315357-JN315361, JQ582660-JQ582662, KC150885,JX966234-JX966239**	31
trnH-psbA of Chloroplast DNA	**JN245983-JN245988, JN315362-JN315369, KC150886-KC150898**	27
D-loop of mt-DNA	**JN603607-JN603629, JN417003**	25
DNA barcode (*COXI*) of mt-DNA	**FJ185310,FJ185309,GU563917, GU563918, FJ171914,FJ171915,JN245989-JN245997, JN417002,JN560176, JX127224-JX127243, JQ913015-JQ913017**	40

AVAILABLE AT: http://www.ncbi.nlm.nih.gov/nuccore/?term=Mahadani%20P

DNA BARCODES
Short communication • DOI: 10.2478/dna-2013-0002 • DNA • 2012 • 35–38

VERSITA

DNA Barcoding:
A tool for species identification from herbal juices

Abstract
Herbal ethnomedicine constituents, being unstructured, are difficult to identify at species level by appearance. Several studies of plant DNA barcoding have generated a huge number of reference sequences in the database (NCBI) for several authentically identified plant species. We tested identification of herbal constituents at the species level in a few ethnomedicine based on the *matK* sequences and following analysis through Basic Local Alignment Search Tool. Fresh leaves and herbal juices of different ethnomedicine samples were collected from the herbalists and recorded by the common vernacular name. PCR products for *matK* barcodes (585 bp - 872 bp) were recovered using a single set of primer and sequenced. The *denovo* sequences showed high similarity (99%-100%) with the conspecific sequences in the database. Therefore, different herbal constituents were readily identified, except a single case, at the lowermost taxonomic level based on *matK* sequence and following the DNA barcoding technique, indicating the novelty of *matK* in ethnobotany research.

Keywords
Ethnomedicine • Herbal juice • DNA barcoding • *matK* • Sequence homology

Pradosh Mahadani,
Sankar K Ghosh*

DNA barcode and Genomics Laboratory,
Department of Biotechnology,
Assam University,
Silchar-788011,
Assam, India

Received 24 September 2012
Accepted 21 December 2012

Introduction

The concept of DNA barcoding has become very popular for species level identification and is based on the species-specific variations between the short DNA sequences from a uniform locality of the genome [1]. In 2009, CBOL Plant Working Group proposed to use defined portions of plastid gene *rbcL* and *matK* as standard DNA barcode for plants and to be supplemented with additional regions as required [2]. But, the low discriminating power of *rbcL* gene has been reported [3]. The goal of DNA barcoding is to distinguish the majority of world species by using one or a few regions of DNA sequence and to produce a large scale reference sequence library of life on the earth. Several studies of plant DNA barcoding have generated a huge number of reference DNA barcode sequences from taxonomically authenticated species [2-7]. Therefore, an approach of similarity search with reference database would be potential to identify species from any unstructured plant part.

The traditional knowledge of using plants for treatment of various ailments in Northeast India is rich due to high diversity of tribes as well as rich diversity of plants [8]. Such rich knowledge is least explored and remained fragmentary. Moreover, there is a trend of not revealing the knowledge by herbal medicine provider to common people. Applications and efficacies of the herbal medicine depend critically on the accuracy in identifying the source plant. Appearance and the conventional morphological identification system do not easily lead to identify the constituent species in juice and powder [9]. Nevertheless, substitutes and adulterants, for profit making, are also in practice that undermined the quality of ethnomedicine. This study was carried to test the application efficiency of *matK* for species level identification of important ethnomedicinal constituents. We generated *matK* sequences from some well known, commonly available and valuable ethnomedicine, and compared with public reference database.

Methods

Fresh leaves (100g) and leaf juices (200ml) of different ethnomedicine samples were purchased from herbalists. As leaves and juices were unidentifiable by appearance, we recorded the common vernacular name in local language (Bengali) and assigned the sample ID to each material (Table 1). Some of the traditional uses of the study samples are like *Aloe vera* juice is taken for curing digestion and inflammation problem, *Cajanus cajan* for curing jaundice; *Cynodon dactylon* for dysentery, *Hibiscus rosa-sinensis* (leaves in the form of paste) as anti-inflammatory agent, *Senna hirsuta* and *Senna obtusifolia* independently for curing skin diseases, and *Acmella oppositifolia* is used in cuts as antiseptic [10]. Here, we were avoiding the admixture sample.

* *E-mail: drsankarghosh@gmail.com*

Table 1. List of sample ID and sample type with their common and scientific name (according to the Flora of Assam), Accession Number (NCBI) and Length of sequences are also given.

Sample ID	Sample type	Common Name in Bengali	Scientific Name (according to Flora of Assam)	Accession No	matK Sequence Length (bp)
AUMP16	Leaf juice	Ghrita kumari	*Aloe vera*	JN228939	781
AUMP28	Juice	Durba ghash	*Cynodon dactylon*	JN228941	872
AUMP21	Leaf juice	Akarkara	*Acmella oppositifolia*	JN228937	823
AUMP2	Leaf juice	Arhar	*Cajanus cajan*	JN228940	771
AUMP4	Young Leaf	Jaba	*Hibiscus rosa-sinensis*	JN228942	585
AUMP61	Young Leaf	Swarnapatri	*Senna hirsuta*	JQ582660	639
AUMP63	Young Leaf	Chakunda	*Senna obtusifolia*	JQ582661	634

40mg wet fresh leaves were homogenized in DNA extraction buffer (50mM Tris HCl pH 8.0, 25mM EDTA pH 8.0, and 150mM NaCl, 2µl/ml β- mercaptoethanol). In case of juice, 3ml sample was centrifuged @14000x g for 1 min and supernatant was discarded [11]. 1ml of DNA extraction buffer and more 20µl/ml β- mercaptoethanol was added into precipitation and incubated in a water bath at 65° C for one hour. Genomic DNA was extracted in less than one hour using Potassium acetate (5M, pH 9.0), Phenol: Chloroform: Isoamyl alcohol (25:24:1), Chloroform: Isoamyl alcohol (24:1). PCR was performed using primer *matK* X F 5'-TAATTTACGATCAATTCATTC-3' and *matK* 5r 5'-GTTCTAGCACAAGAAAGTCG-3' [12]. The PCR mixture contained 20ng genomic DNA, 0.2mM of each dNTPs, 50 pmole of each primer, 0.5 units of high fidelity Taq polymerase enzyme (4328212, Applied Biosystem), 1x buffer and 1.5mM $MgCl_2$. PCR in a reaction mixture of 30µl was prepared with the PCR thermal profile as 94° C for 3 min, 30 cycles at 94° C for 45 sec; 48° C for 45 sec; 72° C for 45 sec and a final extension at 72° C for 10 min. The PCR product was checked by 1.5% agarose gel electrophoresis. The PCR products of expected size were extracted using QIA quick PCR purification kit (QIAGEN, Cat. No. 28704). The purified PCR products were sequenced bidirectionally using automated DNA sequencer (ABI 3700).

Trace files were assembled in Applied Biosystem Sequence Scanner v1.0 and sequences with greater than 2% ambiguous bases were discarded using QV of 40 for bidirectional reads. Manual editing of raw traces and subsequent alignments of forward and reverse sequences enabled us to assign edited sequences for most species. The 3' and 5' terminals were clipped to generate consensus sequences for each sample. GenBank database was searched using megablast during November-December 2011 with default parameter adjusted to retrieving 5000 sequences. In most of the case, this corresponded to the sequence with the high BLAST score. In other cases, the closest match was a shorter target with a higher percent identity. Ambiguous bases in target sequence were considered as matching. A similar procedure was followed for BOLD searches. But, BOLD was less well populated with plant DNA barcode sequences used in ethnomedicine. So, subsequent analysis was not performed in BOLD.

Results and Discussion

Successful amplification was achieved using a single set of primer for the enough length of readable *matK* barcode sequences (585 bp – 872 bp) from selected ethnomedicinal juices and leaves. The BLAST searches by each sample sequence in GenBank revealed the closest matches with the same species and nearest neighbour (NN) of same or different genus. *matK* sequence of sample AUMP16 showed 100% identity with *Aloe vera* and 99% with both *Aloe compressa* and *Aloe capitata. matK* sequence of sample AUMP28 showed 99.9% similarity with *Cynodon dactylon* and 99% with *Cynodon transvaalensis* and *Brachyachne ciliaris*. The samples AUMP61, AUMP21 and AUMP2 showed 99% identity with *Senna hirsuta*, *Acmella oppositifolia* and *Cajanus cajan* respectively while a below 97% NN similarity with same or different genus was also recorded. AUMP63 showed 100% identity with *Senna obtusifolia* and 98% NN identity with *Senna alata*. However, the sample AUMP4 showed 100% identity simultaneously with *Helicteropsis microsiphon* and *Hibiscus rosa-sinensis* and remained inconclusive (Figure 1).

We authenticated the species name not only based on BLAST result but also by cross checking the common vernacular name as per the Flora of Assam [13] (Table 1). Sample AUMP4 *matK* sequence was closest to both *Helicteropsis microsiphon* and *Hibiscus rosa-sinensis*. It may be due to short sequence length (585 bp) in comparison to other barcode sequences. Therefore, according to vernacular name, the sample AUMP4 was submitted to GenBank as *Hibiscus rosa-sinensis* (Table 1). The accuracy of DNA barcoding for species identification relies on sufficient sequence difference between closely related species. However, recognition of an ideal plant DNA barcode locus also largely depends on universality, sequence quality and coverage [2]. Plant genera with possible occurrence of natural hybridization and gene introgression may be quite challenging in search of universal loci for plant DNA barcode [14,15]. Plant DNA barcoding in these cases may be problematic and demanding. Longer sequence or additional barcode markers could be provide a better resolution in case of *Hibiscus rosa-sinensis.* Large sample sizes were required to increase the power of the test with such members, but the limited number

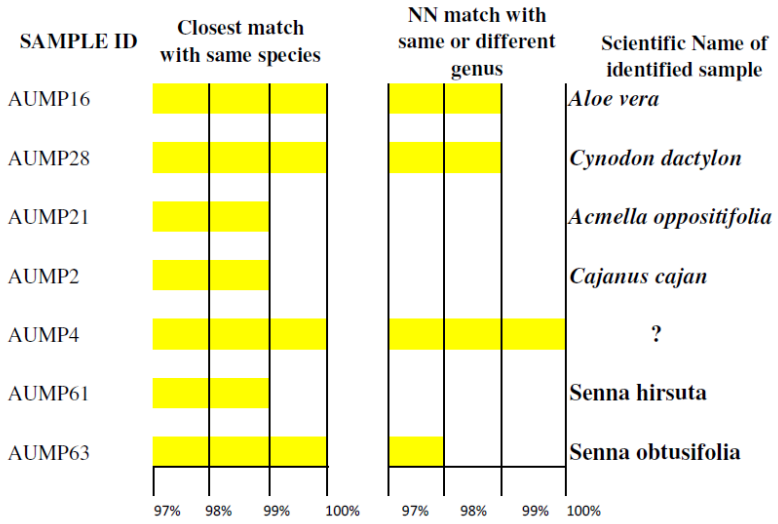

Figure 1. Species identification based on matK barcode. For each sample, sample ID, graphical representation of similarity search result (BLAST) are shown. Color bars depict percentage identity to close match with same species and Nearest Neighbor (NN) in the same or different genus, with scale at bottom.

of *matK* sequences in the database posed a limitation to flag diagnostic nucleotide positions. The *matK* gene showed the relatively small number of position's differences that distinguish many closely related plant species by comparing the closest match with identical species and NN of same or different genus [16]. Single loci *matK* showed enough discrimination power among the studied medicinal plant species which reckoned the other findings [6,7,17]. Therefore, DNA barcoding, using *matK* gene as a potential marker, can be adopted for studying and identifying medicinal plant products that are unidentifiable by morphology alone and for detecting fraudulence that would help in developing ethnobotany research.

Acknowledgement

We acknowledge the Department of Biotechnology, Govt. of India for the infrastructural support and UGC - JRF in Engineering & Technology fellowships to the author (PM). We are also thankful to Dr. Boni Amin Laskar (DBT-RA fellow) for critical comments and editing of the manuscript.

References

[1] Hebert PD, Cywinska A, Ball SL, & deWaard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270(1512):313-321.

[2] Hollingsworth MP, et al. (2009) A DNA barcode for land plants. Proc Natl Acad Sci U S A 106(31):12794-12797.

[3] Chase MW, et al. (2007) A proposal for a standardised protocol to barcode all land plants. Taxon 56:295-299.

[4] Asahina H, Shinozaki J, Masuda K, Morimitsu Y, & Satake M (2010) Identification of medicinal Dendrobium species by phylogenetic analyses using matK and rbcL sequences. J Nat Med 64(2):133-138.

[5] Bruni I, et al. (2010) Identification of poisonous plants by DNA barcoding approach. Int J Legal Med 124(6):595-603.

[6] Gao T, et al. (2011) Identification of Fabaceae plants using the DNA barcode matK. Planta med 77(1):92-94.

[7] Mahadani P, Sharma GD, & Ghosh SK (2013) Identification of Ethnomedicinal plants (Rauvolfioideae: Apocynaceae) through DNA Barocding from Northeast India. Pharmacognosy Magazine (In Press).

[8] Kala CP, Dhyani PP, & Sajwan BS (2006) Developing the medicinal plants sector in northern India: challenges and opportunities. J Ethnobiol Ethnomed 2:32.

[9] Srirama R, et al. (2010) Assessing species admixtures in raw drug trade of Phyllanthus, a hepato-protective plant using molecular tools. Journal of ethnopharmacology 130(2):208-215.

[10] Handique PJ (2009) Medicinal plants of North East India (International Book Distributors, Dehradun).

[11] Ng CC, Chang CC, Wu IC, Kotwal S, & Shyu YT (2006) Rapid molecular identification of freshly squeezed and reconstituted orange juice. Int J Food Sci & Tech 41:646-651.

[12] Ragupathy S, Newmaster SG, Murugesan M, & Balasubramaniam V (2009) DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study

by the hill tribes of the Western Ghats in southern India. Mol Eco Res 9 (Suppl s1):164-171.

[13] Kanjilal UN, Kanjilal PC, Das A, De RN, & Bor NL (1934-1940) Flora of Assam ,Government Press, Shillong, India.

[14] Fazekas AJ, et al. (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Mol Eco Res 9(130-139).

[15] Roy S, et al. (2010) Universal plant DNA barcode loci may not work in complex groups: a case study with Indian berberis species. PloS one 5(10):e13674.

[16] Stoeckle MY, et al. (2011) Commercial teas highlight plant DNA barcode identification successes and obstacles. Sci Rep 1:42.

[17] Lahaye R, et al. (2008) DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci U S A 105(8):2923-2928.

OPEN ACCESS Freely available online

PLOS ONE

The Species Dilemma of Northeast Indian Mahseer (Actinopterygii: Cyprinidae): DNA Barcoding in Clarifying the Riddle

Boni A. Laskar, Maloyjo J. Bhattacharjee, Bishal Dhar, Pradosh Mahadani, Shantanu Kundu, Sankar K. Ghosh*

Department of Biotechnology, Assam University, Silchar, Assam, India

Abstract

Background: The taxonomic validity of Northeast Indian endemic Mahseer species, *Tor progeneius* and *Neolissochilus hexastichus*, has been argued repeatedly. This is mainly due to disagreements in recognizing the species based on morphological characters. Consequently, both the species have been concealed for many decades. DNA barcoding has become a promising and an independent technique for accurate species level identification. Therefore, utilization of such technique in association with the traditional morphotaxonomic description can resolve the species dilemma of this important group of sport fishes.

Methodology/Principal Findings: Altogether, 28 mahseer specimens including paratypes were studied from different locations in Northeast India, and 24 morphometric characters were measured invariably. The Principal Component Analysis with morphometric data revealed five distinct groups of sample that were taxonomically categorized into 4 species, viz., *Tor putitora*, *T. progeneius*, *Neolissochilus hexagonolepis* and *N. hexastichus*. Analysis with a dataset of 76 DNA barcode sequences of different mahseer species exhibited that the queries of *T. putitora* and *N. hexagonolepis* clustered cohesively with the respective conspecific database sequences maintaining 0.8% maximum K2P divergence. The closest congeneric divergence was 3 times higher than the mean conspecific divergence and was considered as barcode gap. The maximum divergence among the samples of *T. progeneius* and *T. putitora* was 0.8% that was much below the barcode gap, indicating them being synonymous. The query sequences of *N. hexastichus* invariably formed a discrete and a congeneric clade with the database sequences and maintained the interspecific divergence that supported its distinct species status. Notably, *N. hexastichus* was encountered in a single site and seemed to be under threat.

Conclusion: This study substantiated the identification of *N. hexastichus* to be a true species, and tentatively regarded *T. progeneius* to be a synonym of *T. putitora*. It would guide the conservationists to initiate priority conservation of *N. hexastichus* and *T. putitora*.

Citation: Laskar BA, Bhattacharjee MJ, Dhar B, Mahadani P, Kundu S, et al. (2013) The Species Dilemma of Northeast Indian Mahseer (Actinopterygii: Cyprinidae): DNA Barcoding in Clarifying the Riddle. PLoS ONE 8(1): e53704. doi:10.1371/journal.pone.0053704

Editor: Vincent Laudet, Ecole Normale Supérieure de Lyon, France

Received August 16, 2012; **Accepted** December 3, 2012; **Published** January 16, 2013

Funding: BAL has received a fellowship (DBT-RA) from the Department of Biotechnology, Government of India. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: drsankarghosh@gmail.com

Introduction

The term 'mahseer' refers to a group of freshwater cyprinid fishes easily distinguishable by relatively larger size of scales on their body compared to the other cyprinid fishes [1,2]. The members of mahseer belong to two genera, viz., *Tor* and *Neolissochilus*. These two genera are distinguished by the presence of a continuous labial groove in *Tor* but interrupted in *Neolissochilus*, and 10–14 gill rakers on the lower arm of first gill arch in the former and 6–9 in the latter [3,4]. They inhabit in the mountain streams and distributed in the range from Pakistan throughout Southern Asia to Southeast Asia up to the Malay Peninsula and the larger Indonesian islands across Sumatra, Borneo and Java [5,6]. However, species composition within each genus varies in different locations, like Southeast Asian species are different from Southern Asian species. Furthermore, within India, many species of mahseer are discontinuously distributed and mostly endemic in the South, Central and Northeast India. Among the mahseer of the Indian subcontinent, *Tor putitora* is widely distributed in Pakistan, India, Nepal and Bhutan; while *Neolissochilus hexagonolepis* is distributed in Nepal, Bhutan, North India and Northeast (NE) India [7,8]. A few studies suggest that the angling of mahseer provides superlative thrills than any other sport fishes except European Salmon [9,10]. They are highly sought-after because of great attraction to recreational anglers and are important components of the Angling-tourism pursuit [11]. In developing countries, there are many instances where the tourism industry has added recreational fishing to their attractions [12]. Owing to the growing value, the mahseer has become popular and considered as a cultural icon of diverse economic, recreation, and conservation

standpoint in rivers of eleven Asian nations [13]. Above all, the mahseer is an integral component of the aquatic ecosystem, serves as an important indicator of its health and supports the livelihood of many rural and indigenous ethnic groups in Asia [14]. However, the important mahseer fishes are threatened in the NE India as well as other distribution areas due to the growing harvest pressure as well as anthropogenic effects [15,16]. The two most threatened species, viz., *Tor putitora* and *Neolissochilus hexagonolepis* are regarded as the flagship species in NE India (http://www.nbfgr.res.in/). The conservation of mahseer has been hampered because the taxonomy of mahseer is most confusing due to the morphological variations they exhibit [17] that poised the understanding of actual species composition, distribution, autecology and biology at large.

Historically, with the pioneering work of Hamilton-Buchanan (1822) [1], many new descriptions of different species of mahseer have been proposed from Indian waters by distinguished naturalists. McClelland (1839) [18] recorded 4 new species from NE India, viz., *Tor progeneius*, *T. macrocephalus*, *Neolissochilus hexagonolepis* and *N. hexastichus*. McClelland, however, admitted difficulty in identifying Hamilton's *Cyprinus* (now *Tor*) *putitora* and particularly emphasized on a large cellular appendage to the apex of the lower jaw for *T. progeneius*, and the color gray on the back and reddish yellow on rest of the body for *N. hexastichus* [18]. The taxonomy of *T. progeneius* had long been in doubtful status, and it has been considered as a junior synonym of *T. putitora* [19]. Sen and Jayaram (1982) [20] characterized *T. progeneius* and elucidated with some new characteristics. Later, Rainboth (1985) [3] noted that *T. progeneius* is confusing to be classified whether within the genus *Neolissochilus* or *Tor*. It was further noted that most of the McClelland's type specimens were misplaced and some constituted curatorial nightmare [3]. Yet, McClelland's descriptions of two distinct species, viz., *Neolissochilus hexagonolepis* and *Tor progeneius* are recognized to be valid; while *T. macrocephalus* and *N. hexastichus* have been considered to be not valid rather the former was synonymized with *T. putitora* and the latter with *T. tor* [5,21].

Thus, the traditional taxonomy of mahseer in NE India has been facing several problems due to (1) lack of morphometric details in original description, (2) presence of very few holotypes of mahseer species, (3) indiscernible morphological nuances in them, and (4) disagreements in recognizing specific morphological characters. Consequently, the taxonomy of a few mahseer species has been extremely chaotic and described severally [2,4,5,20,21,22,23]. The mahseer species composition in the region is poorly understood and the identification of two species, viz., *T. progeneius* and *N. hexastichus*, has been difficult due to inconsistent taxonomic descriptions. Therefore, species level identification of mahseer is needed to be strengthened to facilitate the autecological study of mahseer and to develop conservation strategy for sustainable utilization in recreational fishing based tourism. Genomic approaches of taxon diagnosis have been found to be resourceful to aid traditional taxonomy [24,25]. In this context, the mitochondrial genome is a better target than nuclear genome because it evolved faster and can thus give more information to discriminate close species.Lately, a partial fragment of mitochondrial cytochrome oxidase c subunit I (COI) gene has been proposed to be sufficient singly to differentiate all, or at least the vast majority of animal species [26]. As such, this partial locus (COI) has been extensively tested for its efficacy in fish species identification and recognized as a unique marker of species identification with high confidence and called as "DNA barcode" [27,28,29]. The concept of DNA barcode based species identification is easy, rapid and accurate for being sequencing and web based; as such it has gained great attention worldwide [30,31,32]. Recently, the catfish diversity in NE India has been re-evaluated through DNA barcoding [33]. Therefore, morphological and DNA barcode data in combination can help to resolve the species dilemma of Northeast Indian mahseer, particularly *T. progeneius* and *N. hexastichus*, for effective conservation and management of the species.

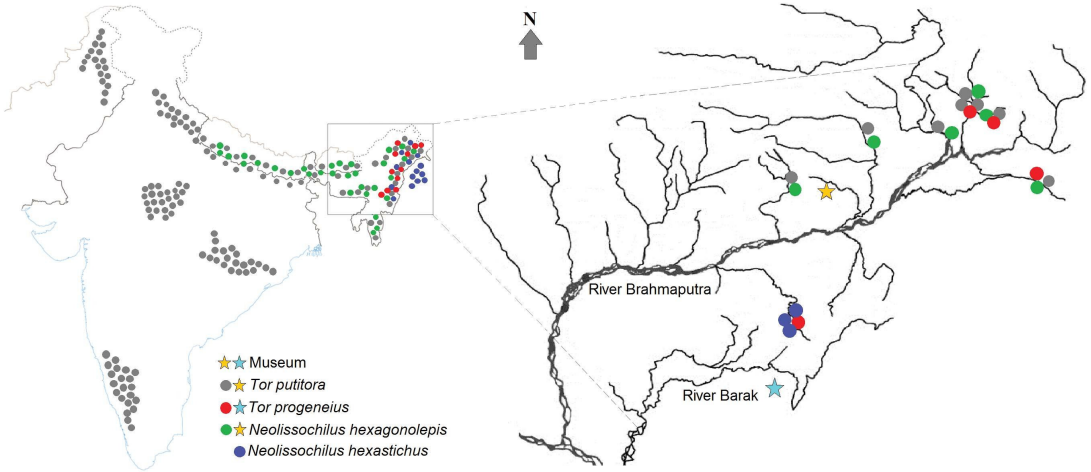

Figure 1. Map of the study site showing the known distribution of the studied species and the collection sites in different river drainages. The figure shows that the Northeastern region of India is drained mostly by River Brahmaputra and partly by River Barak. The studied specimens were collected from the drainages of River Brahmaputra. The topography of the region restricts the convergence of Southeast Asian fish composition with this region.
doi:10.1371/journal.pone.0053704.g001

Table 1. Morphological grouping of the studied organisms along with the corresponding codes.

Group	Nomenclature in practice	Sequence accession number used in molecular analysis/ catalogue number of paratypes in museum	Sample code used in morphological analysis (PCA)
N_2	*Neolissochilus hexastichus*	SGBL-BMF35	A
		JX127237	B
		JX127239	C
		JX127235	D
		JX127236	E
		JX127238	F
		SGBL-BMF36	*
N_1	*N. hexagonolepis*	JX127232	G
		JX127234	H
		JX127231	I
		JX127233	*
		** RGUMF-0036	V
		** RGUMF-0037	W
		** RGUMF-0038	X
T_2	*Tor progeneius*	JX127229	J
		***	K
		JX127228	L
		***	M
		JX127230	N
T_3	*T. putitora*	** RGUMF-0034	Aa
		JX127240	O
		JX127224	P
		JX127241	Q
		JX127242	U
T_1	*T. putitora*	JX127227	R
		JX127226	T
		JX127225	S
		** RGUMF-0035	Y
		***	Z
		***	Ab

The grouping was done based on scatter plot from Principal Component Analysis (PCA) as well as following the authoritative taxonomic keys. Sequence accession numbers in GenBank are used in the presentation of molecular analysis and the sample codes in PCA. Alphabetic sample codes replacing the full name of organisms are ascribed for ease of presentation those however clearly mentioned in Table S2.

*big specimen from market whose morphometric not done.

**paratypes from museum preserved in formaline whose sequencing not done.

***previously identified specimens preserved in formaline whose sequencing not done.

doi:10.1371/journal.pone.0053704.t001

Materials and Methods

Sample Collection

Fish specimens belonging to the group mahseer in the range of sub-adult to adult size were collected through participatory sampling with the marginal fishers engaged in commercial fishing. The specimens were from various locations in the hills and foothills across the Northeast India, particularly in the drainages of River Brahmaputra (Figure 1). The method of sample collection was approved by the Ministry of Science and Technology, Department of Biotechnology, Government of India (vide No. BT/HRD/01/002/2007). Some known voucher specimens within the genera *Tor* and *Neolissochilus* were examined from the Museum of Biodiversity in Rajiv Gandhi University, Arunachal Pradesh (voucher numbers are given in Table 1). The morphometrics of previously identified specimens from collection of *T. putitora* and *T. progeneius*, as well as the type specimens of *T. putitora* and *N. hexagonolepis* were included in the analysis. The type specimens of *T. progeneius* and *N. hexastichus* are not available in the museum. In lieu of examining type specimens of *T. progeneius* a small review on the existing contradictions among the taxonomists regarding the taxonomic descriptions and opinions on the status of the species is given in Supporting Information S1. Concerning the identification of *T. progeneius* and *N. hexastichus*, the original descriptions were emphasized. A total of 19 fresh specimens belonging to 4 species were studied in association with 5 paratypes and 4 previously collected specimens. Muscle tissue samples were invariably collected aseptically from behind of dorsal fin of the fresh specimens and taken in 500 µL of TES buffer (50 mM Tris HCl, 25 mM EDTA and 150 mM NaCl). The whole body

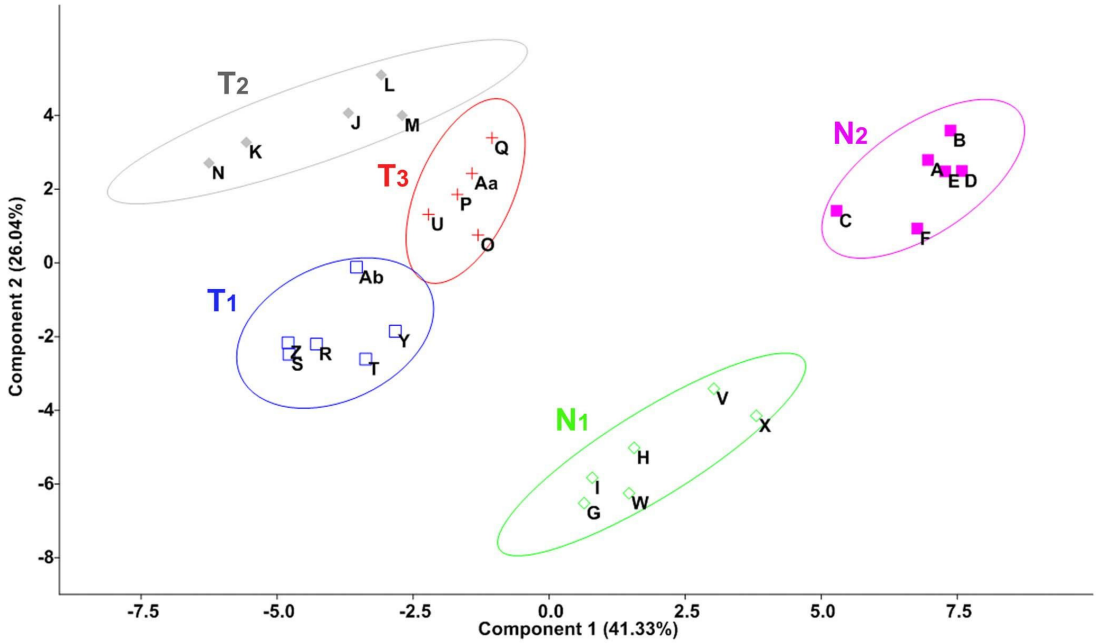

Figure 2. Principal Component Analysis (PCA) on 24 morphometric variables of the study samples including paratypes. The clusters of samples obtained from PCA were assigned to respective taxa based on meristic counts as well as non-quantitative characters of samples following authoritative taxonomic keys. The groups are like T_1, T_2 and T_3 comprising *Tor* congener, and N_1 and N_2 comprising *Neolissochilus* congener.
doi:10.1371/journal.pone.0053704.g002

specimens are preserved and stored at the Department of Biotechnology of Assam (Central) University, Silchar, Assam, India, for frequent examination and record of vouchers (vouchers' details are provided in Table S1).

Taxonomic Identification and Nomenclature

Specimens were categorized systematically based on the taxonomic characters available from the original description as well as subsequent re-descriptions and taxonomic reviews. Altogether 24 morphometric variables along with 6 important meristic counts were measured following standard literatures [23,34] (Figure S1) and the measurements were recorded using digital slide caliper (0.01 mm). The morphological characters those are non-quantitative yet taxonomically relevant, e.g. color pattern on the body and fins, presence or absence of tubercles, appearance and diagonal shape of mouth, etc. were also recorded from all the specimens. The measurements were taken at least three times independently and mode of each parameter was finally considered to minimize the error. The samples were designated into the respective species as per the authoritative taxonomic keys [4,20,23] and the species nomenclature was adopted as per the updated catalogue [8].

PCR Assay and Purification

DNA was extracted with standardized Phenol-Chloroform-Isoamyl alcohol method [35]. COI gene fragment (~655 bp) was amplified using the set of published primers: FishF1-5′TCAACCAACCACAAAGACATTGGCAC 3′ and FishR1-5′TAGACTTCTGGGTGGCCAAAGAATCA 3′[27]. The amplification was performed in 25 µl reaction mixture of 1X PCR buffer, 2 mM $MgCl_2$, 10 pmol of each primer, 0.25 mM of each dNTPs, 0.25 U high-fidelity polymerase and 100 ng of DNA template. PCR conditions were: initial denaturation at 94°C (2 minutes) followed by 30 cycles at 94°C (45 seconds), 50°C (45 seconds) and 72°C (1 minute), and a final elongation at 72°C (8 minutes). The PCR-amplified products were checked in 1% agarose gels containing ethidium bromide (10 mg/ml) and the single uniform band was then purified using QIAquick^R Gel extraction kit (QIAGEN, USA). The amplicons were bi-directionally sequenced in an automated DNA sequencer (ABI 3500, Applied Biosystems Inc., CA, USA).

Sequence Quality Control Measures

Two chromatograms that represent sequences of both the DNA strands were obtained for each sample. The PCR amplified products as well as their corresponding DNA sequences were larger than 600 bp that assured the sequences being not Numts as the limit of Numt hardly reaches 600 bp [36]. The noisy sequences were trimmed at both end and greater than 2% ambiguous bases were discarded, using quality value of >40 for bidirectional reads. BLASTN program was used to compare the sequences retrieved from the two chromatograms [37], and the fragment showing 100% alignment with no gap or indel (insertion/deletions) was selected. In some cases of discrepancy, both the sequences were reviewed and quality value of the sequences were considered to determine the most likely nucleotide using the software SeqScanner Version 1.0 (Applied Biosystems Inc., CA, USA). The selected fragments of the sequence were aligned using ClustalX software [38]. Finally, each of the sequences was compared in NCBI through BLASTN to examine

Table 2. Summary of PCA on 24 morphometric measurements of 28 samples within 4 species.

	PC 1	PC 2
% variance	41.336	26.04
Eigen value	19.4715	12.2662
Variable	Loadings	
SL	−0.3807	−0.1508
PrDL	−0.1599	0.2042
PoDL	0.1929	−0.1689
HtCF	0.298	0.1092
HL	−0.2624	0.3204
HtPF	0.1026	0.09338
HtDF	0.2399	0.3027
HtAF	0.1577	0.1663
HtDS	0.1129	0.1021
DP&V	−0.1125	−0.4824
LnCP	−0.1836	−0.1738
BDdf	0.4392	−0.1007
HDop	0.2464	−0.1186
HDe	0.1665	−0.03199
BWdf	0.1187	−0.1101
HWe	0.1814	0.00083
SnL	−0.1136	0.1563
ED	0.0623	0.07016
LnLF	−0.182	0.363
LHtCP	0.1185	−0.00832
HtVF	0.1842	0.07278
DVF&AF	−0.08024	−0.4201
LnBDF	0.2209	−0.09064
LnBAF	0.03401	−0.05084

Proportion of variance, Eigen values and coefficients (loadings) of the first two principal components (PC1 and PC2) for the % total length of the morphometrics of studied mahseer species.
doi:10.1371/journal.pone.0053704.t002

the complete alignment with the partial coding sequence of fish mitochondrial COI gene. The sequences were translated using the online software ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and aligned through BLASTP to examine whether the partial amino acid codes were coherent with the fish mitochondrial COI gene frame and without any stop codon. In this way, the generated sequences were confirmed to be the fragments of mitochondrial COI gene. All the analyzed sequences were then deposited in GenBank (details of accession numbers are given in Table S1). The sequences were also submitted in a FISH-BOL project entitled "DNA barcoding of Mahseer fishes from Northeast India" and the code name 'MFISH'.

Data Analysis

Morphometry. Principal Component Analysis (PCA), a multivariate statistical procedure commonly used to reveal patterns in measured correlated variables, was used to differentiate the samples into possible groups and any variation among the samples of same species and the paratypes. The morphometric measurements were transformed into percentage of the total body length to develop the relative data of each variable for the samples of different size and species. The analysis was performed using PAST version 2.17 b (http://folk.uio.no/ohammer/past). The PCA output is presented as scatter plot showing the groups of the samples with designated codes.

COI sequence data analysis. The sampled specimens were invariably sequenced and their congeneric sequences were acquired from the databases (GenBank and BOLD) to examine the level of intraspecific variation. Most of the database sequences lack geographical information yet they were assumed to be at least from distant locations. The analysis was based on a total data set of 76 COI barcode sequences of mahseer containing 21 *denovo* sequences and 55 database sequences. Additionally, 2 sequences of *Hypsibarbus wetmorei* and 3 sequences of *Puntius sarana* were acquired from GenBank to represent the out-group in the study. Geographical information and GenBank accession numbers of the developed as well as acquired sequences are given in Table S1. The calculation of Kimura 2-parameter (K2P) congeneric and conspecific distance [39] as well as phylogenetic analysis through Neighbor Joining (NJ) method were performed using MEGA Version 5.1 [40]. The tree topology obtained through NJ method was double-checked by Maximum Likelihood (using MEGA Version 5.1) and Bayesian approach (using MrBayes 3.2.0) [41].

Results

Morphological Characteristics

The PCA yielded 24 components which correspond to the 24 morphometric measurements. Projection of the morphometric data of studied mahseer species on first 2 principal axes showed the separation of the samples into 5 groups at 75% concentration ellipse level (Figure 2). The first 2 principal components contributed to 67.37% of total variance (PC1 = 41.33% and PC2 = 26.04%) (Table 2). The third, fourth and fifth components contributed to 8.57%, 4.93% and 3.55%, respectively, but did not improve the separation of the samples. These 5 groups were categorized into 2 broad groups and each corresponds to a genus, as per the authoritative taxonomic keys. The meristic count of the samples is presented in Table 3 which depicts a prominent difference in number of gill rakers on the lower arm of first arch between the two genera. The rakers were 8–9 in *Neolissochilus* and 13–14 in *Tor*. The other meristics were almost similar in all the samples. In the PCA scatter plot, the samples within the genus *Neolissochilus* further formed two distinct groups, one of which grouped with the paratypes of *N. hexagonolepis* but the other group stood distant indicating both the groups belonging to different species. The samples within the genus *Tor* appeared to be in a single but very stretched out group indicating a wide range of variation. In this group, some samples formed two slightly distant groups, yet each of the groups assembled with at least one of the paratypes of *T. putitora* while the rest few samples formed a slightly separate group and remained away from the paratypes. The non-quantitative characters of samples within *Tor* and the prevailing taxonomic descriptions suggested two possible species name. The groups of samples appeared in PCA were designated as T_1, T_2 and T_3 comprising *Tor* congener, and N_1 and N_2 comprising *Neolissochilus* congener. The constituent samples within each group were given the alphabetic sample code (Table 1), like S, R, T, Y, Z and Ab fall within T_1; J, K, L, M, and N fall within T_2; Aa, O, P, Q and U fall within T_3; G, H, I, V, W and X fall within N_1; and A, B, C, D, E and F fall within N_2. The meristic counts and morphometric data are given in Table 3 and supplementary Table S2 respectively.

Table 3. Important meristic counts of three specimens in each species.

Organism name (Species)	Replicates	Parameters					
		Gill rakers on first arch (upper arm+lower arm)	Scales on lateral line	Dorsal fin rays	Ventral fin rays	Pectoral fin rays	Anal fin rays
Tor putitora	a	2+14	25	9+ii	8+i	15+i	6+i
	b	2+14	25	9+ii	8+i	15+i	6+i
	c	2+14	25	9+ii	8+i	15+i	6+i
Tor progeneius	a	2+14	26	9+ii	8+i	15+i	6+i
	b	2+13	26	9+ii	8+i	15+i	6+ii
	c	2+14	26	9+ii	8+i	15+i	6+i
Neolissochilus hexagonolepis	a	2+8	27	9+ii	8+i	14+i	6+i
	b	2+8	27	9+ii	8+i	14+i	6+i
	c	2+8	27	9+ii	8+i	14+i	6+i
Neolissochilus hexastichus	a	2+9	24	10+ii	7+i	14+i	7+i
	b	2+9	24	10+ii	7+ii	14+i	7+i
	c	2+9	24	10+ii	8+i	14+i	7+i

The table shows that the number of gill rakers on the lower arm of first arch is a very important distinguishing character between the two genera. This character is very easily identifiable and based on this character the first hand classification of mahseer in to respective genera can be easily done.

• Lowercase roman numerals are used to denote the simple rays in fin ray count.

doi:10.1371/journal.pone.0053704.t003

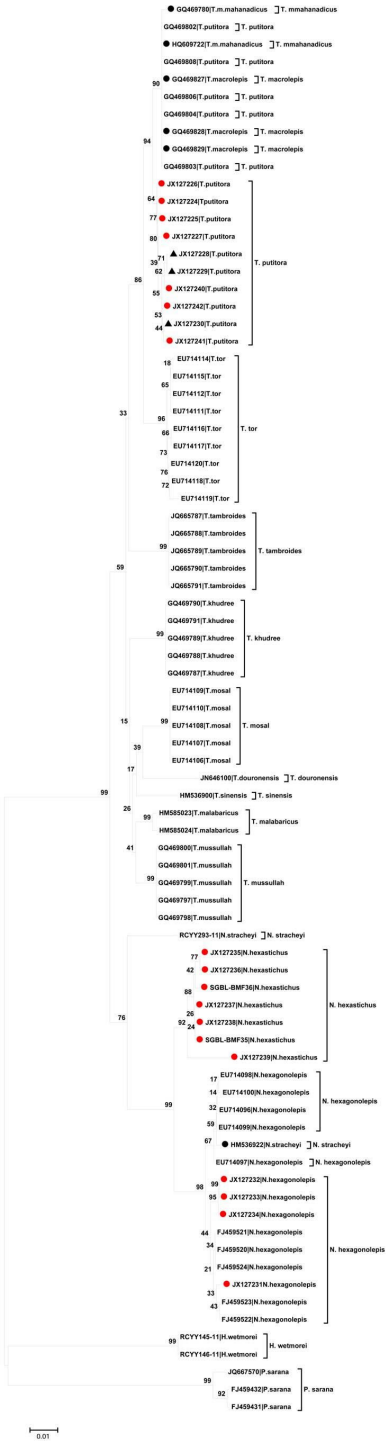

Figure 3. Neighbor Joining (NJ) tree developed using K2P distance among 76 COI sequences of mahseer. In most cases the studied samples (marked as red dots) showed cohesive clustering with their conspecific database sequences. The samples morphologically identified as *Tor progeneius* (accession numbers JX127229, JX127230 and JX127228) clustered conspecific with developed sequences as well as database sequences of *T. putitora.* All samples of *N. hexastichus* clustered cohesive as the same species and distinct from *N. hexagonolepis* sequences. Few database sequences of 3 species revealed aberrant clustering {*Neolissochilus stracheyi* (accession number HM536922), *Tor mosal mahanadicus* (accession numbers HQ609722, GQ469780), *Tor macrolepis* (accession numbers GQ469827-29)}. ● The numbers at the nodes are bootstrap values based on 1000 replications. ● The specimens' GenBank accession number and species name are shown for each taxon. ● Red dots and black triangles correspond to the sequences developed in this study. Black triangles also correspond to the sequences of samples, although morphologically identified as *Tor progeneius* but were found conspecific with *Tor putitora* based on COI sequence data analysis. Black dots correspond to the cases of abnormal clustering.
doi:10.1371/journal.pone.0053704.g003

Tor Congener

The first taxonomic key to differentiate species within *Tor* is based on the relative head length to the body depth. In this study, all *Tor* congener possessed slightly longer head than body depth. There was no stringent variation in meristic counts among the *Tor* congener (Table 3); and both T_1 and T_2 samples were similar in most of the other taxonomic features (Table S2). But, T_3 differed from T_1 and T_2; firstly on having long (up to the margin of maxilla) mental lobe (also called lower labial flap) vs. short/absent, and secondly on having longer upper jaw and with skin like flap extending behind upper lip vs. both the jaws equal and upper lip without a flap. The mental lobe length was 4.29% to 5.91% of total length in T_3 samples vs. 1.47% to 1.99% in T_1 and T_2 samples. The longer upper jaw in T_3 samples correspondingly shared to greater head length and snout length than in T_1 and T_2 samples. It appeared that the presence of both upper and lower lips as being relatively more fleshy and the mental lobe being prominent and long in all the samples of T_3 differentiate them from T_1 and T_2 samples. The observed morphological features of all the samples within T_1 and T_2 bear close affinity with the described features of *Tor putitora.* Therefore, despite minor differences, T_1 and T_2 samples were considered to be belonging to the same species and named accordingly. The particular lip character in T_3 samples resemble with the original descriptions of *Tor progeneius*. It was observed that the three groups though bear minor variation in morphometrics but they are not discernible except the particular differentiating features of *Tor progeneius* that appeared to be unique and very much noticeable. Thus, T_3, T_1 and T_2 samples are tentative considered as morphs and the former is designated as long mental lobed while the latter 2 as short mental lobed.

Neolissochilus Congener

The N_1 samples were distinguished from N_2 due to absence of mental lobe vs. prominent, and interrupted groove behind the lower lip vs. continuous groove. Both these features of N_2 samples resembled with the *Tor* congener. But, the gill rakers in them were 9 vs. 13–14 in *Tor* congener. Tubercles were mostly present on the cheeks in N_1 but entirely absent in N_2. Mouth smoothly rounded in N_2 vs. truncate in N_1, edge of lower jaw blunts in N_2 vs. sharp in N_1. The color of the back, bases of caudal and dorsal as well as the upper part of the head in N_2 samples was greenish gray, reddish yellow on rest of the body, and the tips of the fins red. These observed features in N_2 samples have been originally emphasized

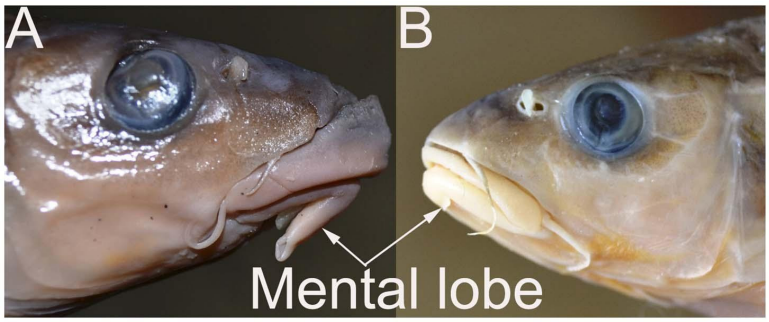

Figure 4. Illustrating the counter appearances of mouth in (A) *T. progeneius* and (B) *T. putitora*. Showing fleshy lips, a semicircular flap extending behind the upper lip, and a fleshy appendix extending from the lower lip up to the margin of maxilla (very long mental lobe) in (A), but absent in (B).
doi:10.1371/journal.pone.0053704.g004

to describe *N. hexastichus* as a distinct species. Thus, the N_2 samples were named as *N. hexastichus* according to the authoritative descriptions. The morphometrics of N_1 samples were mostly similar to both *N. hexagonolepis* and *N. stracheyi*. The N_1 samples were uniquely identified to be *N. hexagonolepis*, based on color pattern having scales coppery colored with a tinge of red above lateral line and fins deep slate paling towards their margins. As per the prevailing taxonomic description, *N. hexagonolepis* is different from *N. stracheyi* due to the absence of a lateral black stripe as in N_1 samples. Thus, following the taxonomic keys, the N_1 samples were named as *N. hexagonolepis*.

DNA Barcoding Analyses

The K2P divergence matrix of the dataset (as shown in Table S3) revealed that the congener of *Tor* maintained divergences in the range of 3.5% to 7.4% with the congener of *Neolissochilus*. The maximum K2P divergence among T_1 and T_2 samples was 0.2% and the comparison of both T_1 and T_2 samples with T_3 samples also revealed a maximum K2P divergence of 0.2%. The maximum divergence of all the samples belonging to T_1, T_2 and T_3 with the closest database sequences of *Tor putitora* was 0.8%. The divergence matrix suggested that all samples of T_1, T_2 and T_3 are conspecific of *T. putitora* in the absence of any database sequence of *T. progeneius*. Therefore, COI gene sequences of these 3 groups were submitted to both GenBank and BOLD under the putative species *T. putitora*. The maximum divergence within N_1 samples was 0.6% while their divergences with the conspecific database sequences were in the range of 0.4% to 0.8%. The divergence matrix suggested that N_1 samples are conspecific of *Neolissochilus hexagonolepis*. The within group divergences of N_2 samples were in the high range up to 0.9% possibly due to a particular sequence. Excluding the particular sequence (accession number JX127239), the within group divergence of N_2 samples remained nil in the absence of any conspecific sequence in the database.

The averages of conspecific and congeneric divergences were determined from the matrix to be 0.5% ±0.2% and 2.8% ±0.7% respectively. In the dataset, the minimum distance between the closest species (closest congener) was 1.5%. Therefore, the closest congeneric divergence among mahseer species was calculated to be 3 times higher than the mean conspecific divergence, which is called as the 'barcode gap'. Based on the barcode gap, T_1, T_2 and T_3 samples were found to be conspecific with *T. putitora*; N_1 samples were conspecific with *N. hexagonolepis*; and N_2 samples including JX127239 were discrete in the absence of any database sequence of *N. hexastichus*.

The NJ as well as Maximum Likelihood (ML) and Bayesian tree based cluster revealed that the congener of *Tor* and *Neolissochilus* formed two related clades while *Puntius sarana* and *Hypsibarbus wetmorei* remained as out-group (Figure 3, Figure S2, and Figure S3). This also revealed that *T. putitora*, *T. tor*, *T. khudree*, *T. sinensis*, *T. mussullah*, *T. mosal*, *T. malabaricus*, *T. douronensis*, *T. tambroides*, *N. hexagonolepis* and *N. stracheyi* clustered separately and are distinct species. All the samples of T_1, T_2 and T_3 clustered in the same clade and nearest to *T. putitora*; N_1 samples clustered with *N. hexagonolepis*; and the 6 N_2 samples clustered in the same clade while the other N_2 sample remained a bit distant. In addition, some database sequences reflected aberrant clustering like, 1) all sequences of *T. mosal mahanadicus* and *T. macrolepis* clustered with *T. putitora* and 2) a single sequence of *N. stracheyi* (accession number HM536922) clustered with *N. Hexagonolepis*.

Discussion

In this study, all the possible mahseer habitats across the Northeast India were surveyed. Altogether three morphologically distinct groups of mahseer within the genus *Tor* and two within *Neolissochilus* were identified from the study site. DNA barcoding analyses however recognized all the three groups belonging to a single species within the genus *Tor* and conspecific of *T. putitora*. The *T. putitora* is a widely distributed species and it has been reported to be exhibiting polymorphism in geographically isolated populations [42]. Among the study samples, T_3 samples possessed long fleshy appendage to the lower lip (mental lobe) while the others lack this feature. This feature corresponds to the original description of *T. progeneius* where this particular feature was specially emphasized for nomenclature [18,21]. This species had been also considered closely allied to *T. tor* in view of its lower lip character; consequently these two species have been synonymized very often [5,19]. *T. progeneius* was however differently described after its original description probably due to lack of original holotype [3] and non-availability of fresh specimens [21]. It was identified to be distinct from *T. tor* due to length of head almost equal to depth of the body in the former vs. length of head considerably shorter than depth of the body in the latter [20]. Subsequently, based on archival specimens (Zoological survey of India, Kolkata; specimens' catalogue details not mentioned), it was characterized to be having 8–10 rakers on the lower arm of first gill arch, tubercles on the cheek and lacking completely a mental

lobe. Based on such characters this species was remarked to be doubtful to place in either in *Tor* or *Neolissochilus* [3]. According to one proposition, there are two types among the yellow finned mahseer: i) the lips are fleshy and the lower one is produced backwards into a long fleshy appendage, and ii) the lips are of normal type and the lower lip does not form an appendage [21]. Based on such descriptions, Hamilton's *Cyprinus* (present *Tor*) *putitora* and *C. mosal* have been stated to be the same species and the nature of their lips was stated to be adaptive characters [5]. Besides, the description of a fan-shaped structure behind upper jaw in *T. progeneius* [21] was stated to be an abnormal formation based on archival specimens (Zoological Survey of India, Kolkata; specimens' catalogue details not mentioned) [5]. In contrary, Menon (1992) [5] described *T. progeneius* to be possessing of 27–31 numbers of lateral line scales on the body. It seems that Menon (1992) was so influenced by this feature of *T. progeneius* that he used it as a taxonomic key to species. Secondly, in contrary to all previous descriptions except Rainboth (1985) [3], Menon (1992) noted the presence of cheek tubercles in *T. progeneius*. On the other hand, according to original description as well as the prevailing adoption of taxonomic character for this species indicate that the number of scales on the lateral line was never more than 26, and the extension of singular appendage from lower lip has been largely emphasized. This species has long been remained unreported, that might be due to the above mentioned morpho-taxonomic perplexity arising from vague and varied presentation of its specific characters incongruent to the original description [18]. All the T_3 samples were observed to be possessing of maximum of 26 lateral line scales, 13–14 gill rakers, the slightly longer head than body depth and particularly the fleshy lips with long angular appendage to the lower jaw (long mental lobe) that is in contrast to the short mental lobe in both T_1 and T_2 samples (Figure 4). The different lower lip structure in T_3 samples could be an adaptive [5,21] or a sexually dimorphic feature [43]. Moreover, different geographical populations of *T. putitora* have been reported for significant Nuclear Organiser Region polymorphism [42] that indicates the possibility of the presence of a polymorphic form of this species in northeast India. Because, the collection site of T_3 samples in the drainages of river Brahmaputra is phylogeographically poorly connected with the other Himalayan streams such as Ganga. Therefore, notwithstanding such noticeable differences in mouth structure, following DNA barcoding results, we conclude that *T. progeneius* is a synonymous species of *T. putitora*. This study contributed 10 replica barcode sequences in GenBank of *T. putitora*. In elsewhere, DNA barcoding approach has been successful in describing different nominal species in one [44]. This study would guide the conservationists to turn away the focus of conservation endeavor from *T. progeneius* to *T. putitora*.

The present study recognized two morphologically distinct groups of mahseer within the genus *Neolissochilus*. Among them, the N_1 and N_2 samples were identified to be *Neolissochilus hexagonolepis* and *N. hexastichus* respectively. DNA barcoding also differentiated both the species with considerable barcode gap and hence their identifications were confirmed. This study added in GenBank 3 replica barcode sequences of *N. hexagonolepis* and 7 new barcode sequences of *N. hexastichus*. The latter species has long been concealed since its first description in around 175 year back [18] due to lack of its morpho-taxonomic details and mis-identification with *T. tor*. The species *N. hexastichus* is though reported from other locations in the Salween basin [45] and Myanmar [46] but there are almost no biological data available on this species. Yet, it was first categorized into 'Vulnerable' [15] and subsequently to 'Near Threatened' status [16]. In this study, two species of mahseer, viz., *T. putitora* and *N. hexagonolepis* were found frequently in all the mahseer habitats in the study area. On the other hand, the species *N. hexastichus* was absent in all the surveyed habitats except a particular river (25.420 N 92.993 E) in the entire study area that raises a serious concern about the future sustainability of this species. Although this river also harbors the other two most common species of mahseer but we observed illegal harvest of fishes through destructive fishing in the river. Thus, the mahseer species in this river are assumed to have been threatened from anthropogenic activities that demands mass awareness. This study would provide benefit to generate life history parameters of *N. hexastichus* for its conservation standpoint and development of aquaculture package of practice for sustainable utilization. Therefore, this study suggests to initiate priority conservation of *N. hexastichus*.

One of the differentiating characters of two genera *Neolissochilus* and *Tor* is based on the presence of labial groove interrupted in the former and continuous in the latter. This generic character was found to be confusing because this difference was not evident in *N. hexastichus*. Therefore, we consider that interrupted labial groove would be confusing to treat as the generic character of *Neolissochilus*. On the other hand, the characteristic difference of the number of gill rakers on the first arm of gill arch was found to be a very pronounced generic character of the two genera that may be emphasized in genus categorization.

The NJ, ML and Bayesian cluster showed that the genera *Puntius* and *Hypsibarbus* remained as out-group with respect to the two genera *Tor* and *Neolissochilus* of mahseer. In another study the two genera of mahseer have been proposed to be in a distinct clade compared to other six different clades within the subfamily Cyprininae [47]. So, the grouping of mahseer in a separate tribe [34,48] appeared justified, but, the particular tribe name is contentious. In the NJ phylogenetic analysis, some sequences, e.g., *T. macrolepis* (2 sequences) and *T. mosal mahanadicus* (3 sequences), though carried distinct names in the database but clustered cohesively with a popular species *T. putitora*. Such a wrong clustering of sequences may arise either due to misidentification or due to the occurrence of synonymous species, such as *T. macrolepis* has been stated to be a synonym of *T. putitora* [11,17]. Besides, the two samples of *Neolissochilus stracheyi* did not cluster with each other and have been possibly misidentified in the database.

In the history of taxonomy, the dawn of DNA barcoding technique has sufficiently helped in troubleshooting of many species identification where morphological characters were overlooked or overemphasized [29]. Yet, the reference database is found to be lacking of information on many extant species of mahseer. Hence, development of both new barcodes and replica barcodes from wide spatial scale would be important to enrich the DNA barcode reference library. New barcodes are particularly essential to achieve the objective of DNA barcoding to complete the digital taxonomic guide of earth's biota, while the replica barcodes from wide geographical ranges would substantiate the range distribution of the extant species.

Supporting Information

Figure S1 Scheme of measurement of morphometric variables on Fish. (adopted from Jayaram (1999) [23].
(TIF)

Figure S2 ML phylogeny. The tagging of the sequences with red and black dots as well as black triangles follow the same description as given for NJ phylogenetic tree in Figure 3.
(TIF)

Figure S3 Bayesian phylogeny. The specimens' GenBank accession number and species name are shown for each taxon. The

sequences highlighted with red and blue colour correspond to the sequences developed in this study while blue coloured sequences alone correspond to the sequences of samples morphologically identified as *Tor progeneius*, but are found conspecific with *Tor putitora* in this study hence, marked as *Tor putitora*. The green coloured sequences correspond to the cases of abnormal clustering.
(TIF)

Table S1 List of the studied species, GenBank Accession of the analyzed sequences and the geographical positions of the sample.
(DOC)

Table S2 Morphometric details of the studied species.
(DOC)

Table S3 Pairwise K2P divergence matrix between the sequences.
(XLS)

Supporting Information S1 Comparison of taxonomic descriptions based on morphology of *T. Progeneius* from time to time.
(DOC)

Acknowledgments

Our humble thanks go to the Department of Biotechnology, Government of India, for providing infrastructure facilities (vide No. BT/HRD/01/002/2007).

Author Contributions

Conceived and designed the experiments: SKG BAL MJB. Performed the experiments: BAL MJB BD PM SK. Analyzed the data: BAL MJB PM SK. Contributed reagents/materials/analysis tools: BD PM SK. Wrote the paper: BAL MJB SKG.

References

1. Hamilton F (1822) An Account of the Fishes found in the River Ganges and its Branches. Edinburg. 405 p.
2. Desai VR (2003) Synopsis of biological data on the *Tor* mahseer *Tor tor* (Hamilton, 1822). FAO Fisheries synopsis, Rome, FAO. 158: 36 p.
3. Rainboth WJ (1985) *Neolissochilus*, a new genus of south Asian Cyprinid fishes. Beaufortia. 35: 25–35.
4. Talwar PK, Jhingran A (1991) Inland fishes of India and adjacent countries. Oxford and IBH publishing, New Delhi. 1 & 2: 1158 p.
5. Menon AGK (1992) Taxonomy of mahseer fishes on the genus *Tor* of Gray with description of new species from Deccan. J. Bombay Nat. Hist. Soc. India. 89: 244–258.
6. Roberts TR (1999) Fishes of the cyprinid genus Tor in the Nam Theun watershed, Mekong Basin of Laos, with description of a new species. Raffles Bulletin of Zoology. 47: 225–236.
7. Bagra K, Kadu K, Sharma KN, Laskar BA, Sarkar UK, et al. (2009) Ichthyological survey and review of the checklist of fish fauna of Arunachal Pradesh, India. Check List. 5(2): 330–350.
8. Eschmeyer WN (ed) (2012) Catalog of Fishes. California Academy of Sciences. Available: http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp. Accessed 2012 Aug 04.
9. Thomas HS (1897) The Rod in India. Thacker and Company, London. 317 p.
10. McDonald JA (1948) Circumventing the Mahseer and other sporting fish in India and Burma. Natraj publication, Dehradun. 306 p.
11. Nguyen TTT (2008) Population structure in the highly fragmented range of *Tor douronensis* (Cyprinidae) in Sarawak, Malaysia, revealed by microsatellite DNA markers. Freshwater Biology. 53: 924–934.
12. FAO (2005) Fisheries and Aquaculture topics. Recreational fisheries. Topics Fact Sheets. Text by Andrew Smith. In: FAO Fisheries and Aquaculture Department [online]. Rome. Updated 27 May 2005. Accessed on 19 July 2012. http://www.fao.org/fishery/topic/14831/en.
13. Nautiyal P (2006) Rising awareness and efforts to conserve the Indian mahseers. Current Science. 91: 1604.
14. FAO (2006) FAO newsletter No.35. http://www.fao.org/docrep/013/a0595e/a0595e00.htm.
15. Molur S, Walker S (1998) Report of the workshop "Conservation, Assessment and Management plan (CAMP) for freshwater fishes of India. Zoo Outreach Organization, Conservation Breeding specialist group, Coimbatore. 156 p.
16. IUCN (2012) IUCN Red List of Threatened Species. Version 2012.1. <www.iucnredlist.org>. accessed on 04 August 2012.
17. Mohindra V, Khare, Praveen Lal KK, Punia P, Singh RK, et al. (2007) Molecular discrimination of five Mahseer species from Indian peninsula using RAPD analysis. Acta Zoologica Sinica 53: 725–732.
18. McClelland J (1839) Indian Cyprinidae. Asiatic Researches Calcutta, Bishop College Press. 19(2): 217–468.
19. Day F (1878) The fishes of India: being a natural history of the fishes known to inhabit the seas and fresh waters of India, Burma and Ceylon. Vols. I & II. Williams and Norgate, London. 778 p.
20. Sen TK, Jayaram KC (1982) The Mahseer Fishes of India: A review. Rec. Zool. Surv. India, Occ. Paper 39: 1–38.
21. Hora SL (1941) The game fishes of India: VIII. Mahseers of the large scaled barbels of India, 6. The Jungha of Assamese, *Barbus (Tor) progeneius* McClelland. Journal of Bombay Natural History Society 12: 526–532.
22. Hora SL, Mukerji DD (1936) Fish of the Naga Hills, Assam. Records of Indian Museum 38: 328–330.
23. Jayaram KC (1999) The freshwater fishes of Indian region. Narendra publishing house, Delhi. 551p.
24. Avise JC (1994) Molecular markers, natural history and evolution. New York: Chapman & Hall.
25. Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends Ecol. Evol. 18: 70–74.
26. Hebert PDN, Stoekle M, Zemlak T, Francis CM (2004) Identification of birds through DNA barcodes. Plos Biology 2: 1657–1668.
27. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding of Australia's fish species. Philos Trans R Soc Lond B Biol Sci 360: 1847–1857.
28. Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology 74: 329–356.
29. April J, Mayden RL, Hanner RH, Bernatchez L (2011) Genetic calibration of species diversity among North America's freshwater fishes. Proc Natl Acad Sci USA 108: 10602–10607.
30. Clare EB, Lim BK, Engstrom MD, Eger JL, Hebert PDN (2006) DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Molecular Ecology Notes 7: 184–190.
31. Smith MA (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc Natl Acad Sci USA 105: 12359–12364.
32. Frezal L, Leblois R (2008) Four years of DNA barcoding: current advances and prospects. Infection, genetics and evolution. Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases 8: 727–736.
33. Bhattacharjee MJ, Laskar BA, Dhar B, Ghosh SK (2012) Identification and re-evaluation of freshwater catfishes through DNA barcoding. PLoS ONE 7(11): e49950. Doi:10.1371/journal.pone.0049950.
34. Rainboth WJ (1996) FAO species identification field guide for fishery purposes, fishes of the Cambodia Mekong. FAO, Rome. 265 p.
35. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.
36. Zhang DX, Hewitt GM (1996) Nuclear integrations: challenge for mitochondrial DNA markers. Trends in Ecology and Evolution 11: 247–251.
37. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.
38. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Current protocols in Bioinformatics/editorial board, Andreas D Baxevanis [et al] Chapter 2: Unit 2.3.
39. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111–120.
40. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 10: 2731–2739.
41. Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542.
42. Singh M, Kumar R, Nagpure NS, Kushwaha B, Mani I, et al. (2009) Extensive NOR site polymorphism in geographically isolated populations of Golden mahseer, Tor putitora. Genome 52: 783–789.
43. Byrkjedal I, Rees DJ, Willassen E (2007) Lumping lumpsuckers: Molecular and morphological insights into the taxonomic status of Eumicrotremus spinosus (Fabricius, 1776) and E. eggvinii Koefoed, 1956 (Teleostei: Cyclopteridae). J. Fish Biol. 71: 111–13.
44. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences, USA 101: 14812–14817.
45. Doi A (1997) A review of taxonomic studies of cypriniform fishes in Southeast Asia. Jap. J. Ichthyol. 44: 1–33.

46. Oo W (2002) Inland fisheries of the Union of Myanmar. In T. Petr and DB Swar (eds.) Cold Water Fisheries in the Trans-Himalayan Countries. FAO Fish. Tech. Pap. 431: 70–76.
47. Yang L, Mayden RL, Sado T, He S, Saitoh K, et al. (2010) Molecular phylogeny of the fishes traditionally referred to Cyprinini sensu stricto (Teleostei: Cypriniformes) Zoologica Scripta 39: 527–550.
48. Devi RK, Indra TJ (1997) Check List of the Native Freshwater Fishes of India. Available: zsi.gov.in/checklist/Native%20freshwater%20Fishes%20of%20India. Accessed 2012 Aug 04.

PM_191_12R4

PHCOG MAG. ORIGINAL ARTICLE

Identification of Ethnomedicinal plants (Rauvolfioideae: Apocynaceae) through DNA Barcoding from Northeast India

Pradosh Mahadani, Gouri Dutta Sharma[1], Sankar K Ghosh

Department of Biotechnology, 1Life Science and Bioinformatics, Assam University, Silchar - 788 011, Assam, India.

Submitted: 08-07-2012 Revised: 28-08-2012 Published: *******

ABSTRACT

Background: DNA barcode-based molecular characterization is in practice for plants, but yet lacks total agreement considering the selection of marker. Plant species of subfamily Rauvolfioideae have long been used as herbal medicine by the majority of tribal people in Northeast (NE) India and at present holds mass effect on the society. Hence, there is an urgent need of correct taxonomic inventorization vis-à-vis species level molecular characterization of important medicinal plants. **Objective:** To test the efficiency of *matK* in species delineation like DNA barcoding in Rauvolfiadae (Apocynaceae). **Materials and Methods:** In this study, the core DNA barcode *matK* and *trnH-psbA* sequences are examined for differentiation of selected ethnomedicinal plants of Apocynaceae. DNA from young leaves of selected species was isolated, and *matK* gene (~800 bp) and *trnH-psbA* spacer (~450 bp) of Chloroplast DNA was amplified for species level identification. **Results:** The ~758 bp *matK* sequence in comparison to the *trnH-psbA* showed easy amplification, alignment, and high level of discrimination value among the medicinal Rauvolfioidae species. Intergenic spacer *trnH-psbA* is also exhibited persistent problem in obtaining constant bidirectional sequences. Partial *matK* sequences exhibited 3 indels in multiple of 3 at 5´ end. Evidently, generated *matK* sequences are clustered cohesively, with their conspecific Genbank sequences. However, repeat structures with AT-rich regions, possessing indels in multiple of 3, could be utilized as qualitative molecular markers in further studies both at the intra-specific and shallow inter-specific levels like the intergenic spacers of CpDNA. **Conclusion:** *matK* sequence information could help in correct species identification for medicinal plants of Rauvolfioideae.

Access this article online

Website: www.phcog.com

DOI: *****

Quick Response Code:

Keywords: Apocynaceae, DNA barcoding, ethnomedicinal, indels, *matK*

INTRODUCTION

DNA barcoding is emerged as powerful technique of species identification and exemplified with its wide application in monitoring and documentation of bio-resource.[1-4] (Hollingsworth, 2011 #17)(Hollingsworth, 2011 #17) The technique utilizes ~650 bp region of mitochondrial COI in animals[5] and various chloroplast regions (*matK, rbcL*, and *trnH-psbA*) in plants.[6-8] The application of the technique emphasizes some thrust areas, like documentation of the important and vulnerable ethnomedicinal plant bio-resources, dealing with which is recently defined as the subject "Ethnobotany Genomics."[9] The principal issues in ethnobotany emphasized the importance of correct species identification and deciphering of indigenous and conventional knowledge of restorative plant usage and their transfer for the promotion of bio-prospect in human health care. Apocynaceae is one of the 10 largest angiosperm families (including Asclepiadaceae) and comprises of several prominent medicinal plants, like Rauvolfioideae subfamily of Apocynaceae is known for the rich source of typical laticiferous tissues, which produce various alkaloids and cardenolides being used in traditional medicines for stomach ulcer, fever, asthma, whooping cough, etc. Similarly, *Catharanthus roseus* is the source of very important drug *viz*. vinblastine, vincristine used in cancer chemotherapy.[10] *Calotropis gigantea* is also a potential candidate source for anti-cancer drugs,[11] and *Allamanda cathartica* possess a remarkable wound healing function.[12]

Indian saga has a long heritage of using numerous medicinal and aromatic plants (MAPs) for human health

Address for correspondence:
Prof. Sankar K Ghosh, Department of Biotechnology, Assam University, Silchar - 788 011, Assam, India.
E-mail: drsankarghosh@gmail.com

care, and the nation is bestowed with rich resources of plant bio-diversity distributed in various ecological conditions. It is the home of about 17,000 of global plant species and expected to be fully explored. It is reported that above 2000 species of ethnobotany plants have been utilizing by various medicines in Northeast India.[13] Amidst, the galaxy of rich traditional knowledge of herbal medicine in use by the majority of tribal people in NE India, there is an urgent need of correct taxonomic inventorization vis-à-vis species level molecular characterization of medicinal plants from this region in the globe. The conventional morphological techniques involve difficulties in species identification from any unstructured plant part. Thus, the ethnomedicinal resources of NE India seem least explored and found fragmentary. It entails the need of intervention of modern tools to characterize the molecular marker of important and vulnerable medicinal plants for correct species-level identification as well as their inventorization.

The DNA barcoding is rapidly evolving, but yet provides full agreement on which region(s) of DNA should be universally used for plants. In the current study, we have explored the effectiveness of *matK* and *trnH-psbA* spacer in differentiation of selected ethnomedicinal plants (*Catharanthus roseus* (L.) G. Don, *Alstonia scholaris* (L.) R. Br., *Thevetia Peruviana* (Pers.) Merrill, *Allamanda cathartica* L. Allamanda, *Tabernaemontana divaricata* (L.) Alston, *Calotropis gigantea* L. R. Br. Ex Ait) belonging to the subfamily Rauvolfiadae of family Apocynaceae inhabiting in NE India. The *matK* is located in the large single-copy region of chloroplast genome, nested between the 5´ and 3´ exons of *trnK, t-RNA* –lysin. In *matK*, rates of substitution among all the 3 codon positions are reported almost equal,[14] leading to the high rate of substitution, which results from non-synonymous mutations, but amino acid replacements occur as chemically-conserved, preserving its structural and biochemical properties.[15] The *trnH-psbA* spacer is among the most variable plastid regions in angiosperms. It is a popular tool for plant population genetic and species-level authentication.[16,17] The study shows the efficiency of *matK* in species delineation like DNA barcoding in Rauvolfiadae, and bears insights of effective utilization of *matK* indels in multiple of 3 for studies both at the intra-specific and shallow inter-specific levels in the entire family Apocynaceae.

MATERIALS AND METHODS

Sample collection, DNA Isolation, and PCR amplification

Young leaves of selected ethnomedicinal plants of Rauvolfioideae were collected aseptically from different sources in Southern Assam, India. All the species examined in the study were carefully identified by expert. About 40 mg, wet young leaves were homogenized in the DNA extraction buffer (50 mM Tris HCl pH 8.0, 25 mM EDTA pH 8.0, 150 mM NaCl, and 2 μL/mL β- mercaptoethanol). Genomic DNA was extracted through successive steps using 5 M Potassium acetate (pH 9.0), Phenol:Chloroform:Isoamylalchol (25:24:1), Chloroform:Isoamylalchol (24:1). To obtain high-quality DNA, free from polysaccharides and other metabolites that might interfere during PCR amplification, purified DNA concentration of each sample was estimated both fluorometrically and by comparison of ethidium bromide-stained band intensities against standard λ DNA. PCR was performed using primers pair, *matK*-F 5´-TAATTTACGATCAATTCATTC-3´, *matK*-R 5´-GTTCTAGCACAAGAAAGTCG-3´ and *trnH-F 5´*-CGCGCATGGTGGATTCACAATCC-3´ and *psbA-R 5´*-GTTATGCATGAACGTAATGCTC-3´ for *matK* and *trnH-psbA,* respectively.[18] The PCR reaction of 30 μl mixture contained 20 ng genomic DNA, 20 pmole each primer, 0.2 mM of each dNTPs, 0.5 units of high fidelity *Taq polymerase* enzyme (Applied Biosystem), 1Xbuffer, and 1.5 mM $MgCl_2$. PCR thermal conditions were 94ºC for 3 minutes, 30 cycles at 94ºC for 1 minute, 48ºC for 45 seconds, 72ºC for 45 seconds in the case of *matK,* and 94ºC for 3 minutes, 30 cycles at 94ºC for 1 minute; 51ºC for 45 seconds, 72ºC for 45 seconds for *trnH-psbA* and a final extension at 72ºC for 10 minutes for both the cases. The PCR products were checked by 1.5% agarose gel electrophoresis.

Purification of PCR Products and DNA sequencing

The PCR products of presumed size were extracted using QIA quick PCR purification kit (QIAGEN, Cat. No.28704). The purified PCR products were sequenced both bi-directionally using automated DNA sequencer (ABI 3700).

Sequence analysis

Raw traces were manually edited, and both forward and reverse sequences were subsequently aligned to generate targeted sequences. The 3´ and 5´ terminals were clipped to generate consensus sequences for each taxon for sequence length of ~ 758 bp (Nt. 520-1278) for *matK* and ~450 bp for *trnH-psbA*. The Open Reading Frame (ORF) for *matK* was checked, and correct amino acid sequences were determined by online software ORF prediction (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). These *matK* and *trnH-psbA* sequences were aligned individually for combined data set using the ClustalX program.[19] The aligned sequences were corrected manually, and nucleotide compositions were calculated using BioEdit program.[20] Neighbor-joining (NJ) method was used for calculating intra- and interspecies divergence. In addition, 20 sequences of *matK* and 13 sequences of *trnH-psbA* intergenic spacer

for same or related taxa of the studied specimen were obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) [Table 1]. The generated sequences of both *matK* and *trnH-psbA* for the studied species of Apocynaceae were subsequently submitted to NCBI.

Phylogenetic analysis

Pair-wise nucleotide sequence divergences were calculated using the Kimura-2-parameter (K2P) model to generate the distance matrices, and the neighbor-joining (NJ) analysis was done in MEGA 4.2 [21] to examine phylogenetic relationship between 14 taxa from a subfamily Rauvolfioideae, and two taxa from the subfamily Asclepiadaceae of Apocynaceae. K2P distances were used following the guidelines of the Consortium for the Barcoding of Life (CBOL) to evaluate performance barcoding locus (http://www.barcoding.si.edu/protocols.html). A total of 1000 bootstraps replicates were calculated for the NJ tree construction.

RESULTS

In this study, we uncovered 8 sequences of the *matK* region and 6 sequences of *trnH-psbA* spacer from the studied specimens, which include the few sequences that have been determined for the first time. The *matK* sequences of *Allamanda cathartica* (JN228933, JN228935) and *Calotropis gigantea* (JN228932), and *trnH-psbA* spacer of *Catharanthus roseus* (JN245984 and JN245989), *A. cathartica* (JN245987), *C. gigantea* (JN245986) *T. peruviana* (JN245983) are the novel sequences contributed from the study [Table 2].

Due to length variations in the *matK* sequences, only 758 aligned nucleotide positions were used in sequence analysis, of which a total of 189 variable and 157 parsimony-informative positions were found. However, the *trnH-psbA* sequences were not included in the subsequent analysis because the alignment was impossible across the Apocynaceae family [Figure 1].

Table 1: *matK* and *trnH-psbA* sequences from NCBI with their Accesion No also given

Species	Subfamily	Accession No. of *matK*	Accession No. of *trnH-psbA*
Catharanthus roseus	Rauvolfioideae	DQ660507, AM295068,	—
Vinca minor	Rauvolfioideae	AM295076, DQ660553	FJ493259
Alstonia scholaris	Rauvolfioideae	FJ449631, Z70189, AJ429321	GQ435037, GQ435038
Alstonia microphylla	Rauvolfioideae	GU135061, GU135060	GU135394, GU135392
Thevetia ahouai	Rauvolfioideae	GQ982112	GQ982387
Thevetia peruviana	Rauvolfioideae	Z70188,	—
Allamanda schottii	Rauvolfioideae	DQ660495	—
Nerium oleander	Rauvolfioideae	EF456295, GQ997641	FJ493258, EU531690, GU135391, FN675803
Tabernaemontana bufalina	Rauvolfioideae	DQ660548	—
Tabernaemontana divericata	Rauvolfioideae	Z70187	—
Plumeria rubra	Rauvolfioideae	Z70191	—
Plumeria cubensis	Rauvolfioideae	DQ660536	—
Carissa ovate	Rauvolfioideae	DQ660506	—
Asclepias curassavica	Asclepiadeae	DQ026716	—
Asclepias incarnata	Asclepiadeae	—	GQ248250, DQ006139
Asclepias syriaca	Asclepiadeae	—	HQ596608

Table 2: List of Plant sample of Apocynaceae examined in this study scientific name, subfamily, Voucher, Accession Number of sequences of *matK* and *trnH-psbA* also given

Species	Subfamily	Sample ID	Accession No. of *matK*	Accession No. of *trnH-psbA*
Catharanthus roseus (L.) G. Don	Rauvolfioideae	AUS-MP-03, AUS-MP-36	JN228930, JN228936	JN245988*, JN245984*
Alstonia scholaris (L.) R. Br.	Rauvolfioideae	AUS-MP-05	JN228931	JN245985
Thevetia peruviana (Pers.) Merrill	Rauvolfioideae	AUS-MP-01	JN228929,	JN245983*
Allamanda cathertica L. Allamanda	Rauvolfioideae	AUS-MP-29, AUS-MP-33	JN228933*, JN228935*	JN245987*
Tabernaemontana divaricata (L.) Alston	Rauvolfioideae	AUS-MP-32	JN228934	-
Calotropis gigantea L. R. Br. Ex Ait	Asclepiadeae	AUS-MP-22	JN228932*	JN245986*

* Sequence submitted first time in GenbanK

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
JN245988        ATTTCGAATT TATTTAGAAT ATTGAGACGA CCAT-TTTCT TTCTT-----
JN245984        ATTTCGAATT TATTTAGAAT ATTGAGACGA CCAT-TTTCT TTCTT-----
FJ493259        ---------- ---CGCGCAT GGTGGATTCA CAAT-CCACT GCCTT-----
JN245987        TATTCTAATT AATTTATAAT ATTTCATACT TCAT--TTCT ATTTT-----
JN245983        TATTTTTTTT TTTTTGAGAT ATTTTAATCT TTATATTTTG ATTTTTGATA
GQ982387        TTCTTTTTTT TTTTTGAGAT ATTTTAATCT TCATATTTGG ATTTTTGATA
GQ435037        AATTCTAATT TATTTAGAAT ATTTCATATT TCAT-TTTCA ATTCTAAATT
JN245985        AATTCTAATT TATTTAGAAT ATTTCATATT TCAT-TTTCA ATT-------
GQ435038        AATTCTAATT TATTTAGAAT ATTTCATATT TCAT-TTTCA ATTCTAAATT
GU135394        AATTCGAATT TATTTCGAAT ATTTAATATT TCAT-TTTCA ATTCT-----
GU135392        AATTCGAATT TATTTCGAAT ATTTAATATT TCAT-TTTCA ATTCT-----
GU135391        AATTCTAATT TATTTAGAAT ATTTCATATT TCAT-TTTCA ATTTTAAATT
EU531690        ---------- -----CGCAT GGGGGATTCA CAAT-CCACT GCCTT-----
FJ493258        ---------- ---CGCGCAT GGTGGATTCA CAAT-CCACT GCCTT-----
FN675803        ---------- ---CGCGCAT GGTGGATTCA CAAT-CCACT GCCTT-----
DQ006139        AAGCTCCATC TATCAATGGC -TAAGATCGT CAGTCTTAGT GTATAGGAGT
GQ248250        AAGCTCCATC TATCAATGGC -TAAGATCGT CAGTCTTAGT GTATAGGAGT
JN245986        ---------- ---------- --TTTATTAA TTTT-TTTTT ATCTCGAAT-
HQ596608        -TTTTCGATT TATTTCCTAT -TGGAATTTA TTACATTTTT TTTAT---TT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
JN245988        AAT-TACTTA A----TTATT ATG------- ------TAGT ATTCTTGGTT
JN245984        AAT-TACTTA A----TTATT ATG------- ------TAGT ATTCTTGGTT
FJ493259        GTACCACTTG G----CTAC- TCCGCCCCCT TCCC------ --TATATTTC
JN245987        -----ATCT- ----TTCCGA GAGA------ ---------T TTTCGAATCT
JN245983        ATATTAAATT AAAATTAAGA AAAAGGATTT TTTTTTTAAT TTTAAAATGT
GQ982387        ATATTAAATT TAAATTAAGA AAAAAGGATT TTTTTTTA-T TTTAAAATGT
GQ435037        CAA--AATTG AAAATGAAGA AAAAATACGA ATTTTTTTTT TT--TGAATT
JN245985        CAA--AATTG AAAATGAAGA AAAAATACGA ATTTTTTTTT ----TGAATT
GQ435038        CAA--AATTG AAAATGAAGA AAAAATACGA ATTTTTTTTT TT--TGAATT
GU135394        ATTCTATTTA GAATTTCGTT TCGACCATTT TCTTATTAGT ATTTCTAGTT
GU135392        ATTCTATTTA GAATTTCGTT TCGACCATTT TCTTATTAGT ATTTCTAGTT
GU135391        CAA--AATTG AAAATGAAGA AAAAATACGA ATTTTTTTTT TTTTTGAATT
EU531690        GATCCACTTG G----CTACA TCCGCCCCCT TCACCCCTTC AGTCTATTTC
FJ493258        GATCCACTTG G----CTACA TCCGCCCCCT TCACCCCTTC AGTCTATTTC
FN675803        GATCCACTTG G----CTACA TCCGCCCCCT TCACCCCTTC AGTCTATTTC
DQ006139        TTTTGAAAAA TAAAGGAGCA AAAATCATCT TCTTGATACA ACAAGAAGGT
GQ248250        TTTTGAAAAA TAAAGGAGCA AAAATCATCT TCTTGATACA ACAAGAAGGT
JN245986        ATTTAAATAA AAAATTTAAT ATTTAGAATA TTTTTAGAAT ATTTTGAATA
HQ596608        CTACAATTTA TAGAATATTT TAAAATA--- ------TTCT ATTTCAATTT

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       110        120        130        140        150
JN245988        TTTATTTCAA ------AGAT --ACAAAGAT TCAAAATAA- ----------
JN245984        TTTATTTCAA ------AGAT --ACAAAGAT TCAAAATAA- ----------
FJ493259        TAA--GATTC CAAATTAAAT ATAATATTAA TT--ACAAAT ----------
JN245987        TTGATAATAT -----ATGAT A-ATATGAAA TTCCAATTA- ----------
JN245983        AAGAAAACTT CACAAAAGAT T-GTGAAGAA CGTAACTTAC TTAACTTAA-
GQ982387        AAGAAAACTT CACAAAAGAT T-GTGAAGAA CGTAACTTAC TTAACTTAAA
GQ435037        TAGAAATCTT CACAAAGGAT T-GGGAAGAA CATAACCTA- ----------
JN245985        TAGAAATCTT CACAAAGGAT T-GGGAAGAA CATAACCTA- ----------
GQ435038        TAGAAATCTT CACAAAGGAT T-GGGAAGAA CATAACCTA- ----------
GU135394        TTTATTTCTA TTTCGGAGAT --ACAAAGAT TCAAAATA-- ----------
GU135392        TTTATTTCTA TTTCGGAGAT --ACAAAGAT TCAAAATA-- ----------
GU135391        TAGAAATCTT CACAAAGGAT T-GGGAAGAA CATAACCTA- ----------
EU531690        TATTTATTTC CTATTCGAAT TTATTTAGAA TTTTATAATT ----------
FJ493258        TATTTATTTC CTATTCGAAT TTATTTAGAA TTTTATAATT ----------
FN675803        TATTTATTTC CTATTCGAAT TTATTTAGAA TTTTATAATT ----------
DQ006139        GATATTGCTC CTT---TATT TTCTTTATAT TTGTTACATT ATCA------
GQ248250        GATATTGCTC CTT---TATT TTCTTTATAT TTGTTACATT ATCA------
JN245986        TTTTTTTATA TTTCTACAAT TTATAGAATA TTTTAAAATA T---------
HQ596608        AATATTTCAA TTTAAATATT TAATAGAAAT TTTTAATAAT ----------

                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
JN245988        ---------- ---------- ------AAAA TAATATTAAT TACAAATTCA
JN245984        ---------- ---------- ------AAAA TAATATTAAT TACAAATTCA
FJ493259        ---------- ---------- ---------- AAAAAAAAAA GTATGAT---
JN245987        ---------- ---------- ------AGAA AAATATTTTT TCTTAA----
```

Figure 1: Showing Alignment of 19 sequences of trnH-psbA of Apocynaceae, containing indels of different regions

Contd...

```
JN245983     ---------- ---------- --------TG TAATATTAAT TACAAAT---
GQ982387     TGTAAAATGT AATCTTACTT AACTTAAATG TAATATTTAT TACAAATTTA
GQ435037     ---------- ---------- -------ATG TAATATTTAT TACAAAT---
JN245985     ---------- ---------- -------ATG TAATATTTAT TACAAAT---
GQ435038     ---------- ---------- -------ATG TAATATTTAT TACAAAT---
GU135394     ---------- ---------- ---------- TAATATTAAT TACAAATAAA
GU135392     ---------- ---------- ---------- TAATATTAAT TACAAATAAA
GU135391     ---------- ---------- -------ATG TAATATTTAT TACAAAT---
EU531690     ---------- ---------- ---------- TCTAATTTAT TTAGAAT---
FJ493258     ---------- ---------- ---------- TCTAATTTAT TTAGAAT---
FN675803     ---------- ---------- ---------- TCTAATTTAT TTAGAAT---
DQ006139     ---------- ---------- -----AAAAT TCAAATATCT CAGAAAT---
GQ248250     ---------- ---------- -----AAAAT TCAAATATCT CAGAAAT---
JN245986     ---------- ---------- --------AA AAAATTCTAT TTCTATT---
HQ596608     ---------- ---------- ---------- TTATATTTAT TTCTATTTAA

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
JN245988     -----AAAAA TGAAAAAATA AGAT------ ACTCAAACCT CA-GAAAAC-
JN245984     -----AAAAA TGAAAAAATA AGAT------ ACTCAAACCT CA-GAAAAC-
FJ493259     ---------A CTCAA-ACCT CAGCAAACTA AAAGTCCTTT GCTTTCTCTC
JN245987     ---------- -----AAGTA TGAT------ ACTCAATCAC AAACAAACCT
JN245983     -----AAAAA AAGAAAAATA TGATCCTCAA TCACGAATGT AA-CGAACCT
GQ982387     CAAATAAAAA AAGAAAAATA TGATACTCAA TCACGAATGT AA-CGAACCT
GQ435037     ---------- -AAATAAATA TGAT-----A GAACGAACCT CA-TAAAATA
JN245985     ---------- -AAATAAATA TGAT-----A GAACGAACCT CA-TAAAATA
GQ435038     ---------- -AAATAAATA TGAT-----A GAACGAACCT CA-TAAAATA
GU135394     -----AAAAA T---AAAGTA TGAT------ ACTCAAACCT CA-TAAAAC-
GU135392     -----AAAAA T---AAAGTA TGAT------ ACTCAAACCT CA-TAAAAC-
GU135391     ---------- -AAATAAATA TGAT-----A GAACGAACCT CA-TAAAATA
EU531690     ---------A TTTACTATTT CATTT--TCA ATTCGATTTT ATTTAGAATT
FJ493258     ---------A TTTACTATTT CATTT--TCA ATTCGATTTT ATTTAGAATT
FN675803     ---------A TTTACTATTT CATTT--TCA ATTCGATTTT ATTTAGAATT
DQ006139     ---------- -AAAAAAGAA AATTTTCGAA AGGAAATTCT AAATAAAAT-
GQ248250     ---------- -AAAAAAGAA AATTTTCGAA AGGAAATTCT AAATAAAAT-
JN245986     ---------- ----TAATAT TAAT------ ATTTCAATTT AA---AAATT
HQ596608     ---ATTGAAA TAAATAATAT TAATTTTTAA ATTCTATTTT ATTTAGAATT

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
JN245988     --GAAAAGTC CCTTGCTTTA TCTGTAATGC AAACAAAAAG AATAAAGATT
JN245984     --GAAAAGTC CCTTGCTTTA TCTGTAATGC AAACAAAAAG AATAAAGATT
FJ493259     TAATGAAAAG AAAGAAGAA- ----AAATTT CTAGAA---- ------AATA
JN245987     CATAAGAGTC CCTTGCTTTA TCTGTAAAGC AACCAATA-- ----AAAATT
JN245983     CATAAAAGTT CCTTGCTTTA TCTGTAATGC AAAGAATT-- ----CAAATT
GQ982387     CATAAGAGTT CCTTGCTTTA TCTGTAATGC AAAGAATT-- ----ACAATT
GQ435037     AATAAAAAAA AAGTCCTTT- ---GTAATAC AAATAA---- -----AAGTT
JN245985     AATAAAAAAA AAGTCCTTT- ---GTAATAC AAATAA---- -----AAGTT
GQ435038     AATAAAAAAA AAGTCCTTT- ---GTAATAC AAATAA---- -----AAGTT
GU135394     --TAAAAGTC CTTTGCTTTC TGTGTAATGC AAAGAAAATA AA--AAAAAT
GU135392     --TAAAAGTC CTTTGCTTTC TGTGTAATGC AAAGAAAATA AA--AAAAAT
GU135391     AATAAAATAA AAGTCCTTT- ---GTAATAC AAATAA---- -----AAGTT
EU531690     TGGTTTCGAC CATTTTATT- ----TATTAT TTTGAA---- ------TATT
FJ493258     TGGTTTCGAC CATTTTATT- ----TATTAT TTTGAA---- ------TATT
FN675803     TGGTTTCGAC CATTTTATT- ----TATTAT TTTGAA---- ------TATT
DQ006139     AGAATTTAAA TATAATTTAA ATAGAAATAA ATATAAATTA TT-AAATATT
GQ248250     AGAATTTAAA TATAATTTAA ATAGAAATAA ATATAAATTA TT-AAATATT
JN245986     AATATTATTT ATTTATTTA- ----TTATTT AAATAATAT- -----TAATT
HQ596608     TGCTTTCGAG AATTTTTTT- -TTGTATTTC TGAGATATT- ----TGAATT

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                       310        320        330        340        350
JN245988     TATATAAAAT ACTATAACTA TA--ATAAAT A--------- --AAAATAAA
JN245984     TATATAAAAT ACTATAACTA TA--ATAAAT A--------- --AAAATAAA
FJ493259     C------GAG AAT------- ------AAAT A--------- --AAAAGAAA
JN245987     TATATAAAAT ACTATTAAAT -----TAAAT A--------- --AAAAGAAA
JN245983     TATCGAAAAT ACTAGAAATT TATCGAAAAT ACTAGAATAA ATAAAATAAA
GQ982387     TATCGAAAAA ACTAGAAT-- ------AAAT A--------- --AAAAGAAA
GQ435037     TATATAAAAT ATTAGAAT-- ------AACT A--------- --AAAAGAAA
JN245985     TATATAAAAT ATTAGAAT-- ------AACT A--------- --AAAAGAAA
GQ435038     TATATAAAAT ATTAGAAT-- ------AACT A--------- --AAAAGAAA
GU135394     TATATAAAAT ACTAGAA--- -----TAAAT A--------- --AAAAGAAA
GU135392     TATATAAAAT ACTAGAA--- -----TAAAT A--------- --AAAAGAAA
GU135391     TATATAAAAT ATTAGAAT-- ------AACT A--------- --AAAAGAAA
```

Figure 1: Contd....

```
EU531690    T------GAT AATGTAAC-- ------AAAC A--------- --AAAATAAA
FJ493258    T------GAT AATGTAAC-- ------AAAC A--------- --AAAATAAA
FN675803    T------GAT AATGTAAC-- ------AAAC A--------- --AAAATAAA
DQ006139    TCTATTAAAT ATTTAAATTG ------AAAT AT-------- -TAAATCGAA
GQ248250    TCTATTAAAT ATTTAAATTG ------AAAT AT-------- -TAAATCGAA
JN245986    TTT---AAAT TCTATTTT-- ------ATTT A--------- ---TAATTTT
HQ596608    TTT----GAT AATGTAAC-- ------AACT A--------- --TAAAGAAA

                ....|....| ....|....| ....|....| ....|....| ....|....|
                      360        370        380        390        400
JN245988    ATAAAGGAGC AATA-CCACC CTCTTGATAG AAGAAGAAGG TGA-TTATTG
JN245984    ATAAAGGAGC AATA-CCACC CTCTTGATAG AAGAAGAAGG TGA-TTATTG
FJ493259    ATAAAGGAGC AATA-CCACC CTCTTGATAG AACACGAAGG TGA-TTATTG
JN245987    ATAAAGGAGC AATA-CCCCT TTCTTGATAG AACAAGAAGG TGA-TTATTG
JN245983    ATAAAGGAGC AATA-CTC-C CTCTTGATAG AACAAGAAGG TGGATTATTA
GQ982387    ATAAAGGAGC AATA-CTC-C CTCTTGATAA AACAAGAGGG TGG-TTATTG
GQ435037    ATAAAGGAGC AATA-CCAAC CTCTTGATAG AACAAGAAGG TGA-TTATCG
JN245985    ATAAAGGAGC AATA-CCAAC CTCTTGATAG AACAAGAAGG TGA-TTATCG
GQ435038    ATAAAGGAGC AATA-CCAAC CTCTTGATAG AACAAGAAGG TGA-TTATCG
GU135394    ATAAAGGAGC AATA-CCACC CTCTTGATAG AACAAGAAGG TGA-TTATTG
GU135392    ATAAAGGAGC AATA-CCACC CTCTTGATAG AACAAGAAGG TGA-TTATTG
GU135391    ATAAAGGAGC AATA-CCAAC CTCTTGATAG AACAAGAAGG TGA-TTATCG
EU531690    ATAAAGGAGC AATA-CCGCC CTCTTGATAG AACAAGAAGG TGA-TTATTG
FJ493258    ATAAAGGAGC AATA-CCGCC CTCTTGATAG AACAAGAAGG TGA-TTATTG
FN675803    ATAAAGGAGC AATA-CCGCC CTCTTGATAG AACAAGAAGG TGA-TTATTG
DQ006139    ATAGAATATT TTCAAATATT CTATAAATTG TAGAAATAAA AAAAATGTAA
GQ248250    ATAGAATATT TTCAAATATT CTATAAATTG TAGAAATAAA AAAAATGTAA
JN245986    CTAAAGGAGC AATA-TCAAC TTCTTGTTCT ATCAAGAAG- ------TTTG
HQ596608    ATAAAGGAGC AAAAATCACC TTCTTGTTCT ATCAAGAAGA TGC-TTTTTG

                ....|....| ....|....| ....|....| ....|....| ....|...
                      410        420        430        440
JN245988    CTCCTTTATT TTTCAAAAAC TCCTATACAC TAAGAACAGG G-TCTTAG
JN245984    CTCCTTTATT TTTCAAAAAC TCCTATACAC TAAGAACAGG G-TCTTAG
FJ493259    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCGAG G-TCTTAG
JN245987    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCTGG G-TCTTAG
JN245983    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCAGT G-TCTTAG
GQ982387    CTCCTTTATT TTTCAATAAC TTCTATACAC TAAGACCAGT G-TCTTAG
GQ435037    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCAGA G-TC----
JN245985    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCAGA G-TCTTAG
GQ435038    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCAGA G-TC----
GU135394    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCAGG G-TCTTAG
GU135392    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCAGG G-TCTTAG
GU135391    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCAGA G-TCTTAG
EU531690    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCGGG G-TCTTAG
FJ493258    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCGGG G-TCTTAG
FN675803    CTCCTTTATT TTTCAATAAC TCCTATACAC TAAGACCGGG G-TCTTAG
DQ006139    TAAATTTCAA TAGGAAATAA ATCGAAAAAT AAATAGAAAT AGAACCAG
GQ248250    TAAATTTCAA TAGGAAATAA ATCGAAAAAT AAATAGAAAT AGAACCAG
JN245986    CTCCTTTATT TTTCAAAAAC TCCTATACAC TAAGACTGGC GGTCTTAG
HQ596608    CTCCTTTATT TTTCAAAAAC TCCTATACAC TAAGACTGGC GATCTTAG
```

Figure 1: Contd....

Also, very few sequences are available from Rauvolfioideae in GenBank. Nucleotide composition of *matK* sequences of Apocynaceae is strong A+T bias (average 65.6% for all codon) where a percentage of T (36.5%) is higher than A (29.1%). The rates of substitution among the 3 codon position were almost equal [Figure 2].

matK sequences exhibited indels in multiple of 3 at 5´ end where a 12 bp insertion (641-652 region) was found in *Tabernaemontana divaricata, Tabernaemontana bufalina, Calotropisgigantea,* and *Asclepias curassavica*; next 12 bp insertion (677-688 region) was found in *Tabernaemontana divaricata* and *Tabernaemontana bufalina*, while the other 6 bp insertion (1124-1129 region) was found only in *Calotropis gigantea* and *Asclepias curassavica* of a subfamily Asclepiadaceae [Figure 3].

To evaluate the degree of DNA polymorphism, sequence divergence between and w ithin species were calculated by Kimura 2-parameter (K2P) that revealed high average inter-specific and low intra-specific distances. Highest inter specific distance was 0.119 between *Catharanthus roseus* and *Calotropis gigantea. Thevetia peruviana, Tabernaemontana divaricata, Allamanda cathartica,* and *Alstonia scholaris* also showed high distance with *Calotropis gigantea*. Minimum inter-specific (0.065) was found between *Catharanthus roseus and Tabernaemontana divaricata*; maximum mean divergence within species (0.029) was found in

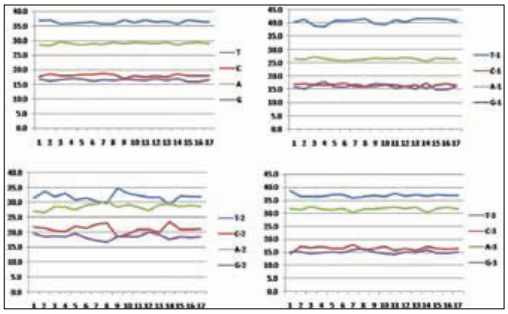

Figure 2: Nucleotide compositions of ~758 bp partial *matK* for the different species of Apocynaceae plants. The frequencies of nucleotide in sequences are present as the total average value for the all the codon positions and for each codon position separately with the accuracy to tenths of a percent. (A, T, G, C shown average value for all codon positions. A-1, T-1, G-1, C-1 shown average value for first codon position. A-2, T-2, G-2, C-2 shown average value for second codon position. A-3, T-3, G-3, C-3 shown average value for third codon position. A+T, A1+T1, A2+T2, A3+A3 represent the average value of A+T bias of total and each codon position.)

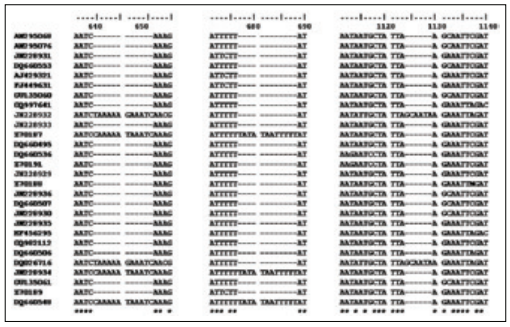

Figure 3: Showing Alignment of 28 sequences of *matK* of Apocynaceae, containing indels of 3 different regions. A 12 bp insertion found in *Tabernaemontana divaricata* (Z70187, JN228934), *Tabernaemontana bufalina* (DQ660548), *Calotropis gigantea* (JN228932), *Asclepias curassavica* (DQ026716) (641-652 region) and in *Tabernaemontana divaricata* (Z70187, JN228934), *Tabernaemontana bufalina* (DQ660548) (677-688 region) and 6 bp insertion in *Calotropis gigantea* (JN228932), *Asclepias curassavica* (DQ026716) (1124-1129 region). * indicate conserve nucleotide

Table 3: Mean divergence (K2P) within (bold number on diagonal) and among (below diagonal) the 6 species of Apocynaceae from southern Assam. (n/c indicates comparable due to only one accession number)

	Species	1	2	3	4	5	6
1	Catharanthus roseus	0.001					
2	Alstonia scholaris	0.068	0.000				
3	Calotropis gigantea	0.119	0.094	n/c			
4	Allamanda cathertica	0.089	0.068	0.109	0.001		
5	Tabernaemontana divaricata	0.075	0.065	0.113	0.080	0.005	
6	Thevetia peruviana	0.083	0.068	0.103	0.070	0.073	0.029

Thevetia peruviana, and minimum mean divergence (0.00) was found in *Alstonia scholaris* [Table 3]. The accuracy of barcoding depends on the barcode gap between intra-specific and inter-specific variation. Sequence variation between species has to be high enough to tell them apart, while the distance within species must be low for them to cluster together.

The different species of Apocynaceae have formed distinctive clusters. Evidently, all the database sequences and the conspecific generated sequences of *Catharanthus roseus, Thevetia peruviana, Tabernaemontana divaricata, Allamanda cathartica,* and *Alstonia scholaris* with Genbank accession numbers are clustered cohesively. However, the members of Asclepiadaceae subfamily, *Calotropis gigantea* and *Asclepias curassavica,* were located at the basal position, hence used as an out group of the phylogenetic tree [Figure 4].

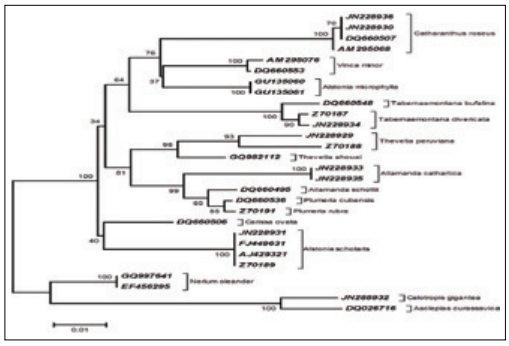

Figure 4: Neighbor-Joining analysis of Kimura2-parameter (K2P) distance of *matK* sequences of Apocynaceae ~758 aligned nucleotide positions of *matK* (Nt. 520-1278) were used in phylogenetic analysis. A total of 1000 bootstrap replicates were calculated for the NJ tree construction.

DISCUSSION

Analyzes of the targeted single loci *matK* (~ 750 bp, Nt. 520-1278) sequences depicted repeat structures with AT-rich regions possessing indels in multiple of 3, and high rate of substitution contributed a considerable number of characters for resolving the phylogeny of

the ethnomedicinal plants of Apocynaceae. Occurrences of indels in *matK* sequences have also been explored to the extent of their applicability as qualitative molecular markers depending upon the size, position, and influence of open reading frame.[22] Several molecular processes are known to create indels, *viz*., polymerase slippages during DNA replication so called slipped-strand mispairing,[23] and due to addition or subtraction of short repeat sequences, which are primarily AT rich.[22] In general, microstructural changes in DNA, such as, insertions and deletions (indels), and inversions in introns and intergenic spacers, have been importantly used both for resolving phylogenetic relationships among the angiosperms[24,25] and for inferring relationships among more closely related taxa.[26] Imperatively, these changes in protein coding gene are very rare phenomenon, because these changes would lead into non-synonymous mutation. But, the observed indels in the presumed barcode region of *matK* happened in multiple of 3 nucleotides, thereby reduced the chances of frameshift mutation and did not interrupt the site of maturase activity in X domain. So, *matK* indels could be utilized as qualitative molecular marker for studies both at the intra-specific and shallow inter-specific levels like the intergenic spacers of CpDNA.

The sequence divergence (K2P) among the studied ethnomedicinal plants of Apocynaceae revealed the highest divergence (0.119) between *Catharanthus roseus* and *Calotropis gigantea* [Table 3]. Moreover, *Calotropis gigantea* being a member of subfamily Asclepiadoideae always consistently high rate of divergences with other 5 studied members of subfamily Rauvolfioideae. Thus, following the notional DNA barcode concept, it can be justifiably infer that use of the partial *matK* sequence having reliable barcode gap as characterized in the study would be appreciably applicable to the species level discrimination of the important ethnomedicinal plants belonging to the family Apocynaceae.

Furthermore, NJ tree showed that the member of Rauvolfioideae subfamily Apocynaceae formed one clade where different species clustered into different subclade. The generated sequences of *Allamanda cathartica* is found closely related to *Allamanda schottii*. It is also close to genera *Thevetia* and *Plumeria*. Although *Alstonia microphylla* and *Alstonia scholaris* are the congeners but placed in different clades, which may be due to polyphyletic nature of Alstonieae.[27] Two sequences from *Neruim oleander* of subfamily Apocynoideae, and two members, *viz. Calotropis gigantea* and *Asclepias curassavica* of subfamily Asclepiadoideae, formed two distinct clades at the basal position of phylogenetic tree. Large sample sizes are required to increase the power of the test in Asclepiadaceae subfamily members, but the poor number of *matK* sequences of Asclepiadaceae in the database remained a limitation of the study, which entail the study using large sample sizes from different geographical location.

Recently, the CBOL Plant Working Group (2009) confirmed and suggested the combination of *matK* with *rbcL* as a universal plant DNA barcode[28] though the low discriminating power of *rbcL* gene is severally reported.[6,8] On the contrary, insertions, deletions, and short sequence repeats were common and often more numerous than single base pair substitution that has been the limitation on the part of *trnH-psbA,* hence remained unable to fulfill the criteria of plant DNA barcoding.[7] Nevertheless, in the present study, intergenic spacer *trnH-psbA* also exhibited persistent problem in obtaining constant bidirectional sequences. Our study showed that species identification of Rauvolfioideae subfamily is possible using phylogenetic analyzes constructed from partial *matK* sequences (Nt. 520-1278), which is comparable to that of the full-length sequences, also had species discrimination power. The observed divergences among the studied species using the partial *matK* sequences maintained a reliable gap, which holds good to the concept of species discrimination through DNA barcoding.[29] Furthermore, the NJ phylogenic tree, based on K2P model, also efficiently distinguished the species under study using the partial *matK* sequence. This gene has been identified as a universal DNA barcode for flowering plants.[30]

CONCLUSION

The *matK* sequences within and among the Rauvolfioideae sub-family have shown indels in multiple of 3, particularly N-terminal regions. The *matK* indels could be utilized as studies both at the intra-specific and shallow inter-specific levels like intergenic spacers of CpDNA. To evaluate the indel containing regions, a more powerful algorithm is needed to calculate the intra- and inter-species comparisons. Our result suggests that *matK* sequence information could help in correct species identification for medicinal plants of Rauvolfioideae and in providing diagnostics for rapid and easier identification of mal species forensics in herbal formulation, which bear insights of similar application in family Apocynaceae.

REFERENCES

1. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLoS One 2011;6:e19254.
2. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A 2005;102:8369-74.

3. Liu Z, Chen SL, Song JY, Zhang SJ, Chen KL. Application of deoxyribonucleic acid barcoding in Lauraceae plants. Pharmacogn Mag 2012;8:4-11.
4. Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP. Commercial teas highlight plant DNA barcode identification successes and obstacles. Sci Rep 2011;1:42.
5. Hebert PD, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc Biol Sci 2003;270: 313-21.
6. Asahina H, Shinozaki J, Masuda K, Morimitsu Y, Satake M. Identification of medicinal Dendrobium species by phylogenetic analyses using matK and rbcL sequences. J Nat Med 2010;64:133-8.
7. Bruni I, De Mattia F, Galimberti A, Galasso G, Banfi E, Casiraghi M, *et al.* Identification of poisonous plants by DNA barcoding approach. Int J Legal Med 2010;124:595-603.
8. Sun XQ, Zhu YJ, Guo JL, Peng B, Bai MM, Hang YY. DNA barcoding the Dioscorea in China, a vital group in the evolution of monocotyledon: Use of matK gene for species discrimination. PLoS One 2012;7:e32057.
9. Newmaster SG, Ragupathy S. Ethnobotany genomics - discovery and innovation in a new era of exploratory research. J Ethnobiol Ethnomed 2010;6:2.
10. Van Der Heijden R, Jacobs DI, Snoeijer W, Hallard D, Verpoorte R. The Catharanthus alkaloids: Pharmacognosy and biotechnology. Curr Med Chem 2004;11:607-28.
11. Wong SK, Lim YY, Abdullah NR, Nordin FJ. Antiproliferative and phytochemical analyses of leaf extracts of ten Apocynaceae species. Pharmacogn Res 2011;3:100-6.
12. Nayak S, Nalabothu P, Sandiford S, Bhogadi V, Adogwa A. Evaluation of wound healing activity of Allamanda cathartica. L. and Laurus nobilis. L. extracts on rats. BMC Complement Altern Med 2006;6:12.
13. Kala CP, Dhyani PP, Sajwan BS. Developing the medicinal plants sector in northern India: Challenges and opportunities. J Ethnobiol Ethnomed 2006;2:32.
14. Hilu KW, Borsch T, Muller K, Soltis DE, Soltis PS, Savolainen V, *et al.* Angiosperm phylogeny based on matK sequence information. Am J Bot 2003;90:1758-76.
15. Barthet MM, Hilu KW. Evaluating evolutionary constraint on the rapidly evolving gene matK using protein composition. J Mol Evol 2008;66:85-97.
16. Liu Y, Zhang L, Liu Z, Luo K, Chen S, Chen K. Species identification of Rhododendron (Ericaceae) using the chloroplast deoxyribonucleic acid PsbA-trnH genetic marker. Pharmacogn Mag 2012;8:29-36.
17. Strochova H, Olson MS. The architechture of the chloroplast psbA-trnH non coding region in angiosperms. Plant Syst Evol 2007;268:235-56.
18. Ragupathy S, Newmaster SG, Murugesan M, Balasubramaniam V. DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India. Mol Ecol Resour 2009;9(Suppl s1):S164-71.
19. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876-82.
20. Hall TA. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 1999;41:95-8.
21. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596-9.
22. Hilu KW, Alice LA. Evolutionary implications of matK indels in Poaceae. Am J Bot 1999;86:1735-41.
23. Kelchner SA. The evolution of non-coding chloroplast DNA and its application in plant systematic. Ann Mo Bot Gard 2000;87:499-527.
24. Graham SW, Reeves PA, Burns AC, Olmstead RG. Microstructural changes in non-coding DNA: Interpretation, evolution and utility of indels and inversions in basal angiosperm phylogenetic inference. Int J Plant Sci 2000;161:S83-96.
25. Ingvarsson PK, Ribstein S, Taylor DR. Molecular evolution of insertions and deletion in the chloroplast genome of silene. Mol Biol Evol 2003;20:1737-40.
26. Golenberg EM, Clegg MT, Durbin ML, Doebley J, Ma DP. Evolution of a noncoding region of the chloroplast genome. Mol Phylogenet Evol 1993;2:52-64.
27. Simoes AO, Livshultz T, Conti E, Endress ME. Phylogeny and systematic of the Rauvolfioideae (Apocynaceae) based on molecular and morphological evidence. Ann Mo Bot Gard 2007;94:268-97.
28. Hollingsworth MP, Forrest LL, Spouge LJ, Hajibabaei M, Ratnasingham S, van der Bank M, *et al.* A DNA barcode for land plants. Proc Natl Acad Sci U S A 2009;106:12794-7.
29. Gao T, Sun Z, Yao H, Song J, Zhu Y, Ma X, *et al.* Identification of Fabaceae plants using the DNA barcode matK. Planta Med 2011;77:92-4.
30. Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, *et al.* DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci U S A 2008;105:2923-8.

Cite this article as: Citation will be included before issue gets online***

Source of Support: Infrastructural support from Department of Biotechnology, Govt. of India. **Conflict of Interest:** No.

Author Queries???

AQ1: Kindly provide full form???