Bionature, 38(4) 2018 : 266-276 ISSN: 0970-9835 (P), 0974-4282 (O) © *Bionature*

MOLECULAR IDENTIFICATION AND PHYLOGENY OF *MUSA* SPECIES FROM NORTH-EASTERN INDIA BY INTERNAL TRANSCRIBED SPACER 2 OF NUCLEAR DNA

GURUMAYUM RANIBALA[1], SOROKHAIBAM SURESHKUMAR SINGH[1*], WAHENGBAM ROBINDRO SINGH[1] AND MOHAMMED LATIF KHAN[2]

[1]Department of Forestry, North Eastern Regional Institute of Science and Technology (Deemed to be University), Nirjuli-791 109 (Itanagar), Arunachal Pradesh, India.
[2]Department of Botany, Harisingh Gour Vishwavidyalaya, Sagar-470003, Madhya Pradesh, India.
Email: suresh@nerist.ac.in, sksorokhaibam@gmail.com

Article Information

Editor(s):
(1) Said Elsayed Desouky Mohammed, Al-Azhar University, Egypt.
(2) Gabriela Civeira, University of Belgrano, Argentina.
(3) Ahmed Medhat Mohamed Al-Naggar, Cairo University, Egypt.
Reviewers:
(1) Hamit Ayberk, Istanbul University, Turkey.
(2) Davi Coe Torres, Universidade Federal do Ceará (UFC), Fortaleza-CE, Brazil.
(3) Lia Hapsari, University of Brawijaya, Malang, Indonesia.

Received: 13th May 2018
Accepted: 7th August 2018
Published: 20th August 2018

Short Research Article

ABSTRACT

The internal transcribed spacer region of the ribosomal RNA gene cluster in the nuclear DNA has been used as one of the most informative molecular markers for identification and phylogenetic analysis of fungi and plants in the last three decades. The aim of the present study was to evaluate the application of ITS2 locus of rRNA region in nuclear DNA as molecular marker in the identification and molecular phylogeny of 15 *Musa* specimens (Musaceae) in North-Eastern India. The sequence characteristic of the ITS2 locus revealed a length variation from a minimum of 210 bp in *Musa mannii* to a maximum of 223 bp in *M. balbisiana* with an average of 214 bp. The GC contents of the ITS2 locus also varied from a minimum of 64.95% in *M. markkui* to 70.97% in *M. balbisiana* with an average of 69.01%. Sequence homology analysis of the ITS2 regions revealed sequence similarity from a minimum of 90% for *M. markkui* followed by 97% for *M. acuminata* and 98% for *M. nagensium* while the remaining species showed 99% to 100% similar sequence identities in public database. The phylogenetic analysis based on ITS2 sequence revealed a monophyletic origin of all taxa under different genera of the family Musaceae. The clade I comprising of 4 representative taxa of the section *Callimusa* (*Australimusa, Callimusa* and *Ingentimusa*) found only in few Southeast Asian countries and Australia were clearly separated from clade II which consists of 21 taxa of the section *Musa* (*Eumusa* and *Rhodochlamys*) including the 15 taxa from North-Eastern India. The clade II was further separated into two sub-clades, *M. balbisiana* group in clade IIa and *M. acuminata* and *Rhodochlamys* group in clade IIb. A new taxon with poorly described genetic and molecular characteristics, *M. puspanjaliae* was found to show an independent lineage of evolution within the sub-clade IIb suggesting closer genetic similarity with *M. acuminata* as compared to the *M. balbisiana*. Low bootstrap values (<50%) were observed in several branches of the phylogenetic tree of Musaceae reconstructed based on ITS2 sequences except in case of clade I (96%) and few sub-clades (51% to 98%). Therefore, it is concluded that ITS2 locus can be employed as one of the potential molecular marker for the identification of *Musa* specimens but inferences of phylogenetic tree based on ITS2 sequence are poorly supported as revealed by low bootstrap values.

Keywords: Arunachal Pradesh; BLASTn; ITS2; *Musa;* Musaceae; phylogeny; *Rhodochlamys.*

INTRODUCTION

The internal transcribed spacers (ITS) of the ribosomal (rRNA) gene cluster in nuclear DNA has been used as one of the most informative sequence region for identification and phylogenetic analysis of fungi and plants in the last three decades. The sequences in the internal transcribed spacer 1 and 2 (ITS1 and ITS2) regions located between 18S rRNA and 26S rRNA genes have been shown to carry conserved parental and species information which allows identification of an organism accurately up to the lowest rank of the organism, i.e. species. These reasons made the ITS sequences as one of the most popular sequences for phylogenetic inference at the generic and infrageneric levels in plants (Alvarez and Wendel, 2003). The ITS region has also been considered as one of the popular molecular signatures leading to their application as DNA barcodes in fungi and plants despite short sequence lengths (Chase and Fay, 2009; Bellemain et al. 2010; Yao et al. 2010; Schoch et al. 2012). Although the use of ITS is accepted as universal DNA barcodes for fungi, its application in the identification of plants and subsequent use as barcode is increasing in the recent past due to poor delineation capacities of plastid genome sequences (*mat*K, *rbc*L, *psb*A-*trn*L, *rpo*C1 and *ycf*5) up to species levels (Chen et al. 2010). There have been numerous studies on the use of ITS region in phylogeny and barcoding of Angiospermae families which indicates that ITS region must be included as one of the potential markers for DNA barcoding in plants (Li et al. 2010; Tripathi et al. 2013). The studies on evaluation of validity of ITS sequences as universal and novel DNA barcode for 50790 plants (angiosperms, gymnosperms, ferns and mosses) have revealed the region as one of the most promising phylogenetic marker (Yao et al. 2010). Other studies on molecular phylogeny and DNA barcoding of plants at family levels have also shown the significance of ITS region as a potential phylogenetic marker and DNA barcode loci (Wolfe et al. 2002; David, 2009; Gao et al. 2010; Xing et al. 2011; Ashfaq et al. 2013). One of the disadvantages of using whole ITS region in the identification of plants is that the occurrences of sequence length variations, especially difficulties in obtaining a complete ITS1 sequence and poor species identification capacity of ITS1 as compared to ITS2 region. Moreover, ITS2 locus has shown to be a promising marker that can be used as a barcode for the identification of plants as compared to whole ITS region (Yao et al. 2010; Han et al. 2013). A similar study also reported 93% successful identification of 6600 medicinal plant samples belonging to 753 genera and 4800 species by ITS2 sequences (Chen et al. 2010). Though most of the studies have been confined to dicot families of the Angiospermae, ITS markers have also been employed for assessing genetic diversity and phylogeny of monocot families including Musaceae (Carreel et al. 2002; Lamare et al. 2017). Musaceae is a monocotyledonous family comprising of three genera, *Musa*, *Musella* and *Ensete*. The taxonomy of Musaceae, particularly *Musa* genus based on the morphological characteristics, have been controversial in the past due to close similarities in vegetative characters which limits its application in the identification of species and phylogenetic analysis (Li et al. 2010; Hribova et al. 2011). There have been recent developments in molecular taxonomy and phylogenetic analysis of the genus *Musa* during the last two decades. The structure and diversity of the ITS region in

87 representatives of the family Musaceae provided the first detailed information on the ITS sequence diversity in the genus *Musa* and observed to contain conserved parental ITS sequence even among the intraspecific banana hybrids (Hribova et al. 2011). A phylogenetic analysis on 39 accessions under 28 species of *Musa* using combined approaches of ITS region and *trn*L sequences revealed evolutionary diversification of this plant group on biogeographical contexts (Liu et al. 2010). These reports revealed that there is wide acceptance by the scientists on using ITS region as an important marker in phylogenetic inference and DNA barcoding of plants including the family Musaceae. North-Eastern Region of India falls under the Indo-Burma and the Eastern Himalayan Biodiversity Hot Spots region with rich floral and faunal diversity. The region is enriched with 27 of the 82 wild species reported around the world. It includes 14 small pseudostem ornamental species under the section *Rhodochlamys* and 13 large pseudostem species of the section *Eumusa* under the genus *Musa* along with several cultivated varieties in traditional home gardens. The region has been considered as one of the important origins of wild *Musa* species (Hakkinen and Sharrock, 2002; Molina and Kudagamage, 2002; Hakkinen, 2005). Five species of *Rhodochlamys* (*Musa argentii, M. arunachalensis, M. kamengensis, M. markkui,* and *M. markkuana)* and two species of *Eumusa* (*M. puspanjaliae* and *M. nagalandiana*) have been recently discovered and described as new species of *Musa* from the region during the last five years. These new species are reported from small and selected areas of the region leaving a scope of high chance for the discovery of several other species in the remaining large unexplored areas in near future. The identification of wild *Musa* species occurring in the region are primarily based on the external morphological characters which are frequently influenced by several environmental factors which sometimes leads to an inaccurate identification due to highly similar morphological characters among the members of the genus. Therefore, an attempt has been made in this study to highlight the importance of molecular identification of selected wild *Musa* species from the North-Eastern Region of India by using ITS2 locus of nuclear DNA to strengthen morphological identification and also elucidate phylogenetic relationship with their distant relatives in the world.

MATERIALS AND METHODS

Sample Collection and Molecular Analysis

A total of 15 *Musa* specimens and one specimen of *Ensete glaucum* were collected from different areas of Arunachal Pradesh and Manipur, two states of North-Eastern India. Eight of the *Musa* specimens belongs to the section *Eumusa* and seven specimens belongs to the section *Rhodochlamys* under the genus *Musa* (Musaceae). The member of the genus *Ensete* (*Ensete glaucum*, Musaceae) was also included to serve as an outgroup species. The details of all the specimens are presented in the Table 1.

Total genomic DNA were extracted from fresh cigar (young rolling) leaves or from stored leave samples in the 2% CTAB buffer using a modification of CTAB method (Rogers and Bendich, 1994). The ITS region of rRNA gene cluster on nuclear DNA was amplified using a pair of primer, ITS1 (5'-TCCGTAGGTGAACCTGCGG-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3') (White et al. 1990). The PCR amplification was carried out in a thermal cycler (GenAmp 2700, ABI, USA) with initial denaturation at

Table 1. Details of *Musa* specimens collected from Manipur and Arunachal Pradesh, North-Eastern India

Sl. no.	Name of the species	Local names	Place of collection	Voucher no.
1	*M. acuminata*	Jahaji (Tall)	Nirjuli, Arunachal Pradesh	AR032
2	*M. flaviflora*	Kas kol	Nirjuli, Arunachal Pradesh	AR033
3	*M. velutina*	Kodum	Nirjuli, Arunachal Pradesh	AR034
4	*M. balbisiana* 1	Bhim kol	Nirjuli, Arunachal Pradesh	AR036
5	*M. cheesmanii*	Kuhi kulu	Itanagar, Arunachal Pradesh	AR038
6	*M. nagensium*	Dup kulu	Miao, Arunachal Pradesh	AR039
7	*M. itinerans*	Langak	Miao, Arunachal Pradesh	AR040
8	*M. mannii*	Miao kulu	Diyun, Arunachal Pradesh	AR041
9	*M. aurantiaca*	Kodok	Miao, Arunachal Pradesh	AR042
10	*M. markkui*	-	Anjaw, Arunachal Pradesh	AR043
11	*M. chunii* 1	-	Miao, Arunachal Pradesh	AR044
12	*M. chunii* 2	-	Salangam, Arunachal Pradesh	AR045
13	*M. puspanjaliae*	Dura kulu	Anjaw, Arunachal Pradesh	AR046
14	*M. laterita*	Laphu Thambal	Bishnupur, Manipur	MN006
15	*M. balbisiana* 2	Changbi changbi	Kodompokpi, Manipur	MN024
16	*E. glaucum*	Laphu Lempra	Leimaram, Manipur	MN030

94℃ for 4 min, denaturation at 94℃ for 45s, primer annealing at 56℃ for 45s, polymerization at 72℃ for 1 min (35 cycles) and final extension of reaction at 72℃ for 5 min (1 cycle). The 25 µl PCR reaction mixture comprised of 1 µl Taq DNA Polymerase (1U, Hi-Media), 2.5 µl of 10 x PCR buffer (with $MgCl_2$), 0.5 µl of 10 mmol dNTPs, 1.0 µl each of 10 pmol forward and reverse primers, 1.0 µl of DNA template (50 ng) and 18.0 µl Milli-Q water. PCR products were purified using the sodium acetate and ethanol precipitation method. The purified PCR products were sequenced (double reads) at SciGenom Labs, Kerala, India. The sequences were assembled, annotated and ITS2 sequences were extracted using the bioinformatics software Geneious Pro 5.6 (Kearse et al. 2012). The ITS2 sequences were used to identify the *Musa* species collected from the study region using the online bioinformatics tool, BLASTn (NCBI) from the public database (Gen Bank).

Phylogenetic Analysis

Phylogenetic analyses were conducted using the MEGA 6.0 software (Tamura et al. 2013). ITS2 sequences of 10 *Musa* species from different parts of the world (*Musa balbisiana*, *M. acuminata subsp. burmanicoides, M. acuminata subsp. malaccensis*, *M. yunnanensis*, *M. siamensis*, *M. rubinea*, *M. peekelii*, *M. textilis*, *M. beccarii*, *M. ingens* and another outgroup species, *Ravenala madagascariensis* (Strelitziaceae)) were retrieved from the public database (GenBank, NCBI). These sequences were analysed together with the 16 sequences from the present study for phylogenetic inferences. The most appropriate model of nucleotide substitution was tested for generation of phylogenetic tree. Models with the lowest BIC scores (Bayesian Information Criterion) were considered to give the best substitution pattern. The evolutionary history was inferred using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993). The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the analysed taxa and the percentage of replicate trees in which the associated taxa clustered together are shown next to the branches (Felsenstein,

1985). Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pair wise distances estimated using the Maximum Composite Likelihood (MCL) approach. The analysis involved 27 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 191 positions in the final dataset.

RESULTS

ITS2 Sequence Characteristics of *Musa* Species from North-Eastern India

The ITS2 sequence characteristics and GenBank Accession Numbers of 15 *Musa* specimens and one *Ensete glaucum* are given in the Table 2. The nucleotide lengths of ITS2 varied from 210 bp to 223 bp among 15 specimens with an average of 214 bp. The ITS2 sequence length of *E. glaucum* was 222 bp. The GC contents varied from the lowest of 64.95% in *M. markkui* to 70.97% in *M. balbisiana* with an average of 69.01%. The GC content of *E. glaucum* was found to be 67.12%.

Identification of *Musa* Species by ITS2 Sequences

The details of BLASTn analysis for the identification of *Musa* species is given in Table 3. The sequence homology analysis revealed the lowest identity of 90% similarity for *M. markkui* followed by 97% for *M. acuminata*, 98% for *M. nagensium* and 99% in 5 species (*M. itinerans, M. mannii, M. velutina, M. aurantiaca and M. puspanjaliae*). The highest of 100% was found in seven species (*M. balbisiana, M. nagensium, M. itinerans, M acuminata, M. mannii, M. velutina* and *M. aurantiaca*). The outgroup species, *E. glaucum* also displayed 99% sequence identity to the subject sequences from the database. The query coverage was 100% for all sequences except *M. markkui* which showed only 97% as compared to other sequences.

Molecular Phylogeny of *Musa* species from North-Eastern India

The total aligned lengths of the 27 sequences of ITS2 were 232 nucleotide positions within which 116 bases (50%) were variable and 43 bases were parsimony informative sites (37.07% of the variable positions). The strict consensus tree strongly revealed monophyletic lineage where members of the family Musaceae descended from a common evolutionary ancestor or ancestral group (Figure 1). Within the family Musaceae, the phylogenetic tree generated 2 main clades comprising of all the *Musa* species against the outgroup taxa, *E. glaucum* and *Ravenala madagascariensis*. Based on the recent sectional reclassification, the clade I comprised of 4 species of the section *Callimusa* (*M. ingens*, *M. beccarii*, *M. peekelii* and *M. textilis*). The clade II comprised of 21 taxa belonging to the section *Musa* which was further branched into sub-clades IIa and IIb. The sub-clade IIa consisted of 7 taxa consisting of *M. balbisiana* (3 taxa including 1 from GenBank), *M. cheesmanii, M. itinerans, M. nagensium* and *M. flaviflora.* This sub-clade may be referred to as *M. balbisiana* group since all the species except *M. flaviflora* carried specific genetic and morphological characters related to *M. balbisiana*. The sub-clade IIb with 14 taxa had been separated into two clans where IIb(i) had only one taxon, *M. puspanjaliae* while clan IIb(ii) comprising of 13 taxa of *Musa* previously described as *Eumusa* (4 taxa) and *Rhodochlamys* (9 taxa) species. This sub-clade may be referred to as *M. acuminata* group because all the species carried specific genetic and morphological characters related to *M. acuminata*. The taxon, *M. puspanjaliae* is a newly discovered

species from Arunachal Pradesh with unique morphological characters and its lonely placement at an intermediate position between *M. balbisiana* and *M. acuminata* groups suggests the requirement of further studies on this particular species.

Table 2. ITS2 sequence characteristics of the Musaceae specimens

Sl. no.	Species	Gen bank Acc. no.	ITS2 (bp)	GC (%)
1	*M. mannii*	KU512934	210	68.57
2	*M. flaviflora*	KU512943	211	69.67
3	*M. nagensium*	KU512925	212	69.34
4	*M. velutina*	KU512935	213	69.01
5	*M. aurantiaca*	KU512936	213	68.54
6	*M. laterita*	KU512938	213	69.48
7	*M. chunii* 1	KU512939	213	68.54
8	*M. chunii* 2	KU512940	213	68.54
9	*M. cheesmanii*	KU512942	213	69.01
10	*M. puspanjaliae*	KU512944	213	69.01
11	*M. acuminata*	KU512933	214	69.16
12	*M. markkui*	KU512937	214	64.95
13	*M. balbisiana 2*	KU512924	216	70.83
14	*M. itinerans*	KU512926	217	69.59
15	*M. balbisiana* 1	KU512923	223	70.97
16	*E. glaucum*	KU512941	222	67.12

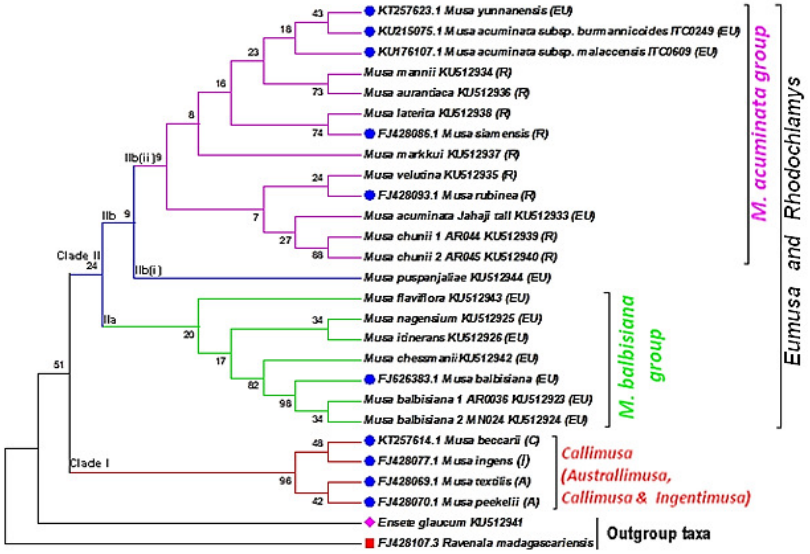

Fig. 1. Phylogenetic tree of *Musa* species from North-Eastern India with *Ensete glaucum* and *Ravenala madagascariensis* as outgroups. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Solid blue circles indicate ITS2 sequences of *Musa* species from GenBank, pink diamond and red square represent outgroup taxa. Red branch shows taxa from three sections, *Australimusa* (A), *Callimusa* (C) and *Ingentimusa* (I). Green, blue and pink branches indicate taxa belonging to *Eumusa* and *Rhodochlamys* sections

Table 3. Result of the BLASTn analysis of ITS2 sequences for the identification of Musaceae specimens

Sl.	Species	Total score	Max. score	Sequence Coverage (%)	E-value	Identity (%)	Acc. No.	Species matched from database
1	*M. balbisiana 1*	412	412	100	2.00E-111	100	KR632991.1	*M. balbisiana* voucher MB4
2	*M. balbisiana 2*	399	399	100	1.00E-107	100	KR632998.1	*M. balbisiana* voucher MB11
3	*M. nagensium*	364	364	100	4.00E-97	98	JF977090.1	*M. nagensium* voucher GXJ562
4	*M. itinerans*	396	396	100	2.00E-106	99	JF977082.1	*M. itinerans* voucher GXJ535
5	*M. acuminata*	363	363	100	2.00E-96	97	FR727861.1	*M. acuminata* isolate 0653con3
6	*M. mannii*	377	377	100	6.00E-101	99	FR727889.1	*M. mannii* isolate 1411con1
7	*M. velutina*	381	381	100	4.00E-102	99	KU215118.1	*M. velutina* voucher ITC0638
8	*M. aurantiaca*	388	388	100	3.00E-104	99	FJ428090.1	*M. aurantiaca*
9	*M. markkui*	274	274	97	8.00E-70	90	FR727943.1	*M.* AAB Group isolate 1135con1
10	*M. laterita*	394	394	100	6.00E-106	100	KU512938.1	*M. laterita* voucher MN006
11	*M. chunii1*	394	394	100	6.00E-106	100	FR727883.1	*M. mannii* isolate 0543con1
12	*M. chunii2*	394	394	100	6.00E-106	100	FR727883.1	*M. mannii* isolate 0543con1
13	*M. cheesmanii*	394	394	100	6.00E-106	100	KU215106.1	*M. cheesmanii* voucher ITC1519
14	*M. flaviflora*	390	390	100	7.00E-105	100	KR633009.1	*M. ornata* voucher MO10 18S
15	*M. puspanjaliae*	377	377	100	6.00E-101	99%	KR559870.1	*M. acuminata* voucher MA11
16	*E. glaucum*	392	392	100	2.00E-105	99%	KU215088.1	*E. glaucum* voucher ITC0775

DISCUSSION

Sequence Characteristics and Species Delineation Capacity of ITS2 Locus in *Musa* Species

The ITS2 sequence lengths (210 to 223 bp) of 15 *Musa* specimens reported in this study are in conformity with previous studies (Yao et al. 2010; Hribova et al. 2011). Hribova et al. (2011) reported the sequence lengths of ITS2 region (between 205 to 227 bp) for 87 accession of *Musa* collected from International Transit Centre (ITC) at Belgium. The average nucleotide length of these *Musa* species (214 bp) observed in this study were found to be shorter than those of monocots (221bp), dicots (236bp), gymnosperms (240bp), ferns (224 bp) and mosses (260 bp) as also reported by Yao et al. (2010). The GC content of the genomic DNA was predicted to significantly affect the genome functioning and species ecology (Smarda et al. 2014). The average GC content (69%) of the ITS2 regions of 15 *Musa* species recorded in this study were found to be remarkably higher than other taxa such as dicots, gymnosperms, ferns and mosses (Yao et al. 2010; Smarda et al. 2014). These finding suggests that ITS2 is highly conserved and have stable genetic component in the genus *Musa*. Therefore, it can be assumed that possible lower rate of mutations likely to occur in the ITS2 region, thereby conserving parental characters while allowing minimum variations to reveal species level identities.

The ITS2 locus could correctly identify 10 (67%) out of the 15 taxa analysed while 5 taxa (33%) could not be correctly identified despite 100% sequence coverage and high percentage of similar sequence identity. The reason for poor identification properties of 5 sequences of ITS2 in this study was due to the unavailability of reference sequences in the public database. However, after submission of these sequences to the GenBank database, a repeated analysis revealed 100% identification of all the species by ITS2 sequence.

Phylogenetic Relationship of *Musa* Species from North-Eastern India

The phylogenetic tree of the inferred from the analysis of ITS2 sequence revealed a monophyletic lineage of all taxa of two genera, *Musa* and *Ensete* within the family, Musaceae. Previous studies on phylogenetic relationship using nuclear and chloroplast DNA sequences revealed similar results of monophyletic origin of Musaceae (Kress and Specht, 2005). Li et al. (2010) and Liu et al. (2010) also confirmed monophyly of Musaceae and their studies provided further monophyletic lineages of the three genera, *Musa, Ensete* and *Musella* into distinct clads. The present study also recorded similar findings where the two genera were distinctly separated into two clades where *Ensete* represented by a single species was separated from other 25 taxa of *Musa* in clade I and clade II. The separation of a single taxon of *Ensete* from *Members* of the genus *Musa* based on the ITS2 sequence directly reflected the morphological and genetic variations of the species among these two genera. There was further separation of the genus *Musa* into clade I and clade II. The clade I consisted of 4 taxa belonging to the sections *Callimusa, Australimusa* and *Ingentimusa* which are classified under the section *Callimusa* at present. The species in clade I are reported to be confined geographically to a few South-Eastern Asian countries (Indonesia, Malaysia, Philippines, Thailand and New Guinea) and Australian region. The remaining 21 taxa in clade II under the sections, *Eumusa* and *Rhodochlamys* are geographically distributed in the South Asian

and Southeast Asian countries (Bangladesh, China, South and North Eastern India, Myanmar, Thailand, etc.). The separation of these two clades revealed that all the taxa in clade I was geographically and genetically different than those of the taxa in clade II. These results are also in support of the recent sectional reclassification of the genus *Musa*, by merging taxa belonging to three sections in the clade I into a single section, *Callimusa* and all the taxa in clade II into a single section, *Musa* (Li et al. 2010; Hakkinen, 2013). The group of 21 taxa in clade II were further separated into two sub-clades, IIa and IIb. The clade IIa comprised of 7 taxa which had genetic and morphological characters of *M. balbisiana,* hence this sub-clade was designated as *M. ballbiisiana* group. The other sub-clade, clade IIb with 14 taxa comprised on all taxa of *Rhodochlamys* and *M. acuminata.* Since this sub-clade consisted of 3 taxa of *M. acuminata,* it was designated as *M. acuminata* group. This finding is in contrast to the reports that clear separation of taxa was not observed between *M. balbisiana* and *M. acuminata* when ITS based phylogenetic analysis is conducted among the members of the genus *Musa* (Li et al. 2010; Liu et al. 2010). The position of the newly described taxon, *M. puspanjaliae* in a separate lineage [(clade II(i)] from the members of *Rhodochlamys* and *M. acuminata* [(clade IIb(ii)] needs further investigation. This taxon was found to possess morphological characters similar to the *M. balbisiana* but ITS2 sequence based analysis revealed more close similarity with *M. acuminata* at genetic level. It can be assumed that this taxon might have evolved parallel to the *M. acuminata* and *M. balbisiana* from a common ancestor. Additionally, the presence of persistent bract and nature of bract lifting of this taxon were closely similar to the members of *Ensete.* All the 15 taxa of *Musa* (*Eumusa* and *Rhodochlamys*) from North-Eastern India were placed together in the clade II. These taxa have shown distant relationship to those belonging to *Callimusa* (*Australimusa, Callimusa* and *Ingentimusa*). It may be noted that there has been no report of the occurrence of any taxon of the section *Callimusa* in the North-Eastern India despite the region is considered as one of the origins of *Musa* species in the world. Observation of low bootstrap values (<50%) in several branches of the phylogenetic tree suggests poor support of important inferences derived from the phylogenetic analysis although low bootstrap values do not mean that the relationships are false (Ghosh and Mallick, 2008). Moreover, it has been argued that the bootstrap values do not “prove” anything conclusively and need to be used cautiously though they are considered to be very useful indicators of the reliability of a phylogenetic tree topology (Higgs and Attwood, 2005).

CONCLUSIONS

The ITS2 sequence characteristics of 15 taxa of *Musa* revealed an average sequence length of 214 bp and 69.01% GC content. The use of ITS2 as molecular marker could identify 67% of the *Musa* species reported in this study. The percentage of identification increased to 100% when the ITS2 sequences of 5 taxa were submitted in GenBank. ITS2 sequence based phylogenetic analysis revealed monophyletic evolution of Musaceae while sectional classifications were clearly observed within the genus, *Musa*. However, it was observed that the bootstrap values in the phylogenetic tree and its branches (except few branches with high bootstrap values) were observed to be low which suggests that poor support of the phylogenetic tree topology and inferences. Therefore, it can be concluded that the identification of taxa belonging to

the genus *Musa* may employ ITS2 locus as one of the potential molecular marker for the species identification but inferences of phylogenetic tree based on ITS2 sequences does not provide a strong support of the important inferences due to low bootstrap values.

ACKNOWLEDGEMENTS

Authors gratefully acknowledge the Department of Biotechnology, Government of India sponsored Bioinformatics Infra-structure Facility (DBT-BIF), Department of Forestry for providing computational facilities of sequence analysis.

AUTHORS' CONTRIBUTIONS

This work was carried out in collaboration between all authors. Authors GR and SSS designed the study, performed the experiments and statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author WRS was involved in collection of samples, literature searches and data analyses of the study. Author MLK supervised experimental design of the study, data interpretation, analysis and revised the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

References

Alvarez, I. and Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29:417-434.

Ashfaq, M., Asif, M., Anjum, Z. I. and Zafar, Y. (2013). Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). Mol. Eco. Res. 13:573-582.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P. and Kauserud, H. (2010). ITS as an environmental DNA barcode for fungi: An *in silico* approach reveals potential PCR biases. BMC Microbio. 10:189.

Carreel, F., Gonzalez de Leon, D., Lagoda, P., Lanaud, C., Jenny, C. and Horry, J. P., et al. (2002). Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. Genome. 45(4):679-692.

Chase, M. W. and Fay, M. F. (2009). Barcoding of plants and fungi. Science. 325(5941):682-683.

Chen, S., Yao, H., Han, J., Liu, C., Song, J. and Shi, L., et al. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLos ONE. 5(1):e8613. Available:http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008613

David, M. S. (2009). DNA barcoding will frequently fail in complicated groups: An example in Wild Potatoes. Amer. J. Bot. 96:177–1189.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39:783-791.

Gao, T., Yao, H., Song, J., Liu, C., Zhu. Y. and Ma, X., et al. (2010). Identification of medicinal plants in the family *Fabaceae* using a potential DNA barcode ITS2. J. Ethnopharmacol. 130:116-121.

Ghosh, Z. and Mallick, B. (2008). Bioinformatics, principles and applications. Oxford University Press, New Delhi. 212-249.

Hakkinen, M. and Sharrock, S. (2002). Diversity in the genus *Musa*- focus on *Rhodochlamys*. INIBAP, annual report 2001. INIBAP, Montpellier, France. 16-23.

Hakkinen, M. (2005) Ornamental bananas- focus on the section *Rhodochlamys*. Bulletin Helic. Soc. Internat. 12(2):1-7.

Hakkinen, M. (2013). Reappraisal of sectional taxonomy in *Musa* (Musaceae). Taxon 62(4):809-813.

Han, J., Zhu, Y., Chen, X., Liao, B., Yao, H. and Song, J., et al. (2013). The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. Bio Med Res. Internal. Doi: 10.1155/2013/741476

Higgs, P. G. and Attwood, T. K. (2005). Bioinformatics and molecular evolution. Blackwell Publishing, UK. 158-194.

Hřibová, E., Čížková, J., Christelová, P., Taudien, S., de Langhe, E. and Doležel, J. (2011). The ITS1-5.8S-ITS2 sequence region in the *Musaceae*: Structure, diversity and use in molecular phylogeny. PLos ONE. 6(3):e17863.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M. and Sturrock, S., et al. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12):1647-1649.

Kress, W. J. and Specht, C. D. (2005). Between cancer and capricon: Phylogeny, evolution and ecology of the primarily tropical Zingiberales. Biologiske Skrifter. 55:459-478.

Lamare, A., Otaghvari. A. M. and Rao, S. R. (2017). Phylogenetic implications of the internal transcribed spacers of nrDNA and chloroplast DNA fragments of *Musa* in deciphering the ambiguities related to the sectional classification of the genus. Genet. Res. Crop Evo. 64(6):1241-1251.

Li, L. F., Häkkinen, M., Yuan, Y. M., Hao, G. and Ge, X. J. (2010). Molecular phylogeny and systematics of the banana family (*Musaceae*) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. Mol. Phylogenet. Evol. 57:1–10.

Liu, A. Z., Kress, W. J. and Li, D. Z. (2010). Phylogenetic analyses of the banana family (*Musaceae*) based on nuclear ribosomal (ITS) and chloroplast (trnL-F) evidence. Molecular phylogeny of the banana family. Taxon. 59(1):20-28.

Molina, A. B. and Kudagamage, C. (2002). The international network for the improvement of banana and plantain (INIBAP): PGR activities in south Asia. In: South Asia Net-work on Plant Genetic Resources (SANPGR) meeting held on December 9-11 at Plant Genetic Resources Center (PGRC), Peradeniya, Sri Lanka. 2002;1-7.

Rogers, S. O., Bendich, A. J. (1994). Extraction of total cellular DNA from plants, algae and fungi. In: Gelvin S. B., Schilperoort R. A. (eds) Plant Molecular Biology Manual. Springer, Dordrecht. D1:1-8.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge,J. L. and Levesque, C. A., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. USA. 109(16): 6241-6246.

Smarda, P., Bures, P., Horova, L., Leitch, I.J., Mucina, L. and Pacini, E., et al. (2014). Ecological and evolutionary significance of genomic GC content diversity in monocots. Proc. Natl. Acad. Sci. USA. 111(39):E4096-102.
Doi: 10.1073/pnas.1321152111

Tamura, K. and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Bio. Evol. 10:512-526.

Tamura, K., Steche, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Bio. Evol. 30:2725-2729.

Tripathi, A. M., Tyagi, A., Kumar, A., Singh, A., Singh, S., Chaudhary, L. B. and Roy, S. (2013). The Internal Transcribed Spacer (ITS) region and *trnh*H-*psb*A are suitable candidate loci for DNA barcoding of tropical tree species of India. Plos ONE. 8(2):e57934.
Available:http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0057934

White, T. J., Bruns, T., Lee, S. and Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR Protocols: A guide to methods and applications. Innis, M., Gelfand, D., Sninsky, J. and White, T. (eds.). Chapter 38. Academic Press, Orlando, Florida. Wiley & Sons, New York. 315-322.

Wolfe, A. D., Datwyler, S. L. and Randle, C. P. (2002). A phylogenetic and biogeographic analysis of the *Cheloneae* (*Scrophulariaceae*) based on ITS and matK sequence data. Sys. Bot. 27:138-148.

Xing, G., Mark, P. S., Paul, P. B., Pang-Chui, S. and Rui-Jiang, W. (2011). Application of DNA barcodes in *Hedyotis* L. (*Spermacoceae*, *Rubiaceae*). J. Sys. Evol. 49:203–212.

Yao, H., Song, J., Liu, C., Luo, K., Han, J. and Li, Y., et al. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals. Plos ONE. 5(10):e13102.
Available:http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013102