Molecular identification and evolutionary relationships between the subspecies of *Musa* by DNA barcodes

Dhivya Selvaraj
Bharathiar University School of Biotechnology and Genetic Engineering
Ashutosh I
Instituto Tecnologico y de Estudios Superiores de Monterrey Campus Ciudad Juarez
Gowtham I
BU-DRDO Centre for Life Sciences
Baskar V
Bharathiar University School of Biotechnology and Genetic Engineering
Baala harini A
Bharathiar University
Mukunthakumar S
Jawaharlal Nehru Tropical Botanic Garden and Research Institute
Sathishkumar R (rsathish@buc.edu.in)
Bharathiar University

Research article

Keywords: Musa sp, Cultivar, Wild species, DNA barcode, Nucleotide diversity, Genetic relationship

Posted Date: February 8th, 2020

DOI: https://doi.org/10.21203/rs.2.22914/v1

Version of Record: A version of this preprint was published at BMC Genomics on September 24th, 2020.
See the published version at https://doi.org/10.1186/s12864-020-07036-5.

Abstract

Background: The banana (Musa sp., AAA) genome is constantly increasing due to high frequency somaclonal variations. Due to its large diversity, a conventional numerical and morphological based taxonomic identification of banana cultivars is laborious, difficult, and often is a subject of disagreements.

Results: In study, we used ITS2 region to identify and determine the genetic relationship between the cultivars and varieties of banana. Herein, a total of 16 banana cultivars were PCR amplified using ITS2 region. In addition, 321 sequences were retrieved from GenBank, USA, and used for the study. The sequences were aligned using Clustal W and genetic distances computed using MEGA V5.1. The significant divergence between the intra- and -specific genetic distances of ITS2 region was observed and the presence of a DNA barcoding gap was obvious. Based on BLAST1 and Distance methods, the results proved that ITS2 region can be successfully identified and distinguished for the cultivar and varieties of banana. Secondly, in this study, ITS2 revealed the relations between the cultivar and varieties of banana.

Conclusion: Thus, from this study, it is clear that ITS2 is not only an efficient DNA barcode to identify the banana species but also a potential candidate to study phylogenetic relationships between subspecies and cultivars. This is the first comprehensive study to categorically distinguish the economically important banana sub-species and varieties using DNA barcodes and to understand its evolutionary relationship

Background

Banana and plantain belong to the family Musaceae and are cultivated throughout tropical and subtropical regions of the world [1]. This is an important food crop next to rice, wheat and corn [2]. The edible Musa species and their hybrids and polyploids originated from two main wild species of Musa acuminata Colla and M. balbisiana Colla, with A and B genomes respectively [3]. The major cultivars belong to the sub groups of Cavendish (AAA), Lujugira (AAA), Figuepomme (AAB), Plantain (AAB), Saba Bluggoe (ABB) and Sucier (AA) [4]. It is a staple edible fruit crop with a good source of potassium and magnesium, which provides health benefits like maintaining the normal blood pressure and protects the heart [5]. The genome is continuously expanding due to high frequency of somaclonal variation, increasing diversity, leading to often subject of disagreements [6].

There have been wide discussions concerning the identities of the progenitors of domesticated banana. M. acuminata and M. balbisiana have been proposed as wild parents of modern banana. There are four wild species, which includes M. Acuminate (donor of A genome), M. balbisiana (donor of B genome), M. schizocarpa (donor of S genome) and M.texilis (donor of T genome) have contributed to the gene pools of bananas. These wild species are extensively distributed in the subtropical and tropical regions of Asia. There are nine subspecies like banksii, burmannica, burmannicoides, errans, malaccenis, microcarpa, siamea, truncataandzebrine are identified under M. acuminata.M. balbisiana exhibits wide morphological

characters. Though the species M.schizocarpa and M.texilis were endemic to Papua New Guinea, which do not show apparent morphological diversification [7].The species Musa nagensium remained unnoticed by botanists and no collections were made more than a century from North-East India. [8] Misidentified M. cheesmanii N.W. simmonds as M. nagensium and provided the photograph of the former. It was rediscovered and detailed description was explained by both [9] and [10]. Similarly, there was a misidentification on AAA genome, which showed similarity to ABB genome [11]. The genetic diversity was observed in the pool of M. acuminata species by using primers of highly repetitive sequence and tandem repeats by [12] [13]. A total number of cultivars in bananas and plantains were estimated to be around 300–1000 and their names are highly confusing, even within a country [14]. Cultivated bananas differ from their relative wild species and multiplied through vegetative propagation, exhibiting high level of morphological diversification. [7] Classification found to be clear, coherent and accepted widely. Though the cultivated banana has socioeconomic importance, where the genetic studies are limited due to continuation of polyploidy, parthenocarpy and complexity in sample collection. Moreover, the correct identification of Musa cultivar is crucial for utilization and also importantly for genetic resources conservation. Traditional methods to identify Musa cultivars relies more on morphological characters [15] but that are often affected by environmental and developmental factors. Phenotypic classification and their genetic relationship among the genotypes are still under debate [16]. Even molecular markers like RAPD [17] did not have a sufficient discriminating power to classify the nine genotypes of Musa [18]. Genomic in situ hybridisation (GISH) are not found to be suitable for high-throughput screening to large breeding populations [19]. But DNA markers were used for identifying dwarf Cavendish derived from micropropagation [20], finger printing can detect the parental genotypes with progeny populations during hybridisation [21]. Plastid subtype identity (PS-ID) through sequence analysis was carried out to show the possible maternal relationship among Musa sp., [22]. PCR-RFLP markers of ribosomal internal transcribed spacers (ITS) were used to determine the Musa genome and hybrids at nursery stage [23]. Therefore, a simple and accurate identification method is very much required for determining the genetic variation between the cultivars of Musa species.

DNA barcoding is a recent technique that uses short and standardised DNA fragment to discriminate the specimens at species level [24] [25] [26]. Many disputed species have been correctly identified [27]. Herbal products have also been authenticated through DNA barcodes [28] [29]. Also some studies reported that DNA barcodes are used even for identifying the herbarium samples [30], intra specific, ecotypes [31] and ornamental species for horticultural industries [32]. Hence, DNA barcoding became an efficient tool for identification with discriminating power at species level [25]. The chloroplasts DNA sequences like matK, rbcL, psbA-trnH, atpF-atpH and internal transcribes spacer (ITS) region of nuclear ribosomal DNA have been proposed as potential plant barcodes [33]. The internal transcribed spacer 2 (ITS2) is located between the ribosomal 5.8S and 28S, which is actively involved in the regulation of the transcription of active ribosomal subunits and it is essential for pre-rRNA processing [34]. Using the conserved regions, it is easy to design a universal primer, PCR amplify and DNA sequencing of amplicons will reveal the variability that can be used to distinguish the closely related species. Due to this universality, currently ITS2 was considered as standard barcode for authenticating the medicinal plants [35] [36] [37]. Moreover,

recently it also clearly identified the different varieties, imported teas [38] and small millet land races [39]. In this context, DNA barcoding analysis was performed for the banana cultivars and wild Musa accessions using internal transcribed spacer region ITS2 for better understanding of the origin and domestication of cultivated banana and to clear the confusions in varietal synonyms.

Results

PCR success rate and DNA sequencing

The amplification and sequence success rate of the ITS2 sequences from sampled specimens of Musa sp., was found to be 100%. The lengths of the ITS2 sequences used for the analyses were in the range of 325-375 bp, with an average of 345 bp. The mean GC content was 60.3%, with a range of 58.3-69%.

Genetic diversity

Genetic divergences were estimated using six metrics like average inter-specific distance, the minimum inter-specific distance, theta prime, average intra-specific distance, coalescent depth and theta. The region *ITS2* exhibits significant divergences at the inter-species level (Table 2) at the level of cultivars and varieties level. At the intra-specific level, relatively lower divergences were observed for all the corresponding metrics.

Assessment of barcoding gap

Interspecific versus intraspecific divergence were analysed by examining the distribution of genetic distance at a scale of 0.008 distance units. Only a slight overlap in inter and intraspecific variation was observed (Fig 1). The interspecific distance was in the range of 0.002 – 0.184 equaled 0.002 for only 0.26% and the proportion of inter-specific genetic distance < 0.135 was about 8.33%. The intraspecific distance ranged from 0.000 to 0.135, and most *Musa* species with more than two samples in our study had a unique sequence (58.93%) in the *ITS2* region. The results indicate that evident barcoding gap between inter and intra specific divergence, thus*ITS2* provides a useful region to authenticate different *Musa* species.

Efficacy of ITS2 for authentication

ITS2 showed 97.7% and 95.8% identification success rates at the species level for 345 samples of *Musa* using BLAST1 and nearest genetic distance respectively. Nearly 15 cultivar and wild species were identified that are shown in the Table 3. Thus, ITS2 region exhibited high identification efficiency.

Sequence analysis and species discrimination

ITS2 sequences were collected and evaluated using MEGA (Fig 2). As a result, over 95.6% of species had larger inter-than intra-specific diversity; therefore, there were relatively clear species boundaries for ITS2 sequences. Only two species were exceptions *M.schizocarpa* and *M. acuminata* x *M.textilis* had very less

variability of about 0.035%. ITS2 region shows higherpolymorphic sites representing higher genetic diversity in between sup species and cultivars of *Musa.* Unique haplotypes of *Musa* species and sub species were identified by using restriction enzymes like *MSeI, pstI* and *AvaII* respectively shown in the Table 5.

Nucleotide polymorphism and neutrality tests

DNA polymorphism analyses shows rich genomic variations in *Musa* accessions, with the total number of polymorphic sites being 112 in cultivated bananas in A genome and 33 in *B genome*. Nucleotide diversity (π and θ) for all cultivated and wild *Musa* accessions were estimated for silent, non-synonymous and total sites independently. Summaries of nucleotide diversity data for two *ITS2* regions are given in Table 4. Reduced levels of polymorphism emerged as a general property of cultivated bananas relative to their wild progenitors. It represents that sub*species* has slightly higher nucleotide diversity than wild and cultivated species. Thus, these findings suggest that the cultivars would not have undergone any severe genetic bottleneck during the initial domestication process. The triploid genome AAA and AAB groups also hold high levels of nucleotide diversity, representing the historical population sizes are large. The ABB genome of cultivated banana shows higher nucleotide diversity than that of*M. balbisiana* (Table 4). We found that nucleotide diversity at non-synonymous sites *ITS2* region was reduced in the Agenome of wild species represented as ps shown in Table 4. No polymorphic sites were observed within the cultivar and subspecies. However, it was found that the genetic diversity of AAA *genome* was 4-6folds higher than A genome cultivars. Additionally, the patterns of nucleotide variations in *ITS2 region* were examined for deviation from neutral equilibrium evolution using the Tajima's neutrality (D) test. Thus no significant departure from the neutral model was observed.

Phylogenetic analysis

The morphological classification of *Musa* species is based on [11] and [41]. To analyse the phylogenetic relationship of *Musa* cultivar with wild species nearly 103 species were studied using Neighbour Joining (NJ) method shown in Sup Fig 1. *Musa* species for this study includes 60 cultivars, 5 wild species and 9 subspecies shown in the Table 3. Among 103 sequences, 31 species where taken as representative for the comparative analysis for cultivar and wild samples from the laboratory source with subspecies and hybrids from the GenBank (Fig 3). The phylogenetic tree (Fig 3) consists of three main clades A, B and C. In clade A, cultivars red banana and robusta found to be evolutionary related with wild species of *M.balbisiana*, *M. textilis* and *M. schizocarpa* respectively and it is highly closer with subspecies *M acuminata subsp. truncata*. The clade B consists of 5 cultivars *Pisang linin*, *M. acuminata var. flava isolate,*Chemmatti, Grand naine, Nadan found to be closer related with subspecies *M. acuminata subsp. microcarpa* and *M. Banksii* respectively and clade C consist of 4 cultivars *Njalipoovan, Matti* and *Kunnan* are closely related with samples wild species of *M. acuminata, M. balbisiana* and *M. accuminatavar. Burmanicoides* respectively.

From (Fig 4), the Clade I was the most complex with 50 cultivars belonging to the wild species *Musa balbisiana*. Six cultivar species *(M.violascens, M. splendida, M. hirta, M. campestris, M. gracilis, M.salaccensis*) were grouped on same taxon I. The cultivar *M. beccarii* and its variety were lying on taxon II. The cultivars *M.peekeii* and *M. ingens* are distinct from the cultivar*M.maclayii,* which shares equal taxon III. The variety *M.monticola* and *M. barioensis* lie on the same taxon IV. The four cultivar varieties, *M. textilis, M. jackeyi, M. peekeli, M. troglodytanum* were grouped under taxon V and the varieties *M. coccinea* and *M.lutea* found to be at the neighbouring section of taxon V. The cultivar/varieties *M. rosea, M. serpentine, M.rubra, M.lateritia, M.textilis* shares same allelic profiles with the wild varieties of *M. balbisiana*and*M. accuminata*. The cultivar/varieties like *M. zaifei, M.siamensis* and also varieties obtained by us, like *Robusta, Red banana, cultivar Pisang linin, Nadan, Chemmatii, Grandnine* shares allelic profile with wild species *M.accuminata*. The seven cultivars *M.manii, M.ornata, M.yunnanensis, M. tonkinensis, M. itinerans, M. formosana, M. Viridis*grouped under the same section and shared the closer relationship with the wild species *M.balbisana*. The single cultivar species *M.nagensium* separated from the wild species of *M.balbisana.* Clade II consist of 5 taxon representing the four cultivar varieties like *Njalipoovan, Matti, Kannan, Wild M.accuminata Var burmanicoides AA,* which shares genetic relationship with wild species of *M.accuminata*. Clade III consist single taxon *M.basjoo,* which is distinct from wild species of *M.accuminata* and *M. balbisiana* as shown in Fig 4.

Data analysis using restriction enzymes

The restriction enzymes *MSel* and *AvaII* provide the best discriminatory power to differentiate the haploids of *Musa* species using ITS2 sequences. *MSel* shows single restriction sites for 11 genome of wild and cultivar species and 3 shows two restriction sites at different sites and *AvaII* one, two and three restrictions respectively.

Discussion

A common problem for banana researchers and horticulturist in south East Asia is the presence of numerous cultivar names and synonyms in different languages in the regions. Better knowledge on synonyms may promote banana trade and commerce. A rapid and reliable method for species and cultivar recognition is vital to certify the fruits and plantlets of Musa sp., and to preserve banana germplasm resources. To our knowledge, this is the first time that the DNA barcoding has been utilised in identification of cultivar species of Musa using large sample size. An ideal DNA barcode should include higher inter-specific but low intra-specific divergence in order to discriminate different species [26] as in many other previous studies for various cultivar species of colletotrichum isolates [61], fig [43], grapevine [44], pineapple [45]. In this study, ITS2 possess sufficient variable region between the cultivar species for determination of genetic divergence with high discriminatory ability. Morphological characters resolve M. acuminata population only at high land regions but low land shows overlapping and no distinguishing patterns were observed [47]. PCR-RFLP of ITS region using RsaI restriction endonuclease enzyme was used on 68 banana accessions, which showed consistent polymorphic banding DNA patterns between the wild and cultivars species of M. Acuminate [48]. Similarly, acceptable structure diversity and

molecular phylogeny were observed when an ITS1-5.8S-ITS2 region was used for the species of Musaceae [49]. Through this study, we propose that ITS2 can be an ideal DNA barcode candidate for Musa sp., however, it should be noted that the phylogeny of the Musaceae remains still controversial. The taxonomic position of the species M. beccarii remains uncertain and the species M. ingensis still undetermined [50]. Hence, the taxonomic assignment of cultivars of Musa and its discriminating power of ITS2 was conclusive.

The phylogenetic tree shows three clades A, B and C. Clade A consists of two clusters, the specimen B genome of M.balbsiana seems to be closer with M.acuminata subsp. Siamea, whereas sub spburmannica and burmannicoides grouped separately (Fig. 3). The cultivar red banana and robusta of AAA genome was closer with sub species M acuminata subsp. truncata by BLAST1 and using distance-based identification methods the cultivar red banana showed 99% and 99.1% as Musa acuminata subsp. malaccensis respectively.

In the clade B, the species M. textilis, M. banksii x Musa schizocarpa and M. schizocarpa are grouped under same cluster. The hybrid M. banksii x Musa schizocarpa was closely related with wild species M. schizocarpa. The cluster cultivar species Pisang linin is closely related with M. acuminata subsp. microcarpa and the sub species M. banksii. BLAST1 and Distance method, the A genome of cultivar Pisang linin shows 99% and 98.9% similarity with Musa acuminata var. zebrina respectively. The cultivar Chemmatti AA and Grand nine found to be closely related and identified as Musa campestris and Musa acuminata with similarity of 97% and 94% respectively. The cultivar nadan shows 96% similarity with subspecies M.banksii. In the clade C the wild species M. balbisiana, M.acuminata, the subspecies M. accuminatavar. burmanicoides AA and the cultivar species Njalipoovan1 and Matti of A genome grouped under the same cluster. The cultivar Njalipoovan1 and Matti was identified as Musa acuminata subsp. Malaccensis and M.acuminata var. Zebrine respectively. The species M balbisiana isolate LHP3 grouped as M balbisiana isolate LHP3. Thus, cultivar used in this study where inherited from wild species M. acuminata and its subspecies respectively.

Conclusion

In summary, our study clearly demonstrated that ITS2 is an ideal DNA barcode to identify Musa sub-species or cultivars and in reconstruction of the genus Musa phylogeny. However, more Musa species should be included in future to verify whether these findings hold good even if closely related taxa are newly included. In conclusion, DNA barcoding provided much useful genetic information about very complex Musa species, which will be very useful for germplasm management and in resource protection.

Methods

Plant Materials

From GenBank, 800 sequences were obtained, out of which707 sequences belong to 46 species and 65 sequences to wild isolates. We have sequenced 28 *Musa* samples out of which 12 species were from the regions of Western Ghats of India and were submitted to GenBank its accession number is given in the Table 1 and its images are shown in supplementary Figure 1. Nearly 46 annotated species and subspecies of GenBank sequences were used for the study are shown in supplementary Table 1.

DNA Extraction, Amplification and Sequencing

Fresh, young leaves of sampled specimens were collected and genomic DNA was isolated by following [51]. The *ITS2* region was amplified using the following pair of universal primers ITS-2F, 5'-ATGCGATACTTGGTGTGAAT-3', and ITS -3R, 5'-GACGCTCTCCAGACTACAAT-3'. Primers were synthesised by Integrated DNA Technologies, USA. PCR was carried out in 25µL volume containing 1XPCR Buffer, 2.5mM Mg^{2+}, 0.4 mM dNTPs, 0.5 µM of each primer, 1 U Taq DNA polymerase (Genei, India), and 30 ng genomic DNA template. The amplification was performed in a Gradient Master Cycler (Eppendorf, Germany) with a PCR program: 94°C for 4 min, followed by 35 cycles of 94°C for 45 s, 56°C for 45 s, 72°C for 1.5 min, and a final extension at 72°C for 10 min. The PCR products were sequenced by ABI-3130 Genetic Analyzer (Bioserve, India).

Sequence and genetic relationship analysis

The original sequences were analysed using MEGA [52] (Tamura et al. 2007). The ITS2 sequences were subjected to Hidden Markov Model analysis to remove the conserved 5.8S and 28S DNA sequences [26]. The ITS2 sequence were aligned using Clustal W [53] and the genetic distance computed using MEGA 5.1 according to the Kimura 2- parameter (K2P) model [52]. The average intra-specific distance, the minimum intra-specific distance and theta prime were used to represent inter-specific divergences using the K2P model [25] [35]. The average intra-specific distance, coalescent depth and theta were calculated to evaluate the intra-specific variation [26]. The distributions of inter- *versus* intra-specific variability were compared using the DNA barcoding gaps [54]. Wilcoxon two sample tests were performed as indicated previously [55]. Two methods for species identification including BLAST1 and the nearest distance method were used to evaluate the species authentication efficacy [56]. ITS2 sequences of *Musa* species in this study were used as query sequences. BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to search for the reference database for each query sequence. In nearest distance method, correct identification means that the hit in our database based on the smallest genetic distance is from the same species as that of query. Ambiguous identification means that several hits from our database were found to have the same smallest genetic distance to the query sequence. Incorrect identification means that the hit based on the smallest genetic is not from the expected species [35]. The discriminatory power of ITS2 sequences was calculated using MEGA.

To understand the wild parents of cultivated bananas, the Neighbour joining (NJ) method for phylogenetic inference was carried out in MEGA version 5.1 [52], using Kimura's 2- parameter distances [57] (Kimura 1980). Gaps were treated as missing data and bootstrap values for the NJ trees were

obtained from 1000 replicates. We evaluated overall nucleotide diversity and also for AAA, AAB, AA, Wild species, sub species, cultivar species respectively. Genetic analysis of sequence polymorphism was performed using MEGA. Number of segregating sites (S), number of haplotypes (H), Tajima's D was determined [58]. In addition, we surveyed nucleotide diversity (π) [59] and theta (θ) [60] for total, silent and nonsynonymous sites independently, whereas insertion/deletions (indels) were not included in this analysis.

Data analysis using restriction enzymes

ITS2 sequence data of 15 specimens were aligned and restriction patterns were predicted as shown in the Table 5 using NEBcutter. Restriction fragments were predicted and compared for choosing the best discriminatory enzymes for Haplotypes discrimination.

Abbreviations

ITS
Internal transcribed spacer
matK
maturase K
MEGA
Molecular Evolutionary Genetics Analysis
NJ
Neighbour joining
K2P
Kimura 2- parameter

Declarations

Consent for publication

Not applicable

Availability of data and materials

All the data analysed in this study are included in this article and its supplementary information files with accession numbers at NCBI database (https://www.ncbi.nlm.nih.gov/)

Acknowledgments

Authors are thankful to Bharathiar University for the research facility. SR is thankful to UGC - SAP and DST-FIST. DS is grateful to University Grant Commission (UGC) Women's Post-Doctoral Research Award (PDFWM -2015-17-TAM-34021).

Funding

This research is supported by UGC-SAP, DST-FIST and UGC Women's Post-Doctoral Research Award (PDFWM -2015-17-TAM-34021).

Author information

Affiliations

[1]Plant Genetic Engineering Laboratory, Department of Biotechnology, Bharathiar University,

Coimbatore - 641046, India.

[2]Technologico de Monterrey, Centre of Bioengineering, Epigmenio Gonzalez #500, France. San Pablo, Campus Queretaro, Queretaro, Mexico.

[3]Plant molecular farming Laboratory, DRDO-BU Centre for Life Sciences, Bharathiar University, Coimbatore, India – 641046

[4]Biotechnology and Bioinformatics Division, Jawaharlal Nehru Tropical Botanic Garden & Research Institute, Palode, Thiruvananthapuram, Kerala- 695 562, India

Contributions

DS conceived the idea and designed the experiments. GI, BV and MS collected the samples. BA and GI performed the experiments. DS analysed the data and wrote the manuscript. SR finalized the manuscript. DS, GI, BV and SR have read and approved the final manuscript.

Corresponding authors

Correspondence to R. Sathishkumar

Ethics declarations

Not applicable

Consent for publication

Not applicable

Competing Interests

The authors declare that they have no competing interest

Additional information

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

Additional file 1. Number of taxa used in this study

Additional file 2. Supplemental Fig 1: Different varieties of banana used in this study

References

1. Li L-F, Wang H-Y, Zhang C, Wang X-F, Shi F-X, Chen W-N, et al. Origins and Domestication of Cultivated Banana Inferred from Chloroplast and Nuclear Genes. PLoS One Public Library of Science; 2013;8:1–11. https://doi.org/10.1371/journal.pone.0080502
2. Perrier X, Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, et al. Multidisciplinary perspectives on banana (Musa spp.) domestication. PNAS. 2011;5.
3. Opara UL, Jacobson D, Al-Saady NA. Analysis of genetic diversity in banana cultivars (Musa cvs.) from the South of Oman using AFLP markers and classification by phylogenetic, hierarchical clustering and principal component analyses . J Zhejiang Univ Sci B Hangzhou: Zhejiang University Press; 2010;11:332–41. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2865835/
4. Aurore G, Parfait B, Fahrasmane L. Bananas, raw materials for making processed food products. Trends Food Sci Technol. 2009;20:78–91. http://www.sciencedirect.com/science/article/pii/S0924224408002483
5. Sampath KKP., Bhowmik D., Umadevi S, Duraivel S. Traditional and Medicinal Uses of Banana. J Pharmacogn Phytochem. 2012;1:51–63. http://www.phytojournal.com/vol1Issue3/Issue_sept_2012/9.1.pdf
6. Ploetz RC. Fusarium Wilt of Banana. Phytopathology 2015;105:1512–21. http://apsjournals.apsnet.org/doi/10.1094/PHYTO-04-15-0101-RVW
7. Simmonds NW, Weatherup STC. Numerical taxonomy of the wild bananas (Musa). New Phytol. 1990;115:567–71.
8. Uma S, Sathiamoorthy S, Durai S. Banana Indian Genetic Resource and Catalogue. India. National Research Centre for Banana (NRCB), Tiruchirapalli; 2005.

9. Gogoi R. Musa nagensium var. Hongii Hkkinen-a new addition to the Flora of India. Taiwania. 2013;58:49–52.
10. Joe A, Sreejith PE, Ashfak OA, Sabu M. Regarding the identity, rediscovery and taxonomic history of Musa nagensium (Musaceae) from India. Rheedea. 2014;24:5–11.
11. Davey MW, Gudimella R, Harikrishna JA, Sin LW, Khalid N, Keulemans J. “A draft Musa balbisiana genome sequence for molecular genetics in polyploid, inter- and intra-specific Musa hybrids”. BMC Genomics. England; 2013;14:683.
12. Jarret RL, Vuylsteke DR, Gawel NJ, Pimentel RB, Dunbar LJ. Detecting genetic diversity in diploid bananas using PCR and primers from a highly repetitive DNA sequence. Euphytica. 1993;68:69–76. https://doi.org/10.1007/BF00024156
13. Lebot V, Aradhya KM, Manshardt R, Meilleur B. Genetic relationships among cultivated bananas and plantains from Asia and the Pacific. Euphytica 1993;67:163–75. https://doi.org/10.1007/BF00040618
14. Ploetz RC. Assessing threats posed by Destructive Banana Pathogens. Acta Hortic International Society for Horticultural Science (ISHS), Leuven, Belgium; 2009. p. 245–52. https://doi.org/10.17660/ActaHortic.2009.828.25
15. Simmonds NW, Weatherup, S.T.C.1990. Numerical taxonomy of the wild bananas (Musa). *New Phytologist* 115:567–571
16. Gawel NJ, Jarret RL, Whittemore AP. Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of Musa. Theor Appl Genet 1992;84:286–90. https://doi.org/10.1007/BF00229484
17. Venkatachalam L, Sreedhar R V, Bhagyalakshmi N. Genetic analyses of micropropagated and regenerated plantlets of banana as assessed by RAPD and ISSR markers. Vitr Cell {&} Dev Biol - Plant. 2007;43:267–74. https://doi.org/10.1007/s11627-007-9028-7
18. Fauré SJL Noyer J-PHorry F, Bakry C, Gonzàlez-de-León D, Lanaud. A molecular marker based linkage map of diploid bananas (Musa acuminata). Theor Appl Genet. 1993;87. https://doi.org/10.1007/BF00215098
19. D’Hont A, Paget-Goya A, Escoute J, Carreel F. The interspecific genome structure of cultivated banana, Musa spp. revealed by genomic DNA in situ hybridization. Theor Appl Genet. 2000;100. https://doi.org/10.1007/s001220050024
20. Damasco OP, Graham GC, Henry RJ, Adkins SW, Smiths MK, Godwin ID. Random amplified polymorphic DNA (RAPD) detection of dwarf off-types in micropropagated Cavendish (Musa spp. AAA) bananas. Plant Cell Rep. 1996;16:118–23. https://doi.org/10.1007/BF01275464
21. Crouch JH, Crouch HK, Tenkouano A, Ortiz R. VNTR-based diversity analysis of 2x and 4x full-sib Musa hybrids. Electron J Biotechnol. 1999;2. https://doi.org/10.2225/vol2-issue3-fulltext-1
22. Umali RP, Kameya N, Nakamura I. Development of PCR-based fingerprinting tool in banana (Musa sp., AAA) and conversion of negative to positive DNA marker. HortScience. 2002;37:1108–11.

23. Nwakanma DC, Pillay M, Okoli BE. PCR-RFLP of the ribosomal DNA internal transcribed spacers (ITS) provides markers for the A and B genomes in Musa L. Theor Appl Genet. 2003;108. https://doi.org/10.1007/s00122-003-1402-1
24. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proceedings Biol Sci. England; 2003;270:313–21.
25. Chen S, Yao H, Han J, Liu C, Song J, Shi L, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One. United States; 2010;5:e8613.
26. Feng S, Jiang M, Shi Y, Jiao K, Shen C, Lu J, et al. Application of the Ribosomal DNA ITS2 Region of Physalis (Solanaceae): DNA Barcoding and Phylogenetic Study. Front Plant Sci Frontiers Media S.A.; 2016;7:1047. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4949264/
27. Gutteridge A, Burns M. The Application of DNA Molecular Approaches for the Identification of Herbal Medicinal Products. J Assoc Public Anal. 2013;41:53–66.
28. Han J, Pang X, Liao B, Yao H, Song J, Chen S. An authenticity survey of herbal medicines from markets in China using DNA barcoding. Sci Rep Nature Publishing Group; 2016;6:18723. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4703975/
29. Shanmughanandhan D, Ragupathy S, Newmaster SG, Mohanasundaram S, Sathishkumar R. Estimating Herbal Product Authentication and Adulteration in India Using a Vouchered, DNA-Based Biological Reference Material Library. Drug Saf. New Zealand; 2016;39:1211–27.
30. Hartvig I, Czako M, Kjær ED, Nielsen LR, Theilade I. The Use of DNA Barcoding in Identification and Conservation of Rosewood (Dalbergia spp.). PLoS One. Public Library of Science; 2015;10:1–24. https://doi.org/10.1371/journal.pone.0138231
31. Gismondi A, Leonardi D, Enei F, Canini A. Identification of Plant Remains in Underwater Archaeological Areas by Morphological Analysis and DNA Barcoding. Adv Anthropol 2013;03:240–8. http://www.scirp.org/journal/PaperDownload.aspx?DOI=10.4236/aa.2013.34034
32. Elansary H, Ashfaq M, Ali HM, Yessoufou K. The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry. PLoS One Public Library of Science; 2017;12:1–15. https://doi.org/10.1371/journal.pone.0172170
33. 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. United States; 2008;105:2923–8.
34. Jørgensen A, Stothard JR, Madsen H, Nalugwa A, Nyakaana S, Rollinson D. The ITS2 of the genus Bulinus: novel secondary structure among freshwater snails and potential new taxonomic markers. Acta Trop 2013;128:218—225. https://doi.org/10.1016/j.actatropica.2012.05.009
35. Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, et al. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. J Ethnopharmacol. Ireland; 2010;130:116–21.
36. Yao H, Song J, Liu C, Luo K, Han J, Li Y, et al. Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. PLoS One, Public Library of Science; 2010;5:1–9. https://doi.org/10.1371/journal.pone.0013102

37. Gu W, Song J, Cao Y, Sun Q, Yao H, Wu Q, et al. Application of the ITS2 Region for Barcoding Medicinal Plants of Selaginellaceae in Pteridophyta. PLoS One, Public Library of Science; 2013;8:1–8. https://doi.org/10.1371/journal.pone.0067818
38. Lee S-C, Wang C-H, Yen C-E, Chang C. DNA barcode and identification of the varieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences. J food drug Anal. China (Republic : 1949-); 2017;25:260–74.
39. Ragupathy S, Dhivya S, Patel K, Sritharan A, Sambandan K, Gartaula H, et al. DNA record of some traditional small millet landraces in India and Nepal. 3 Biotech Berlin/Heidelberg: Springer Berlin Heidelberg; 2016;6:133. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4903100/
40. De Jesus ON, Silva S de O e., Amorim EP, Ferreira CF, de Campos JMS, Silva G de G, et al. Genetic diversity and population structure of Musa accessions in ex situconservation. BMC Plant Biol 2013;13:41. https://doi.org/10.1186/1471-2229-13-41
41. Bhat K V, Jarret RL, Rana RS. DNA profiling of banana and plantain cultivars using random amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP) markers. Electrophoresis. Germany; 1995;16:1736–45.
42. Feng S, Jiang Y, Wang S, Jiang M, Chen Z, Ying Q, et al. Molecular Identification of Dendrobium Species (Orchidaceae) Based on the DNA Barcode ITS2 Region and Its Application for Phylogenetic Study. Iriti M, editor. Int J Mol Sci MDPI; 2015;16:21975–88. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4613292/
43. Castro C, Hernandez A, Alvarado L, Flores D, A ADCRS, Rica C, et al. DNA Barcodes in Fig Cultivars (Ficus carica L .) Using ITS Regions of Ribosomal DNA , the psbA-trnH Spacer and the matK Coding Sequence. 2015;95–102.
44. Galbacs Z, Molnar S, Halasz G, Kozma P, Hoffmann S, Kovacs L, et al. Identification of grapevine cultivars using microsatelite-based DNA barcodes. Vitis. 2009;48:17–24.
45. Hidayat T, Abdullah FI, Kuppusamy C, Samad AA, Wagiran A. Molecular Identification of Malaysian Pineapple Cultivar based on Internal Transcribed Spacer Region. APCBEE Procedia 2012;4:146–51. http://www.sciencedirect.com/science/article/pii/S2212670812001959
46. Wu B, Zhong G, Yue J, Yang R, Li C, Li Y, et al. Identification of Pummelo Cultivars by Using a Panel of 25 Selected SNPs and 12 DNA Segments. PLoS One. Public Library of Science; 2014;9:1–12. https://doi.org/10.1371/journal.pone.0094506
47. Javed MA, Chai M a K, Othman RY. Morphological characterization of malaysian wild banana. Biotropia (Bogor). 2002;1:21–37.
48. Hapsari L, Lestari DA. Fruit characteristic and nutrient values of four Indonesian banana cultivars (Musa spp.) at different genomic groups. Agrivita. 2016;38:303–11.
49. Hřibová E, Čížková J, Christelová P, Taudien S, de Langhe E, Doležel J. The ITS1-5.8S-ITS2 Sequence Region in the Musaceae: Structure, Diversity and Use in Molecular Phylogeny. PLoS One Public Library of Science; 2011;6:1–11. https://doi.org/10.1371/journal.pone.0017863
50. Simmonds NW. The evolution of the bananas. London: Longmans Green; 1962.

51. Khanuja SPS, Shasany AK, Kumar S. Rapid Isolation of DNA from and Fresh samples of Plants Producing large amounts of secondary Metabolites and Essential Oils. Plant Mol Biol Report 1999;17:1–7. http://link.springer.com/10.1023/A:1007528101452
52. Tamura K, Dudley J, Nei M, Kumar S. Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24. https://doi.org/10.1093/molbev/msm092
53. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinforma. United States; 2002;Chapter 2:Unit 2.3.
54. Meyer CP, Paulay G. DNA Barcoding: Error Rates Based on Comprehensive Sampling. PLOS Biol Public Library of Science; 2005;3. https://doi.org/10.1371/journal.pbio.0030422
55. Selvaraj D, Sarma RK, Shanmughanandhan D, Srinivasan R, Ramalingam S. Evaluation of DNA barcode candidates for the discrimination of the large plant family Apocynaceae. Plant Syst Evol 2015;301:1263–73. https://doi.org/10.1007/s00606-014-1149-y
56. Ross HA, Murugan S, Li WLS. Testing the reliability of genetic methods of species identification via simulation. Syst Biol. England; 2008;57:216–30.
57. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. Germany; 1980;16:111–20.
58. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. United States; 1989;123:585–95.
59. Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics. United States; 1983;105:437–60.
60. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. United States; 1975;7:256–76.
61. Afanador-Kafuri L, Minz D, Maymon M, Freeman S.Characterization of colletotrichum isolates from tamarillo, passiflora, and mango in Colombia and identification of a unique species from the genus. Phytopathol,2003; 93:579–587

Tables

1. Number of chromosomes and ploidy levels of *Musa species* /Cultivar names and its GenBank Accessions

S.NO	Accession Number	Species/ Cultivar Names	Number of Chromosomes	Ploidy	Cultivated Clones/ Subspecies
1	KY710751	cultivar Nadan	33	3n=3x	Cultivated clones
2	KY710752	*Musa acuminata subsp. burmannicoides*	22	2n=2x	Sub Species
3	KY710753	*Musa acuminata*	22	2n=2x	Wild Species
4	KY710754	*cultivar Njalipoovan*	33	3n=3x	Cultivated Clones
5	KY710755	*Musa balbisiana*	22	2n=2x	Wild Species
6	KY710756	*Musa acuminata*	22	2n=2x	Wild Species
7	KY710757	*cultivar Pisang*	33	3n=3x	Cultivated Clones
8	KY710758	*cultivar Chemmatti*	33	3n=3x	Cultivated Clones
9	KY710759	*Musa acuminata*	33	3n=3x	Wild Species
10	KY710760	*cultivar Kannan*	33	3n=3x	Cultivated Clones
11	KY710761	*cultivar Njalipoovan*	33	3n=3x	Cultivated Clones
12	KY710762	*cultivar Grande Nain*	33	3n=3x	Cultivated Clones
13	KY710763	*cultivar Matti*	33	3n=3x	Cultivated Clones
14	KY710764	*Red Dacca (Red Banana)*	33	3n=3x	Cultivated Clones
15	KY710765	*Robusta*	33	3n=3x	Cultivated Clones

Table 2. Genetic distance determination using *ITS2* region for 277 samples of 47 *Musa* cultivars

Measurement	Kimura 2- parameters (K2P) Value
All inter-specific distance	0.194 ± 0.076
Theta prime	0.187 ± 0.064
The minimum inter-specific distances	0.122±0.046
All intra-specific distances	0.035±0.006
Theta	0.081± 0.004
Coalescent depth	0.070±0.018

Table 3. Identification efficiency for *ITS2* using BLAST1 and DISTANCE methods

Sample Name	Sample Identified	Methods of Identification	Correct Identification (%)	Ambiguous Identification (%)
Robusta	*M. aurantiaca*	*BLAST1* *DISTANCE*	97 97.06	3 2.94
Nadan	*M. banksii*	*BLAST1* *DISTANCE*	96 96.04	3 3.06
Wild Musa accuminatavar.burmanicoides AA	*Musa ABB*	*BLAST1* *DISTANCE*	99 99.7	1 0.3
Wild Musa accuminata AA	*M. acuminata subsp. malaccensis*	*BLAST1* *DISTANCE*	99 99.37	1 0.63
Njalipoovan AB	*Musa acuminata subsp. malaccensis*	*BLAST1* *DISTANCE*	98 98.45	2 1.55
Wild Musa balbisiana BB	*Musa laterita*	*BLAST1* *DISTANCE*	98 98.41	2 1.59
Wild Musa balbisiana BB	*Musa balbisiana*	*BLAST1* *DISTANCE*	98 97.68	2 2.32
Wild Musa accuminata	*Musa accuminata*	*BLAST1* *DISTANCE*	96 95.81	4 4.19
Cultivar Pisang linin	*Musa acuminata var. zebrina*	*BLAST1* *DISTANCE*	99 98.9	1 1.1
Cultivar Chemmatti - AA	*Musa campestris*	*BLAST1* *DISTANCE*	97 96.7	3 3.3
Wild Musa accuminata	*Musa laterita*	*BLAST1* *DISTANCE*	95 95.15	5 4.85
Cultivar Kannan BA	*Musa acuminata subsp. malaccensis*	*BLAST1* *DISTANCE*	99 99	1 1
Grandnaine	*Musa acuminata*	*BLAST1* *DISTANCE*	94 94.11	6 5.89
Cultivar Matti AA	*Musa acuminata var. zebrina*	*BLAST1* *DISTANCE*	99 98.21	1 1.79
Red Banana	*Musa acuminata subsp. malaccensis*	*BLAST1* *DISTANCE*	99 99.1	1 0.9

Table 4. Summary of nucleotide diversity and neutrality test statistics for ITS2

Taxonomic groups	Genome	Ha	S	Ps	Θ	π	D
A genome	AAA	39	283	0.632	0.531	0.531	-0.18425
	AAB	8	75	0.115	0.076	0.074	0
	AA Wild Species	4	16	0.089	0.049	0.049	0.123
	A Sub species	13	132	0.589	0.189	0.73	-0.833
	A Cultivar	48	132	0.663	0.149	0.113	-0.845
	Over All	112	638	0.417	0.198	0.299	-0.347
B Genome	BB	6	23	0.125	0.054	0.052	-0.235
	ABB	27	233	0.373	0.125	0.120	-0.5
	Over All	33	263	0.498	0.179	0.172	-0.735

Table 5. Haplotypes discrimination of ITS2 *sequence* using Restriction Enzyme Mapping

Species Name	*MSe*I	*pst*I	*Ava*II
Robusta	363	-	293, 324
Nadan	63	-	333, 353
Wild Musa accuminatavar.burmanicoides AA	50	32	88, 119, 139
Wild Musa accuminata AA	28	-	84, 97
Njalipoovan AB	-	-	35, 66
Wild Musa balbisiana BB	5,444	-	354, 374
Wild Musa balbisiana BB	5		354, 374
Wild Musa accuminata	11	-	111
Cultivar Pisang linin	370	-	280, 300
Cultivar Chemmatti - AA	54, 411	-	321, 341, 372
Wild Musa accuminata	28	-	84, 97
Cultivar Kannan BA	148	-	186, 217, 237
Grandnaine	107, 116	-	-
Cultivar Matti AA	77	-	115, 146
Red Banana	353	-	283, 314

Figures

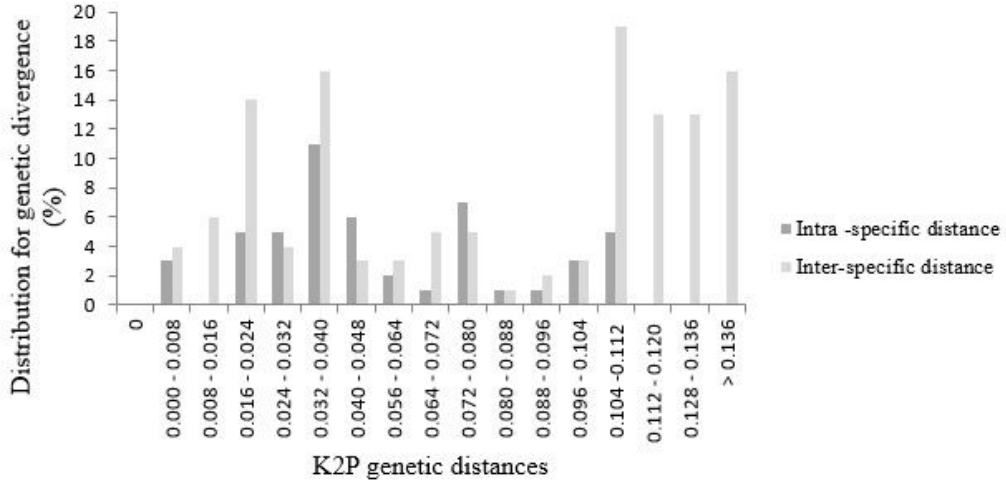

Figure 1

The Relative distribution of inter-specific divergence between congeneric musa species and intra-specific variation ($p < 0.001$)

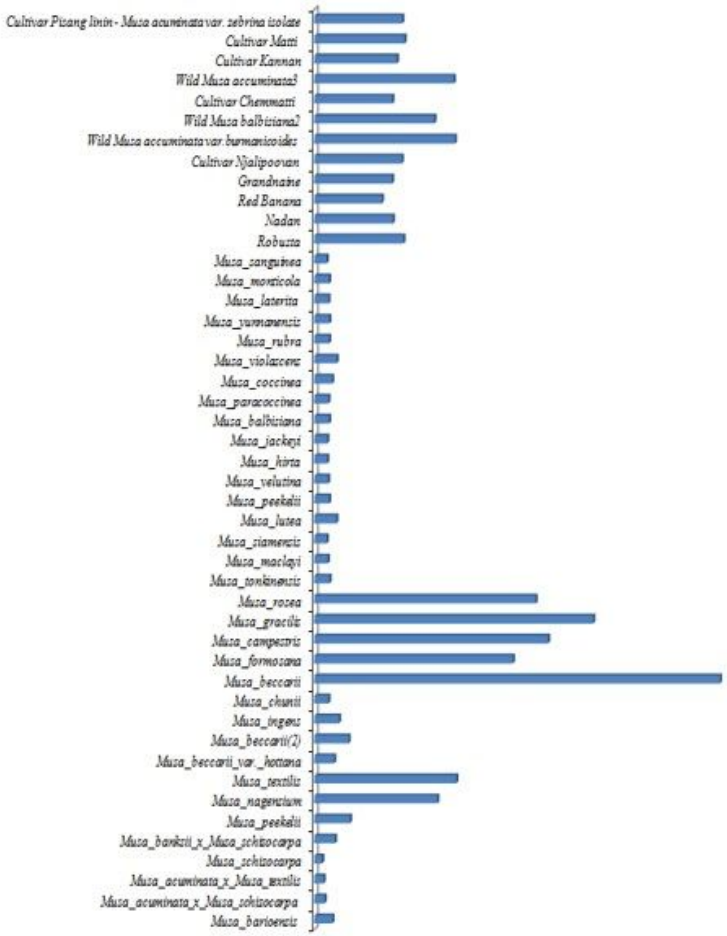

Figure 2

The heterogeneity of individual taxa by ITS2 based on 46 Musa sub-species by MEGA. The left side shows the complete list of Musa species used in this study. The right side depicts heterogeneity between and within species, where the values are calculated by using similarity matrix for biomarker ITS2 with different OTUs (Operational taxonomic Units).

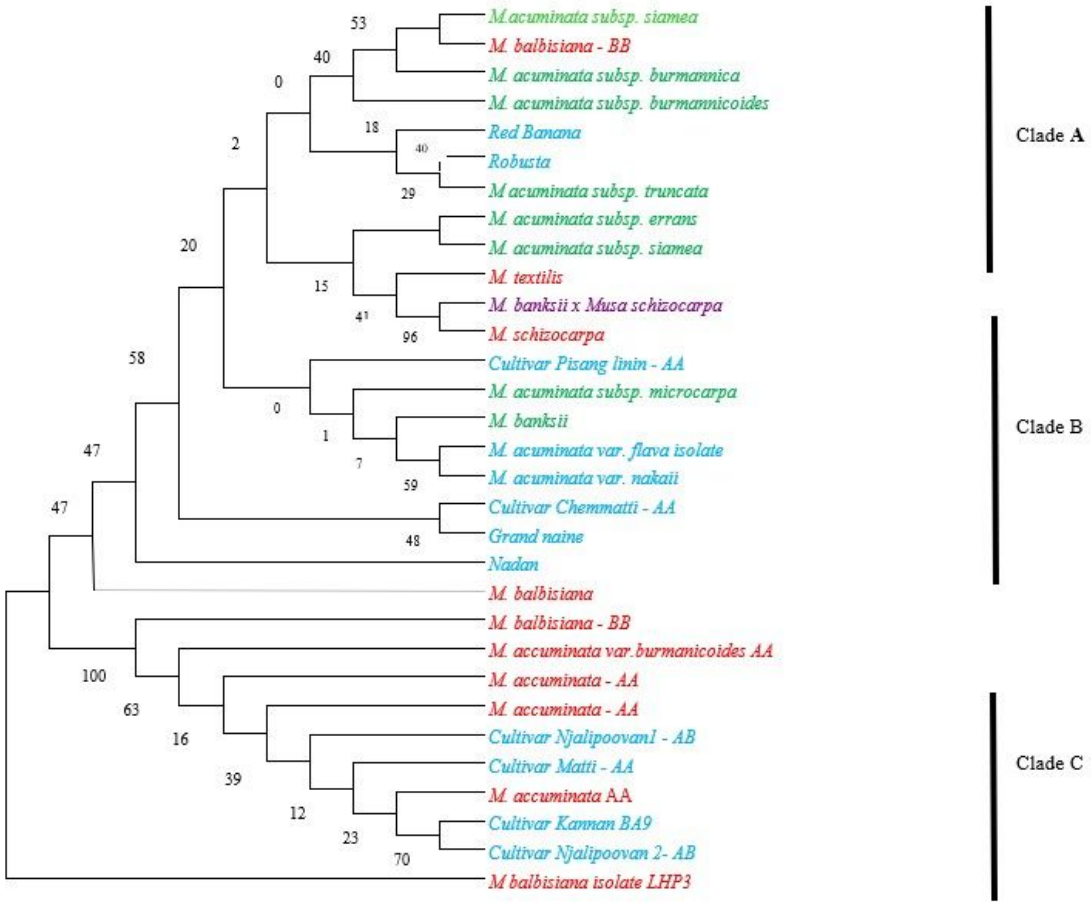

Figure 3

Neighbour Joining (NJ) tree for Musa accessions using ITS2 region. Numbers are bootstrap percentage above 50%. Numbers are bootstrap percentage above 50%. Capital letters following each accession name indicate the previously- recognized genome composition of the cultivar. The appearance of an accession more than once represents distinct sequence cloned from the same cultivar. Red indicates wild species, green indicates sub species and blue indicates cultivar

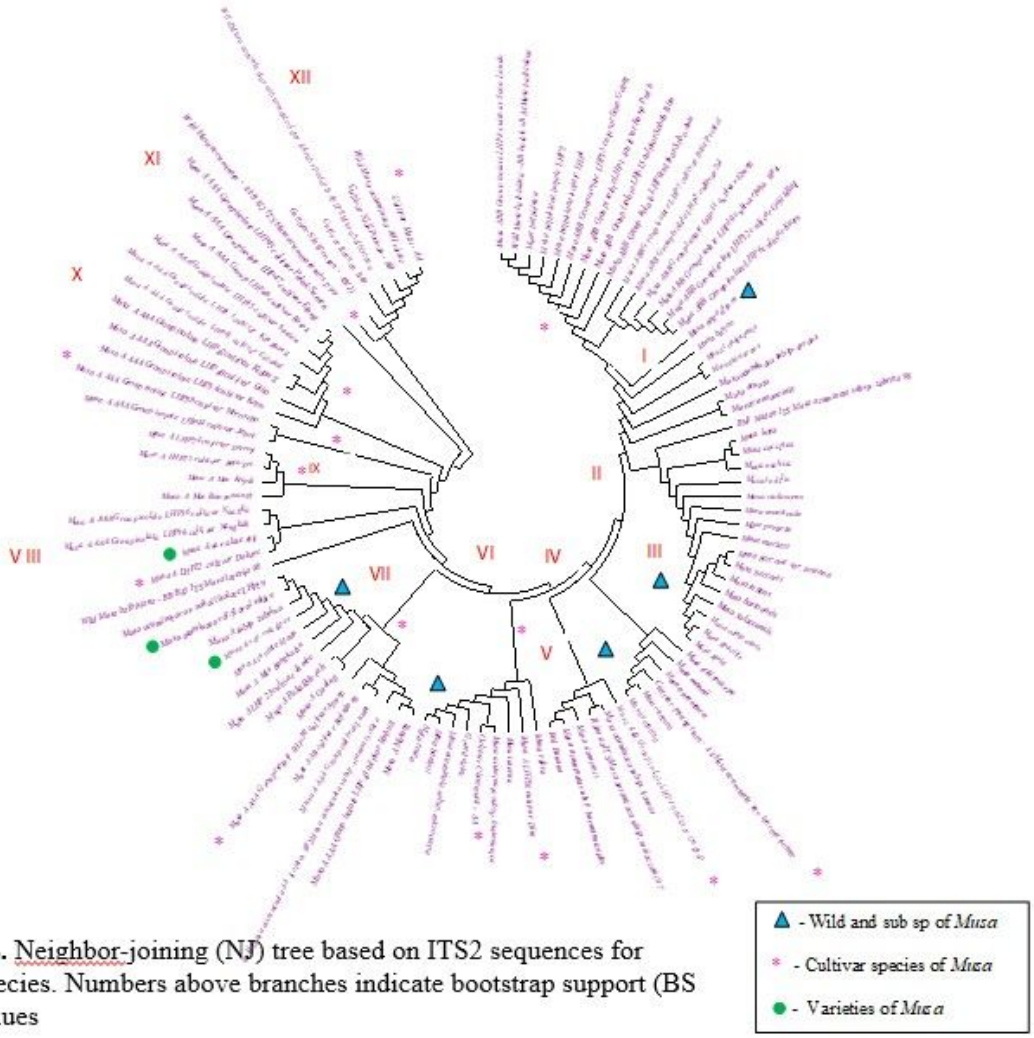

Figure 4. Neighbor-joining (NJ) tree based on ITS2 sequences for Musa species. Numbers above branches indicate bootstrap support (BS ≥ 50) values

Figure 4

Neighbor-joining (NJ) tree based on ITS2 sequences for Musa species. Numbers above branches indicate bootstrap support (BS ≥ 50) values. Legend - Wild and sub sp of Musa, * - Cultivar species of Musa, - Varieties of Musa

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable191119.docx
- S1FIG.jpg
- SupplementaryTable1Numberoftaxausedinthisstudy.docx