Eco. Env. & Cons. 26 (April Suppl. Issue) : 2020; pp. (S7-S13)
Copyright@ EM International
ISSN 0971–765X

DNA barcoding: Study of Bananas (*Musa* spp.) Wild and cultivars group from East Java inferred by *rbc* L gene sequences

Rojaunnajah Kartika Ainiyah, Verina Wahyunindita, Windi Nur Pratama, Intan Ayu Pratiwi, Edy Setiti Wida Utami and Sucipto Hariyanto*

Department of Biology, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia

(Received 27 September, 2019; Accepted 10 January, 2020)

ABSTRACT

Diversity of local banana wild and cultivars in Indonesia, especially in East Java Province has high with various local names and diverse morphological characteristics. Consortium for the Barcoding of Life (CBoL) has been recommended the *rbc*L as the plant barcode. The aims of this study were to evaluate genetic diversity and genetic relationship among *Musa* species i.e. *M. balbisiana* (BB w) i.e. Pisang Klutuk Wulung; *M. acuminata* (AA/AAA cv) i.e. Pisang Orlin and Solokoto Brenggolo; and *M. paradisiaca* (AAB and ABB cv) i.e Pisang Sewu, Susu, and Raja using *rbc*L plastid barcode loci gene sequences. Six banana specimens were collected from wild habitats and cultivations from four regencies in East Java. In this study, we amplified and sequenced the chloroplast DNA regions *(rbc*L*)* of bananas to study sequence analysis to allow a more efficient utilization among different *Musa* species. The results support used *rbc*L loci for barcoding of genus *Musa*. *rbc*L primer showed good amplification in six banana accessions with the sequence lenght in *Musa* spp. was 743 bp. However, molecular confirmation using *rbc*L sequences showed high conservation level (0.830) and low genetic variability. About 520 nucleotides were considered monomorphic (invariable) and 112 sites of DNA sequences were considered as polymorphic which comprised of 100 sites singleton variable and 12 parsimony informative sites. The results clearly showed that the use of these conserved DNA sequences as barcode primer would be an accurate way for species identification and discrimination. Reconstruction of the phylogenetic tree based on the *rbc*L gene showed that there were two clades, the first clade consisted of *M. acuminata* (w and cv), *M. velutina*, and *M. ornata* and the second clade consisted of *M. balbisiana* (w) and *M. paradisiaca* (cv). Our study provides to recommend of the *rbc*L barcode in *Musa* spp.. The finding would be applicable in conserved diversity of the banana accessions to be used as basic data for banana conservation management.

Key words : *Banana, DNA Barcoding, rbc* L*, Genetic.*

Introduction

Bananas belong to the genus *Musa* L., family Musaceae Juss., and a member of the order Zingiberales (Lamere and Rao 2015). Banana is the fourth most important crop in the developing world (Lamere and Rao 2015), with its production at a staggering 117,9 million tons in 2015 (FAO 2017). It is estimated that there are not less than 1000 banana cultivars were exist and known with various local names and have wide ranges of morpohological appearance (Hapsari *et al.,* 2015). However, the

presence of many local names and some possible synonyms among the cultivars are become problems for taxonomists and horticulturist (Hapsari *et al.,* 2017).

Simmond and Shepherd (1955) presented the theory of origin of edible bananas. Wild seeded fertile banana species *M. acuminata* Colla (donor of A genome) and *M. balbisiana* Colla (donor of B genome) are believed to be the ancestral parents of main modern banana cultivars. The vast majority of the cultivated bananas are diploids (2*n*=2×=22), triploids (2*n*=3×=33) or tetraploids (2*n*=2×=44) which have derived from inter- and intraspecific crosses involving two diploid (2n=22) wild species viz. *M. acuminata* Colla and *M. balbisiana* Colla (Simmonds and Shepherd, 1955). Cultivated bananas differ from their wild relatives in being seedless and parthenocarpic, although cultivated bananas reproduce through vegetative propagation, they exhibit a high level of morphological diversification in fruit size, shape and colour. Wild bananas species are inedible and reproduce sexually with numerous fertile seeds in the fruits, whereas sometimes reproduce via asexual budding (Li and Ge, 2017).

Analysis of genetic diversity in plant species including bananas, has an important role in the conservation of genetic resources plant. The progress of molecular biology supports various studies regarding genetic diversity in bananas as has been done by Hasan and Khasim (2018) used DNA barcoding to analyze phylogenetic relationships among 11 *Musa* varieties based on Internal regional sequences Transcribed Spacer 2 (ITS 2) which is the nuclear barcode locus and Hapsari *et al.* (2015) identified the genome and the kinship of the banana *(Musa)* from East Java assessed with the Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP) of the Internal ITS nuclear ribosomal DNA.

One of the molecular markers is DNA barcoding, involves sequences standard regions of DNA as a tool for species identification. However, not yet there is agreement on which area or locus to use for barcode on land plants (CBoL, 2009). CBoL (2009) recommended locus *rbc*L (ribulose-1,5-biphosphate carboxylase/oxygenase) for plant barcoding DNA. DNA barcode markers from chloroplast and/or nucleus genome (CBOL, 2009), etc. are necessary to conduct for a better understanding of the banana species.

The chloroplast rbcL gene encodes the large subunit of Ribulose-l,5-bisphosphate carboxylase/oxygenase (Rubisco) which basically contains all the catalytic active sites of the enzyme (Zhao *et al.,* 2017) which involved in the first major step of carbon dioxide fixation. The *rbc*L exon size approximately at full length 1,400 bp provides many characters that can be utilized in phylogenetic analysis (Hapsari *et al.,* 2019). The universal primer *rbc*L has high conservation level which valuable for studies of genetic diversity and phylogenetic (Hapsari *et al.,* 2019).

Based on this, it can be seen that little information is available and has never been reported before about chloroplast DNA (cpDNA) barcode loci such as *rbc*L in *Musa* spp. (i.e. *M. acuminata* cv. and *M. paradisiaca cv.*), although CBoL recommended locus *rbc*L as the plant barcode. The purposes of this study were to evaluate genetic diversity and genetic relationship of *M. balbisiana* (BB w), i.e. Pisang Klutuk Wulung; *M. acuminata* (AA/AAA cv), i.e. Pisang Orlin and Solokoto Brenggolo; and *M. paradisiaca* (AAB and ABB cv), i.e Pisang Sewu, Susu, and Raja using *rbc*L plastid barcode loci gene sequences.

Materials and Methods

Plant materials

Plant materials in this research was collected from four regencies in East Java, Indonesia, representing of *M. balbisiana, M. acuminata,* and *M. paradisiaca*. Descriptive information of the plant material are listed out in Table 1.

Molecular procedure

Molecular analysis was conducted at Molecular Genetics Laboratory of Biology Department, Airlangga University, Surabaya, Indonesia. The fresh young leaf sample was taken for molecular analysis. Total genomic DNAs were extracted using Promega Wizard® Genomic DNA Purification Kit (Madison, WI, USA) followed the manufacturer's protocols for plant.

Amplification of the *rbc*L marker was accomplished using primer pairs of *rbc*L -1F (5'- ATG TCA CCA CAA ACA GAA AC-3') and *rbc*L-724R (5'-TCG CAT GTA CCT GCA GTA GC-3') referred to CBoL (2009). PCR reactions were conducted in a 35 µL volume contains of 17,5 µL of Go Taq® Green Master Mix (2x) from Promega (Madison, WI, USA) (DNA polymerase, dNTPs and $MgCl_2$), 2 µL each of forward and reverse primers, 6.5 µL of Nuclease-

Table 1. List of *Musa* spp. employed in the study

Sample no.	Local name	Species	Locality/ origin	Wild (w)/ cultivar (cv)	Ploidy
S1	Klutuk Wulung	*M. balbisiana*	Pasuruan	w	BB
S2	Orlin	*M. acuminata*	Gresik	cv	AA
S3	Sewu	*M. paradisiaca*	Gresik	cv	AAB
S4	Solokoto Brenggolo	*M. acuminata*	Pacitan	cv	AA/AAA
S5	Susu	*M. paradisiaca*	Kediri	cv	ABB
S6	Raja	*M. paradisiaca*	Kediri	cv	ABB

Free Water (NFW), and 7 µL of DNA template. PCR thermal cycling program used for *rbcL* amplification consists of initial denaturation temperature at 95 °C for 5 minutes; followed by 40 cycles of denaturation for 30 seconds at 95 °C, annealing for 40 seconds at 57 ºC, and extension for 1 minute at 72 ºC. Final extension carried out for 7 minutes at 72 ºC. Amplified products were then purified and sequenced at 1st BASE Laboratories, Singapore.

Data analysis

*rbc*L sequences of additional accessions of *Musa* spp. and outgroup were retrieved from GenBank and analyzed together with our acquired sequences to maximize the species coverage for the phylogenetic analysis (Table 3). Sequence homology of the amplified sequences was detected using Basic Local Alignment Tool (BLAST). DNA sequences of *rbc*L were evaluated using BioEdit software. Multiple sequences alignments were performed using ClustalW program followed by visual adjusment (MEGA 7.0.21 software). Phylogenetic reconstructions were performed using MEGA 7.0.21 based on evolution model of Kimura 2 Parameter (K2P) using Maximum Parsimony (MP) algorithms. Bootstrap support was categorized as strong (>85 %), moderate (70-85 %), low (50-69 %), and very low (<50 %). Genetic diversity was analyzed with DnaSP ver. 6.12.03.

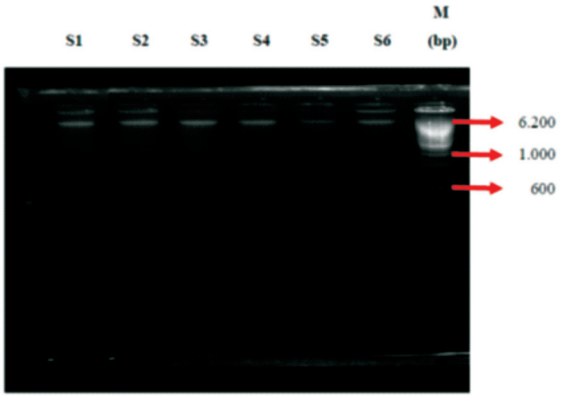

Fig. 1. Electrophoregram of whole genome DNA in *Musa* spp.

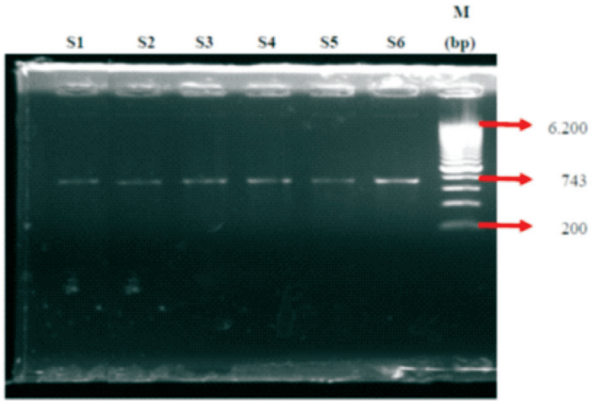

Fig. 2. Electrophoregram of PCR *rbcL* products of *Musa* spp

Results and Discussion

Genetic characterization

DNA barcoding is a technique for characterizing species of organisms using a short DNA sequence from a standard and agreed-upon position in the genome. The sequence lenght of *rbc*L in *Musa* spp. was 743 bp. Electrophoregram of whole genome DNA of *Musa* spp. showed in Figure 1 and PCR *rbcL* products showed in Figure 2.

Amplification of the *rbc*L marker was successfully carried out to six specimens examined (Table 1). Visualization on 1% agarose gel electrophoresis was shown by the presence of a specific DNA band in the sample lane at the length of approximately ±743 bp. All of the samples showed the good quality with thick DNA bands indicating that the processes of *rbc*L gene amplification of six samples were successfully done.

Success in DNA barcoding depends on choice of

a DNA region (barcode locus) with an optimum level of "universality" in terms of Polymerase Chain Reaction (PCR) amplification, sequencing, and sequence alignment, and "resolvability" with respect to the ability of its sequence to differentiate conspecifics from congenerics. In PCR based amplification, quality of DNA plays an important role (Wattoo *et al.,* 2016).

The findings of current study suggest that the use of universal primers (*rbc*L) for DNA barcoding is successful for amplification, identification and discrimination of above mentioned *Musa* spp.. The amplification success rates were 100% for all the six specimens. The *rbcL* primer pairs resulted sharp and thick bands that were required for reliable DNA sequencing. Thus, using DNA barcodes primers systems is reliable and fast system for identification of banana.

Statistical simulation of BLAST and genetic diversity

Sequence homology of the amplified sequences was detected using BLAST. Statistical simulation of BLAST sequence in *Musa* spp. showed in Table 2.

The query cover is the percentage of our sequences aligned to sequences in GenBank database. Per. Ident is the value of the percentage of identity or match between the query sequence and the database sequence. E-value is an estimated value that provides a statistically significant measure for both sequences.

The query cover with *rbc*L, sequence homology of Klutuk Wulung was 100%; Orlin was 96%; Sewu, Solokoto Brenggolo, and Susu were 99%; and Raja was 97%. Per. Ident of Klutuk Wulung was 99.60%; Orlin was 84.85%; Sewu and Susu were 99.33%; Solokoto Brenggolo was 99.46%; then Raja was 98.93%. E. Value for five samples were 0.0.

Direct sequencing on *rbc*L amplicons resulted in DNA sequences with length of 641 bp to 770 bp. The full length of *rbc*L sequences is 1,400 bp (Hapsari *et al.,* 2019); however, according to NCBI data, some *rbc*L markers available were produced shorter amplicons of 500 to 950 bp on Zingiberales, such as in *Musa* spp. 500-530 bp (Hiariej *et al.,* 2015); *Zingiber* spp. 950 bp (Vinitha *et al.,* 2014); *Heliconia* spp. 709-721 bp (Hapsari *et al.,* 2019).

The total aligned and selected *rbc*L sequences of specimens examined were 754 bp. It shows high conservation level (0.830) with 520 nucleotides were considered monomorphic (invariable) and 112 sites of DNA sequences were considered as polymorphic which comprised of 100 sites singleton variable and 12 parsimony informative sites (sites that contain at least two types of nucleotides and present at least twice). Whilst, 122 sites were considered as

Table 2. Statistical simulation of BLAST sequence in *Musa* spp.

Sample no.	Local name	Query lenght (bp)	Significant alignment with	Query cover (%)	Per. Ident (%)	E. value
S1	Klutuk Wulung	746	*M. balbisiana* (Klutuk Wulung)	100	99.60	0.0
S2	Orlin	641	*M. ornata*	96	84.85	2e-175
S3	Sewu	750	*M. balbisiana* (Klutuk Wulung)	99	99.33	0.0
S4	Solokoto Brenggolo	750	*M. ornata*	99	99.46	0.0
S5	Susu	750	*M. ornata*	99	99.33	0.0
S6	Raja	770	*M. balbisiana* (Klutuk Wulung)	97	98.93	0.0

Table 3. Species/cultivar name of *rbc*L sequences were retrieved from GenBank

No.	Species/cultivar name	Family	Bases	Accession no.
1.	*Musa balbisiana* (Klutuk Ijo)	Musaceae	1 to 676	MK238285.1
2.	*Musa acuminata* ssp. *flava*	Musaceae	1 to 676	MK238286.1
3.	*Musa ornata*	Musaceae	1 to 1464	NC_042874.1
4.	*Musa velutina*	Musaceae	1 to 1389	LT576835.1
5.	*Ravenala madagascariensis*	Strelitziaceae	1 to 676	MK238283.1
6.	*Phenakospermum guyannense*	Strelitziaceae	1 to 676	MK238284.1

alignment gaps or missing data. The 100 singleton variable site positions comprised two variants (86, 87, 149, 150, 151, 161, 162, 165, 166, 170, 182, 183, 187, 191, 208, 225, 242, 252, 272, 276, 290, 304, 317, 332, 333, 340, 341, 367, 383, 386, 387, 389, 392, 413, 419, 420, 426, 430, 437, 449, 452, 457, 458, 460, 476, 491, 510, 513, 514, 540, 547, 557, 560, 582, 583, 590, 593, 594, 595, 596, 600, 601, 604, 605, 618, 619, 637, 638, 641, 646, 651, 653, 655, 657, 658, 660, 661, 662, 666, 672, 673, 674, 676, 677, 678, 681, 682, 701, 704, 706, 707, 708, 710, 714, 715, 717, 719, 721, 722, and 723). The parsimony informative sites consisted 9 parsimony-informative sites two variants (73, 233, 287, 521, 624, 633, 654, 669, and 670) and 3 parsimony-informative sites three variants (548, 679, and 683). Parsimony informative sites are useful for reconstructing phylogenetic tree, meanwhile, singleton variation are non-informative sites that cannot provide information about which are parsimonous tree. Nonetheless, the singleton variation sequences specific to certain species may be proposed as identification barcodes (Hapsari *et al.,* 2018).

Clustering analysis result

Genetic relationship was determined by the value of similarity or distance coefficient, the higher value of similarity or fewer distance coefficients so that relationship between genome groups and/or group members are closer. *rbc*L sequences of additional accessions of *Musa* spp. and outgroup were retrieved from GenBank and analyzed together with our acquired sequences to maximize the species coverage for the phylogenetic analysis. Figure 3 represents the phylogenetic tree constructed using Maximum Parsimony (MP) algorithms.

The construction of phylogenetic tree was conducted using MEGA 7.0.21 program with Maximum Parsimony (MP) method. The construction of phylogenetic tree aims to determine the relationship of among several *Musa* spp.. In this study, the phylogenetic analysis of the ingroup species inferred by *rbc*L sequences data showed strong (94), moderate (82), and very low bootstrap (4-25). *Musa* spp. closely related to Strelitziaceae meanwhile, the separation of the outgroup Strelitziaceae was supported by high bootstraps (96) (Figure 3). The higher of bootstrap values (70-100) suggest higher confidence level of the phylogenetic trees while the lower bootstrap values have high possibility of branching rearrangement (Hapsari *et al.,* 2019).

Tree topology of MP algorithm was separated into two clades. First clade consisted of Orlin (S2), *M. velutina*, *M. ornata*, *M. acuminata* ssp. *flava*, Susu (S5), and Solokoto Brenggolo (S4). The second clade consisted of Klutuk Wulung (S1), Raja (S6), Sewu (S3), and *M.balbisiana* (Klutuk Ijo).

The closest genetic relationship were found between *M. acuminata* (AA/AAA cv) (i.e. Orlin and Solokoto Brenggolo), *M. acuminata* (AA w) (i.e. *flava*), *M. velutina*, and *M. ornata*. *M. acuminata* cv/w

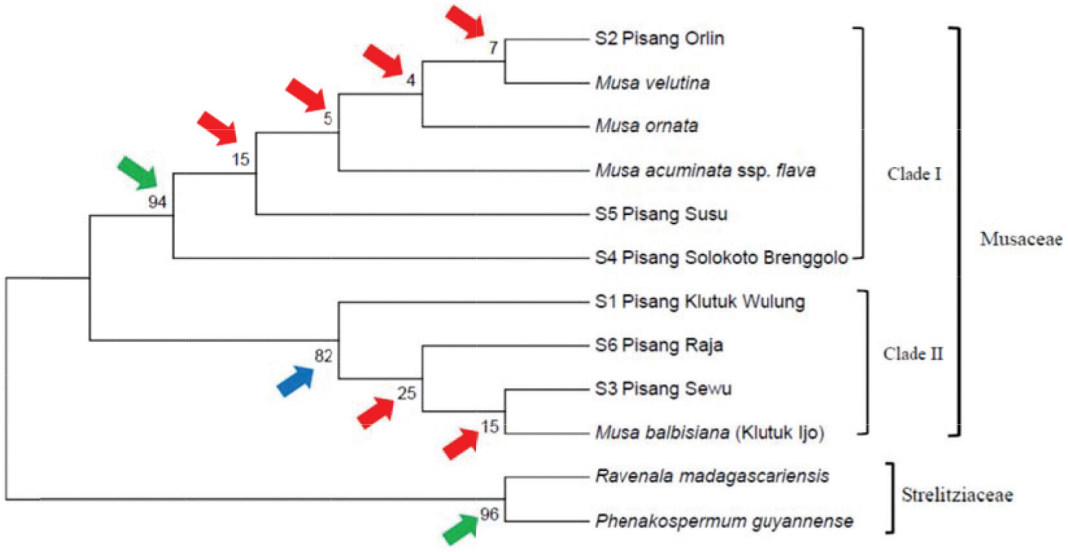

Fig. 3. The phylogenetic tree of *Musa* spp. based on rbcL gene using the Maximum Parsimony (MP) method. Notes: red arrow = branch with very low bootstrap, blue arrow = branch with moderate bootstrap, and green arrow = branch with high bootstrap.

belong to section *Eumusa* and ornamental bananas (*M. velutina* and *M. ornata*) belong to section *Rhodoclamys*. *Rhodochlamys* and *Eumusa* have basic chromosome number as $n=x=11$ (Lamere *et al.*, 2016). Previous research conducted by Lamere *et al.*, (2016) has suggested sect1ion *Rhodochlamys* and *Eumusa* are not distinct genetically two warrant resolution into two different section.

M. balbisiana (i.e. Klutuk Wulung and Klutuk Ijo) and *M. paradisiaca* (i.e. Sewu and Raja) clustered in one group. *M. paradisiaca* is the accepted name for the hybrid between *M. acuminata* and *M. balbisiana*. The results of this study are consistent that ABB-denominated cultivars have a character expression between two elders but are closer to the elders of *M. balbisiana* which donor for two genomes B.

Conclusion

The conclusion showed that, the use of *rbc*L primers for DNA barcoding is successful for amplification, identification and discrimination in *Musa* spp. However, molecular confirmation using *rbc*L sequences showed high conservation level (0.830) and low genetic variability. Reconstruction of the phylogenetic tree based on the *rbc*L gene showed that there were two clades, the first clade consisted of *M. acuminata* (w and cv), *M. velutina*, and *M. ornata* and the second clade consisted of *M. balbisiana* (w) and *M. paradisiaca* (cv).

Acknowledgements

The authors gratefully acknowledge Ministry of Research, Technology and Higher Education, Republic of Indonesia for funding the molecular part of this study with Master's Research Fellowship grant number 1520/UN3/2019. Sincere thanks and great appreciation were also acknowledge to Mrs. Lia Hapsari, S.P., M.Si. (Researcher in Purwodadi Botanic Garden, Research Center for Plant Conservation and Botanic Gardens, Indonesian Institute of Science) for the knowledge and guidance thas has been given, also Molecular genetic Laboratory of Biology Department, Airlangga University for the molecular and genetic facilities.

References

CBoL (Consortium for the Barcoding of Life) Plant Working Group. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Science of the United States of America (PNAS). 106 : 12794–12797. https://doi.org/10.1073_pnas.0905845106

FAO. 2017. *Banana Market Review and Banana Statistics 2015-16.* Rome. [Online]. [download on 2nd October 2018]. Available at http://www.fao.org/-economic/est/estcommodities/bananas/en/.

Hapsari, L., Wahyudi, D., Azrianingsih, R. and Arumingtyas, E.L. 2015. Genome dentification of bananas *(Musa* L.) from East Java Indonesia assessed with PCR-RFLP of the internal transcribed spacers nuclear ribosomal DNA. *International Journal of Biosciences.* 7(3) : 42-52. http://dx.doi.org/10.12692/ijb/7.3.42-52

Hapsari, L., Kennedy, J., Lestari, D.A., Masrum, A. and Lestarini, W. 2017. Ethnobotanical survey of bananas (Musaceae) in six districts of East Java, Indonesia. *Biodiversitas.* 18 (1) : 160-174. https://doi.org/10.13057/biodiv/d180122

Hapsari, L., Azrianingsih, R. and Arumingtyas, E.L. 2018. Genetic variability and relationship of banana cultivars (*Musa* L.) from East Java, Indonesia based on the Internal Transcribed Spacer region nrDNA sequences. *J Trop Biol Conserv.* 15 : 101-120.

Hapsari, L., Trimanto and Wahyudi, D. 2019. Species diversity and phylogenetic analysis of *Heliconia* spp. collections of Purwodadi Botanic Garden (East Java, Indonesia) inferred by *rbc*L gene sequences. *Biodiversitas.* 20(5) : 1266-1283. https://doi.org/10.13057/biodiv/d200505

Hasan, A.S. and Khasim, S.M. 2018. DNA barcoding for differentiating the 11 varietas of *Musa* species. *International Journal of Agricultural Science and Research (IJASR)* 8(4) : 121-130.

Hiariej, A., Arumingtyas, E.L., Widoretno, W. and Azrianingsih, R. 2015. Genetic kinship of tongkat langit banana (*Musa troglodytarum* L.) from Moluccas based on *rbc*L gene sequence. *Indian J Genet.* 75 (4) : 526-528. https://doi.org/10.5958/0975-6906.2015.00085.1

Lamere, A. and Rao, S.R. 2015. Efficacy of RAPD, ISSR and DAMD markers in assessment of genetic variability and population structure of wild *Musa acuminata* colla. *Physiol Mol Biol Plants.* 1-10. https://doi.org/10.1007/s12298-015-0295-1

Lamere, A., Otaghvari, A.M. and Rao, S. R. 2016. 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 Resour Crop Evol*. 1-11. https://doi.org/10.1007/s10722-016-0433-9

Li, L.F. and Ge, X.J. 2017. Origin and domestication of cultivated banana. *Ecological Genetics and Genomics.* 1-2. http://dx.doi.org/10.1016/j.egg.2016.10.001

Simmonds, N.W. and Shepherd, K. 1955. The taxonomy

and origins of the cultivated bananas. Botanical Journal of the Linnean Society 55: 302–312. https://doi.org/10.1111/j.1095-8339.1955.tb00015.x

Vinitha, M.R., Kumar, U.S., Aishwarya, K., Sabu, M. and Thomas, G. 2014. Prospects for discriminating Zingiberaceae species in India using DNA barcodes. *Journal of Integrative Plant Biology*. 56 (8) : 760-773. https://doi.org/10.1111/jipb.12189.

Wattoo, J.I., Saleem, M.Z., Shahzad, M.S., Arif, A., Hameed, A. and Saleem, M.A. 2016. DNA barcoding: amplification and sequences of *rbc*L and *mat*K genome regions in three divergent plant species. *Adv. Life Sci.* 4(1): 03-07.

Zhao, B., Li, J., Yuan, R. and Mao, S. 2017. Adaptive evolution of the *rbc*L gene in the genus *Rheum* (Polygonaceae). *Biotechnology & Biotechnological Equipment*. 31(3): 493-498. http://dx.doi.org/10.1080/13102818.2017.1288072.