

Assessment of Genetic Diversity and Relatedness of Local Banana (*Musa* spp.)
Cultivars in Malaysia using Simple Sequence Repeat (SSR) Markers
(Penilaian Kepelbagaian Genetik dan Perkaitan Kultivar Pisang Tempatan (*Musa* spp.) di Malaysia menggunakan
Penanda Ulangan Jujukan Ringkas (SSR))

AMER IZZAT SAMSUDIN*, SHARMILAH VETARYAN, HOOI WEI YENG & LEE YANG PING

Precision Agriculture and Genomics Department, FGV Innovation Centre, FGV R&D Sdn Bhd, 71760 Bandar Enstek,
Negeri Sembilan, Malaysia

Received: 26 January 2024/Accepted: 27 May 2024

ABSTRACT

Banana (*Musa* spp.) is an economically important crop widely consumed all over the world. Understanding the genetic diversity and relatedness between various banana cultivars is important to continuously improve its fruit characteristics, nutritional value, and disease resistance. This study aimed to develop a DNA fingerprinting panel using polymorphic simple sequence repeat (SSR) markers for the genetic characterisation of 11 banana cultivars in Malaysia. A total of 10 polymorphic SSR markers were developed into a multiplex PCR reaction and capillary electrophoresis to uniquely profile our collection of banana cultivars. The developed fingerprinting panel successfully amplified a total of 95 alleles, with 6 to 13 alleles per SSR marker. The average SSR marker polymorphic information content (PIC) value is 0.812, indicating the informativeness of the panel. Analysis of molecular variance (AMOVA) shows that 97% of the molecular variance from our banana collection is due to inter-variety genetic diversity, while the remaining 3% is due to intra-variety genetic diversity. A population structure analysis groups our collection of banana varieties according to the presence of at least one *M. balbisiana* (B) genome in their genetic makeup. The Cavendish variety, however, showed a distinct structure compared to the other cultivars. This SSR fingerprinting panel provides valuable insights into the genetic diversity and relatedness between banana cultivars in Malaysia. It has the potential to assist future banana breeding initiatives and serve as an effective quality control measure for verifying varieties in a tissue culture facility involved in banana planting materials production.

Keywords: Banana cultivars; DNA fingerprinting; genetic characterisation; polymorphic SSR markers

ABSTRAK

Pisang (*Musa* spp.) adalah tanaman ekonomi yang penting dan banyak dimakan di seluruh dunia. Memahami kepelbagaian genetik dan hubung kait antara kultivar pisang adalah penting bagi terus menambahbaik ciri buahnya, nilai nutrisinya dan ketahanannya terhadap penyakit. Kajian ini bertujuan untuk membangunkan satu panel pencapjarian DNA menggunakan penanda ulangan jujukan ringkas (SSR) untuk pencirian genetik 11 kultivar pisang di Malaysia. Sejumlah 10 penanda SSR yang polimorfisme dibangunkan dalam tindak balas PCR multipleks yang dipadankan dengan elektroforesis kapilari bagi menentukan profil kumpulan kultivar pisang kami. Panel pencapjarian DNA yang dibangunkan ini berjaya mengamplifikasi secara keseluruhannya 95 alel, dengan jumlah 6 hingga 13 alel bagi setiap penanda SSR. Nilai purata kandungan maklumat polimorfik (PIC) bagi penanda SSR adalah 0.812, menunjukkan kebergunaan maklumat panel tersebut. Analisis variasi molekul (AMOVA) menunjukkan bahawa 97% variasi molekul daripada koleksi pisang kami adalah disebabkan oleh kepelbagaian genetik antara varieti, manakala baki 3% lagi adalah disebabkan oleh kepelbagaian genetik dalam varieti. Analisis struktur populasi mengumpulkan koleksi varieti pisang kami berdasarkan kehadiran sekurang-kurangnya satu genom *M. balbisiana* (B) dalam kandungan genetik mereka. Varieti Cavendish pada masa yang sama menunjukkan struktur yang berbeza apabila dibandingkan dengan kultivar lain. Panel pencapjarian SSR ini memberikan maklumat berharga tentang kepelbagaian genetik dan hubung kait antara kultivar pisang di Malaysia. Selain itu, panel ini berpotensi digunakan untuk membantu program biak baka pisang. Ia juga boleh digunakan sebagai salah satu alat kawalan kualiti untuk pengesanan varieti di fasiliti kultur tisu yang terlibat dalam pengeluaran bahan tanaman pisang.

Kata kunci: Kepelbagaian genetik; kultivar pisang; penanda SSR polimorfisme; pencapjarian DNA

INTRODUCTION

Banana (*Musa* spp.) is one of the most important crops worldwide and is widely consumed all around the globe. Its cultivation and trade value play important roles in the livelihoods of many communities and the economic growth of many developing countries. According to the Food and Agriculture Organization of the United Nations (FAO), it has been estimated that more than 100 billion bananas are eaten globally every year. In terms of total global production of food crops, it ranks just behind wheat, rice, and maize. The economic importance of bananas is accredited to their properties of being a good source of energy, highly nutritious, easily accessible, and can be planted all year round.

The majority of banana varieties available on the market have undergone some degree of domestication. Domesticated banana originated from an inter-specific and intra-specific hybrid between two wild banana species; *Musa acuminata* (A genome, $2n = 2x = 22$) and *Musa balbisiana* (B genome, $2n = 2x = 22$). The hybridisation of these *Musa* species has led to the creation of various banana varieties, including the diploid types (AA, AB, BB), triploid types (AAA, AAB, ABB), and tetraploid types (AAAB, AABB, ABBB) (Simmonds & Shepherd 1955). Domesticated bananas are extensively cultivated via asexual propagation methods, allowing them to retain desirable fruit characteristics. Completion of the draft genome for *M. acuminata* and *M. balbisiana* recently has provided better insight into these species (D'Hont et al. 2012; Wang et al. 2019). It is important to study the hybridisation history and genetic diversity of these 2 genomes. The knowledge can be used to assist breeding programs to make further improvements on banana fruit characteristics, nutritional value, and resistance to diseases.

One of the ways to understand the genetic diversity of bananas is through the usage of molecular markers. In bananas, various genetic studies have been conducted using different types of DNA markers such as random amplified polymorphic DNA (RAPD), simple sequence repeat (SSR), single nucleotide polymorphism (SNP), and amplified fragment length polymorphism (AFLP) (Cenci et al. 2021; Hippolyte et al. 2010; Howell et al. 1994; Wang et al. 2007). SSR in particular, is a marker that is highly polymorphic, co-dominant, highly reproducible, and relatively affordable for routine genotyping. This makes it suitable to be used for the development of DNA fingerprinting in bananas. A combination of SSR markers has been utilised for DNA fingerprinting in crop plants

of economic importance, such as coffee, sweet potato, and tomato (Castellana et al. 2020; Meng et al. 2018; Pruvot-Woehl et al. 2020).

DNA fingerprinting can also benefit banana cultivation, especially for commercial farms heavily reliant on the large-scale supply of high-quality planting materials. The tissue culture cloning process, widely adopted and commercialised by many companies, allows the production of a huge number of plantlets from a single sucker to meet the demands of farmers (Suman 2017). Given the importance of the initial starting material, it is imperative to manage the selection process tightly. This will ensure a true-to-type production of banana varieties and the highest conformity to the declared variety. DNA fingerprinting, alongside the traditional selection-based physical properties, emerges as a critical quality control component in achieving these goals.

The objectives of this study were to create an SSR fingerprinting panel to assess the genetic diversity and relatedness of important banana cultivars in Malaysia. The insights gained from this study aim to advance the current understanding of bananas' genetic diversity and further enhance their future cultivation and trade value. Moreover, the panel will serve as a valuable tool in verifying the authenticity of varieties produced in a tissue culture production facility.

MATERIALS AND METHODS

PLANT MATERIALS AND DNA EXTRACTION

A total of 131 leaf samples were collected from 11 banana cultivars maintained at Pusat Penyelidikan Pertanian Tun Razak (PPPTR), Pahang, Malaysia. The banana varieties are locally known in Malaysia as *Saba*, *Nipah*, *Tanduk Lang*, *Rastali*, *Nangka*, *Raja*, *Emas*, *Lemak Manis*, *Jari Buaya*, *Berangan*, and Cavendish. The ploidy of each banana cultivar was determined based on a compilation made by Valmayor et al. (2000). These cultivars are well-maintained, and occasionally, their sword suckers are used as starting materials in cloning activity for commercial purposes. DNA was extracted from approximately 35×10 mm banana leaf using a modified CTAB method, where the sample was ground in 200 μ L CTAB buffer (0.1 M Tris-HCl pH 8, 0.02M Na₂EDTA, 1.4M NaCl, 1% PVP-40, 2% CTAB, 0.2% β -mercaptoethanol) using TissueLyser II (QIAGEN, Germany). Another 800 μ L CTAB buffer was added to the mixture and incubated in water bath at 65

°C for 30 min (Ying & Zaman 2006). The remaining of the protocol followed the general CTAB extraction method (Doyle & Doyle 1990). DNA concentration was quantified using NanoDrop Spectrophotometer ND-1000 (ThermoScientific, USA) and diluted to 15 ng/μL for PCR work.

SSR POLYMORPHISM SCREENING AND FINGERPRINTING PANEL DEVELOPMENT

Thirty SSR markers and their respective primer sequences were retrieved from four scientific papers (Hippolyte et al. 2010; Ning et al. 2007; Ravishankar et al. 2012; Rotchanapreeda et al. 2016). All forward primers of the 30 SSR markers were added with an M13 sequence [5'-GGAAACAGCTATGACCAT-3'] at their 5' end. All reverse primers were added with pig-tail sequence [5'-GTTTCTT-3'] at their 5' end. The M13 sequence [5'-GGAAACAGCTATGACCAT-3'] was also synthesised independently and tagged with 5' 6-FAM (Fluorescein). All primers were produced by Integrated DNA Technologies, USA.

The polymorphism of each SSR marker was screened on representative samples of the banana cultivars. PCR for each SSR marker amplification was conducted in a 10 μL singleplex reaction, by which every PCR run consists of 1X MyTaq HS Mix (Bioline, UK), 0.2 μM of M13-forward primer, 0.2 μM of pig-tailed reverse primer, 0.2 μM of 5' 6-FAM-M13 primer, 150 ng DNA, and sterile water. PCR was carried out using a VapoProtect Mastercycler (Eppendorf, Germany) with the following configuration; initial denaturation of 94 °C for 5 min, followed by 35 PCR cycles of 94 °C for 30 s, annealing at 55 °C for 90 s, extension at 72 °C for 1 min, and final extension of 72 °C for 30 min. The PCR products were run using fragment analysis setup on 3500xL Genetic Analyzer and subsequently analysed using Gene Mapper v 5.0 software (Applied Biosystems, USA). SSR markers suspected of having extra non-unique PCR amplification on the screened cultivars, and showed the presence of four or more alleles were filtered out. The remaining markers were ranked in terms of alleles polymorphism and the ability to distinguish different banana varieties. Ten best polymorphic markers were selected for the development of the SSR fingerprinting panel.

In an effort to streamline resources and enhance efficiency, a multiplex PCR reaction of the ten chosen SSR markers was created to establish the SSR fingerprinting panel. This approach allows for the simultaneous amplification of ten SSR markers in a single-well reaction,

rather than conducting ten separate reactions. In contrast to using the M13 sequence, the forward primers for the multiplex fingerprinting panel were directly labeled at their 5' end with fluorescent dyes of either 5' 6-FAM (FAM), 5' HEX (HEX), and 56-ROXN (ROX) (Integrated DNA Technologies, USA). For marker Ba30, the reverse primer sequence was modified to prevent its amplicons from overlapping with products of another marker tagged with the same fluorescent dye. The concentration of all forward and reverse primers was optimised to get the best PCR amplification profile readable using Gene Mapper v 5.0 software (Applied Biosystems, USA).

CLONING AND SANGER SEQUENCING

The multiplex PCR reaction is designed to detect different SSR alleles based solely on fragment lengths. It is crucial to confirm that the variations between each detected allele are due to the differing number of SSR repeats. Cloning and Sanger sequencing methods were therefore employed to visualize and confirm the presence of the SSR repeat motifs within the PCR amplicons. One representative SSR marker from each source of scientific paper was selected, and PCR was amplified using their respective forward and reverse primers, with a similar PCR setup as the previous step. PCR products containing the SSR markers were run on 1% agarose gels, excised, and purified using Clean-up Gel Extraction kits (Macherey-Nagel, Germany) according to the manufacturer's instructions. The ligation of these PCR products into pJET1.2/blunt plasmid (ThermoScientific, USA) was conducted using the CloneJET PCR cloning kit (ThermoScientific, USA). The vector was transformed into ECOSTM competent *E. coli* cells (Yeastern Biotech, Taiwan) according to the manufacturer's recommendation. Transformed *E. coli* colonies were selected by screening them using colony PCR with the following configuration; initial denaturation of 95 °C for 3 min, followed by 25 PCR cycles of 94 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 1 min. Plasmids containing the correct SSR markers insert were extracted from these *E. coli* cells using the NucleoSpin® Plasmid EasyPure extraction kit according to the manufacturer's instructions (Macherey-Nagel, Germany). The plasmids were Sanger sequenced using the difficult template sequencing protocol by 1st BASE (Apical Scientific, Malaysia). A nucleotide BLAST search was conducted using the cloned sequences to locate the SSR position in *M. acuminata* subsp. *malaccensis* genome assembly (accession number: PRJEA82777) (D'Hont et al. 2012).

DATA ANALYSIS

Heterozygosity (He) and the polymorphic information content (PIC) value of the ten selected markers were calculated using the following formula proposed by Nei (1987) and Botstein et al. (1980), respectively.

$$He = 1 - \sum_{i=1}^n p_i^2$$

n = number of alleles; p_i = frequency of the i^{th} allele

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \left(\sum_{i=1}^n p_j^2 \right)^2 + \sum_{i=1}^n p_i^4$$

where n is the number of alleles; and p_i and p_j is the allele frequency in population i and j , respectively.

Each SSR genotype data was converted into binary format. The presence of an allele was scored as 1, and the absence of the allele was scored as 0. Using DARwin 6 software, dissimilarity analysis was calculated using the dice coefficient method, and a phylogenetic tree was built using the weighted neighbour-joining method with 10,000 bootstraps (Perrier & Jacquemoud-Collet 2006). Genetic distance, principal coordinate analysis (PCoA), and analysis of molecular variance (AMOVA) were computed using GenAIEx 6 (Peakall & Smouse 2006). To identify the population's genetic structure, the binary data was analysed using a Bayesian clustering method in STRUCTURE v 2.3.4 (Pritchard, Stephens & Donnelly 2000). Five independent runs were used to estimate the K value range of 1-10, with each run consisting of a burn-in period of 20,000 followed by 200,000 Markov chain Monte Carlo (MCMC) repetitions. Structure Harvester v0.6.94 was used to determine the optimum number of clusters based on the deltaK values (Earl & VonHoldt 2012).

RESULTS

MARKER ANALYSIS

Our SSR fingerprinting panel was used to screen 131 samples from 11 banana varieties in Malaysia. A total of 95 alleles were detected, with a range of 6 to 13 alleles per marker. The average number of alleles per marker is 9.5. He value is in the range of 0.769 – 0.878, with an average of 0.832. The markers yielded an average PIC value of 0.812. Marker Ba25 has the lowest PIC value of

0.738, and marker Ba8 has the highest PIC value of 0.867. This highlights marker Ba8 as the most polymorphic within our banana collection. All 10 SSR markers in the panel have PIC values of 0.5 to 1, indicating the informativeness of the panel. The combination of these 10 SSR markers has produced a unique SSR fingerprinting profile for each banana variety used in this study (Table 1).

All SSR markers selected for the panel contain dinucleotide repeat motifs. Only marker Ba19, on top of this, also includes a trinucleotide repeat motif within the PCR amplicon. When referred to *M. acuminata* subsp. *malaccensis* strain Doubled-haploid Pahang (DH-Pahang) genome assembly, the SSR markers are located in six out of the eleven banana chromosomes. Three SSRs are located in chromosome A06, two in each of the chromosomes A03 and A08, while one SSR originated from chromosomes A02, A09, and A11, respectively (Table 1).

SSR REGION SEQUENCE

A representative marker from each of the SSR sources was amplified, cloned and Sanger sequenced to visualize and compare the differences in their simple sequence repeat motifs. These markers are Ba8, Ba19, Ba21, and Ba30 (Figure 1). For marker Ba8, which has the dinucleotide repeat motif of $(GA)_n$, our *Berangan* material has an allele of $(GA)_{17}$. The alleles in *M. acuminata* subsp. *malaccensis* assembly and *M. acuminata* genomic clone MA4-7D8 from Ravishankar et al. (2012) are $(GA)_{13}$ and $(GA)_{12}$, respectively. For marker Ba19, Rotchanapreeda et al. (2016) identified a combination of dinucleotide repeat of $(GA)_{17}$ and trinucleotide repeat of $(CTC)_4$ in their *M. balbisiana* clone Tani Pa Tat Luang BBCT28.1. The similar SSR region in *M. acuminata* subsp. *malaccensis* assembly shows no difference in the SSR motif. However, in our *Berangan* material, the dinucleotide repeats for marker Ba19 is $(GA)_{12}$. There are two nucleotide substitutions within the Ba19 trinucleotide repeat in our *Berangan* material, but this does not change its total number of bases compared to *M. acuminata* subsp. *malaccensis* assembly and *M. balbisiana* clone Tani Pa Tat Luang BBCT28.1.

For marker Ba21, Ning et al. (2007) did not provide any sequence information regarding their material. Nevertheless, the alignment of our *Nangka* material with *M. acuminata* subsp. *malaccensis* assembly shows a repeat motif of $(AG)_{16}$ and $(AG)_{12}$, respectively. Concerning marker Ba30, our *Berangan* material exhibited a repeat motif of $(TC)_{22}$. The SSR region's length of *M. balbisiana* material from Hippolyte et al. (2010)

and the *M. acuminata* subsp. *malaccensis* assembly was found to be the same. Interestingly, in the *M. balbisiana* material from Hippolyte et al. (2010), a nucleotide substitution from A to T was observed at the start of the SSR repeat motif, which results in its SSR starting position being earlier than others.

ANALYSIS OF MOLECULAR VARIANCE (AMOVA)

The AMOVA result shows that the genetic diversity among the different banana cultivars accounts for 97% of the total molecular variance in our collection (Table 2). Only as little as 3% variance occurs within the same banana variety at $p < 0.001$. The variance component is as low as 0.408 within the same variety and is 14.927 among different varieties. The results indicated very low genetic diversity within the same banana variety in our collection. Most of the individuals within the same variety are identical to one another.

DIVERSITY AND RELATEDNESS OF THE 11 BANANA CULTIVARS

The SSR genotype data estimated the population structure of 11 banana varieties. The analysis was done in STRUCTURE v 2.3.4 based on a K value of 1 to 10. The two best values of K determined by the software are K=2 and K=7 (Figure 2(a)). For K=2, the Bayesian analysis separates the Cavendish variety from other cultivars. For K=7, *Nipah*, *Saba*, and *Tanduk Lang* clustered into

the same structure. *Nangka* and *Rastali* have the same population structure, while *Emas* and *Lemak Manis* are diploid AA with the same population structure (Figure 2(b)).

A 2D principal coordinate analysis (PCoA) was plotted based on Nei's genetic distance, which was calculated in GenAlEx 6 (Figure 3). The result indicates that the Cavendish banana appears to be clustered further away from the other varieties. Other banana varieties can be seen separated based on their genomic composition, specifically, whether they contain at least one copy of the *M. balbisiana* genome. The 1st PCoA component explains 26.7% of the genotype data, while the 2nd and 3rd components explain 21.3% and 12.4%, respectively.

The SSR genotype data was utilised to construct a phylogenetic tree. Similar to the PCoA, the resulting tree groups the banana varieties according to the presence and absence of a copy of the *M. balbisiana* genome (Figure 4). Varieties that carry at least one *M. balbisiana* genome, such as *Saba*, *Nipah*, *Tanduk Lang*, *Rastali*, *Nangka*, and *Raja*, cluster together in the same clade. Diploid and triploid *M. acuminata* including *Berangan*, *Lemak Manis*, *Emas*, and *Jari Buaya*, also cluster together. Cavendish forms a separate cluster on its own but is closely related to the AA or AAA varieties. The clustering formed here provides valuable insight into the genetic relationship of banana varieties in Malaysia.

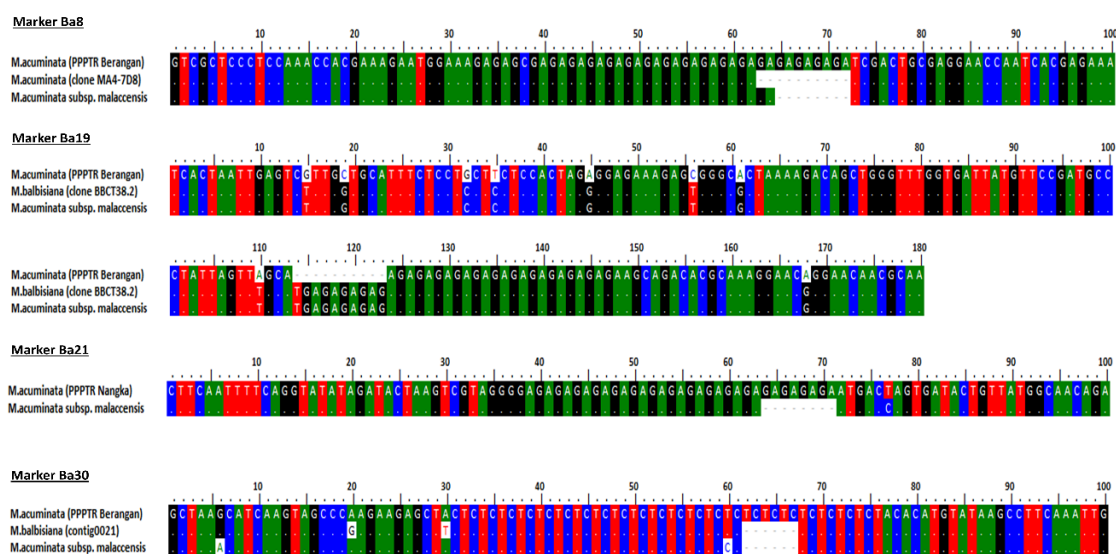


FIGURE 1. Multiple sequence alignments of four SSR markers used for the development of the fingerprinting panel. Each SSR was selected from different sources. The different lengths of the SSR repeat motifs are proven to be the reason behind SSR polymorphism

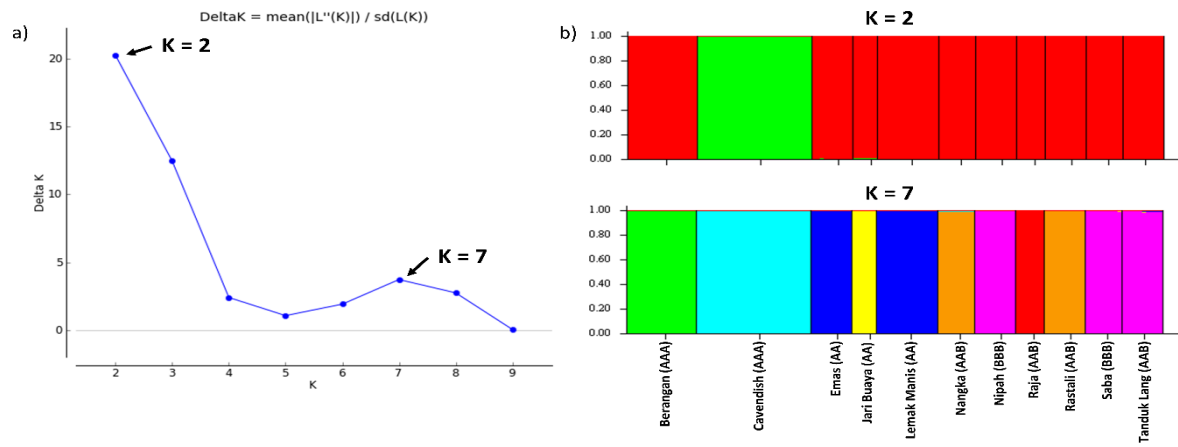


FIGURE 2. (a) Estimates of the best number of population structures based on DeltaK values, using Structure Harvester v0.6.94 (Earl & VonHoldt 2012). The analysis identified two and seven as the optimal numbers for population structure. (b) Barplots of banana varieties determined by STRUCTURE software v 2.3.4. Each column represents a single banana variety. The y-axis indicates Q values. Varieties with identical colours belong to the same population structure at each respective K value

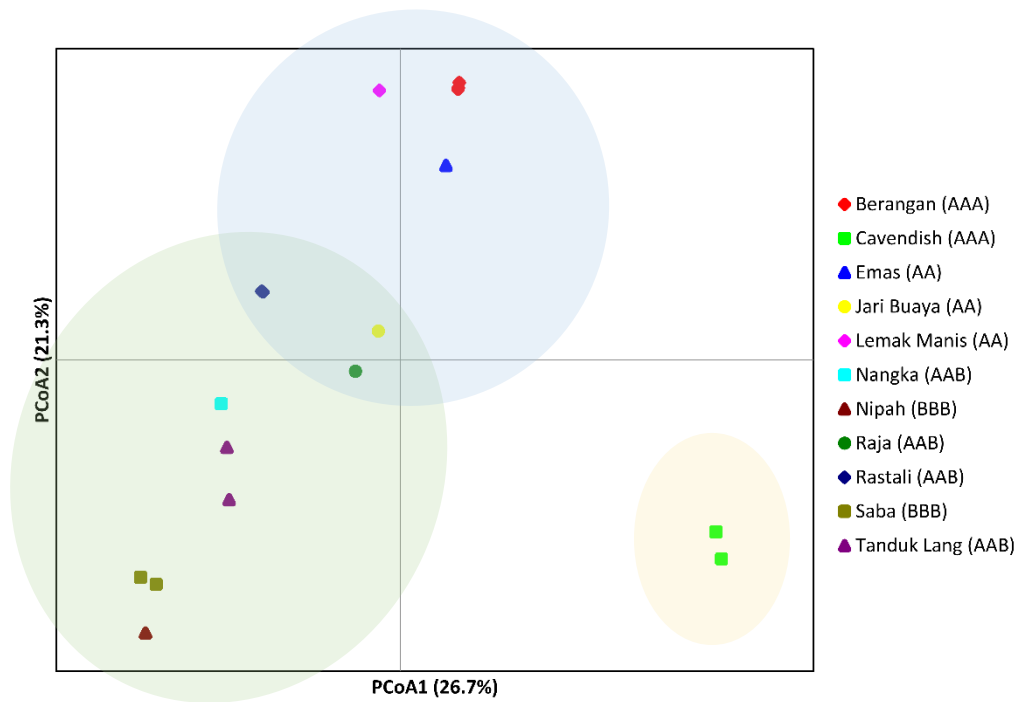


FIGURE 3. A 2D principal coordinate analysis of the 11 Malaysian banana varieties plotted based on Nei's genetic distance. The ploidy of each banana variety is shown in the figure legend. Varieties such as *Nangka*, *Nipah*, *Raja*, *Rastali*, *Saba*, and *Tanduk Lang*, with all having at least one copy of the *M. balbisiana* genome, are clustered towards the bottom left of the plot

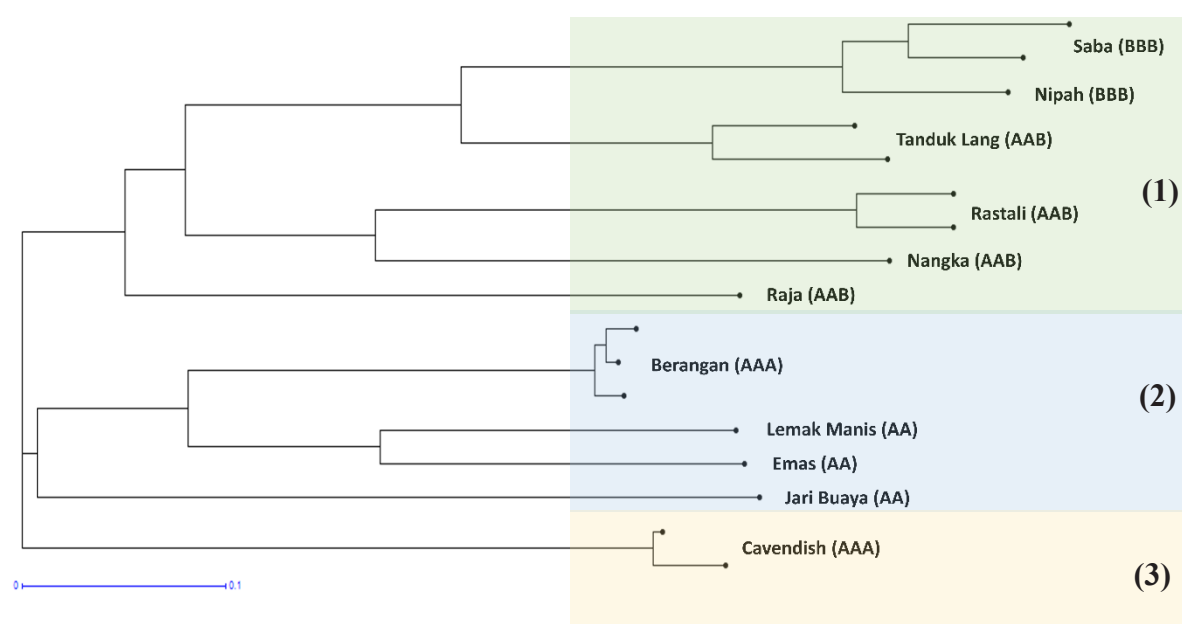


FIGURE 4. Phylogenetic tree of 11 Malaysian banana varieties constructed using the weighted neighbour-joining method. There are three clusters formed; (1) varieties with at least one *M. balbisiana* genome, (2) varieties with only *M. acuminata* genome, and (3) Cavendish variety forming a distinct cluster

DISCUSSIONS SSR MARKERS

The SSR fingerprinting panel developed here amplified a total of 95 alleles on 11 Malaysian banana cultivars, with a range of 6 to 13 alleles per marker. With that, the panel differentiated banana cultivars of high economic importance in Malaysia. This panel was initially developed with prior knowledge of SSR polymorphism in previous studies. In comparison to those studies, Ravishankar et al. (2012) amplified a total of 88 alleles from 26 polymorphic SSR markers, with a range of 2 to 6 alleles per locus. This was tested on 15 banana germplasm accessions in India. Rotchanapreeda et al. (2016) managed to amplify a total of 471 polymorphic alleles using 24 SSR markers, with a range of 2 to 41 alleles per locus. The higher number of polymorphic alleles was due to the higher number of samples tested in the work. Ning et al. (2007) tested 10 SSR markers on 216 banana accessions from China and France. A total of 92 alleles was amplified, with a range of 5 to 15 alleles per locus. Another similar study reported by Creste et al. (2003) amplified a total of 67 alleles using 11 SSR markers on 35 banana accessions from Brazil.

Botstein et al. (1980) suggested that genetic markers are highly informative if they have a PIC value above 0.5, moderately informative if their PIC value is between 0.25 and 0.50, and less informative if their PIC value is less than 0.25. Typically, for SSR markers, the average PIC value within the range of 0.5 to 0.9 would be suitable for plant genetic analysis as this indicates a high level of polymorphism. Therefore, combining multiple polymorphic markers into the same SSR fingerprinting panel will improve its resolution power. The 10 SSR markers used in our fingerprinting panel demonstrate an average PIC value of 0.812, with the lowest and highest PIC values recorded as 0.738 and 0.867, respectively. This implies that our fingerprinting panel is well-suited to distinguish banana varieties from various sources genetically. Similar principles of SSR polymorphism have also been used in the development of fingerprinting technology for other crop species, such as potato and wheat, to generate genetic profiles of various cultivars (Prasad et al. 2000; Spanoghe et al. 2015). The average PIC value of the SSR markers for wheat cultivars in Prasad et al. (2000) study was 0.71, which, like our study,

TABLE 1. Characteristic of ten SSR markers developed into the SSR fingerprinting panel

Marker name	Original marker name	Source	Repeat Motif	Allele sizes	No. of observed Allele	PIC	He	Chromosome location in <i>M. acuminata</i> subsp. <i>malaccensis</i>	Coordinate in <i>M. acuminata</i> subsp. <i>malaccensis</i> ^a
Ba3	Ma_GSS103	Ravishankar et al. (2012)	(AG) ₂₈	348 - 432	12	0.8230	0.8380	A06	9,565,573 (plus)
Ba4	Ma_GSS141	Ravishankar et al. (2012)	(CT) ₁₂	187 - 226	10	0.8384	0.8548	A11	157,549 (plus)
Ba6	Ma_GSS160	Ravishankar et al. (2012)	(TC) ₁₅	339 - 378	7	0.7938	0.8180	A03	10,992,588 (minus)
Ba7	Ma_GSS198	Ravishankar et al. (2012)	(CT) ₉	148 - 169	6	0.7519	0.7825	A08	45,206,798 (plus)
Ba8	Ma_GSS203	Ravishankar et al. (2012)	(GA) ₁₂	135 - 175	13	0.8670	0.8775	A02	21,076,477 (plus)
Ba19	BB_CT-38.2	Rotchanapreeda et al. (2016)	(GA) ₁₇ (CTC) ₄	261 - 288	10	0.8448	0.8601	A09	7,968,112 (minus)
Ba21	16a	Ning et al. (2007)	(AG)	168 - 243	12	0.8264	0.8415	A06	40,858,345 (plus)
Ba25	mMaCIR130	Hippolyte et al. (2010)	(TG) ₁₇	304 - 307	6	0.7377	0.7689	A03	27,061,295 (plus)
Ba26	mMaCIR168	Hippolyte et al. (2010)	(CA) ₇	252 - 276	7	0.7733	0.8028	A08	38,836,353 (plus)
Ba30	mMaCIR241	Hippolyte et al. (2010)	(TC) ₂₀	446 - 490	12	0.8589	0.8719	A06	40,674,991 (minus)

^aplus and minus refer to the forward and reverse strands of *M. acuminata* subsp. *malaccensis* genome assembly, respectively

TABLE 2. Analysis of molecular variance within and among Malaysian banana varieties

Variation source	Degree of freedom	Sum square	Mean square	Estimated variance	%	P value
Among variety	10	1737.853	173.785	14.927	97%	< 0.001
Within variety	120	49.002	0.408	0.408	3%	< 0.001
Total	130	1786.855		15.335	100%	

falls within the 0.5 to 0.9 range, indicating the markers' suitability and informativeness in distinguishing different cultivars of the same crop species. Additionally, our method of detecting and scoring SSR alleles is similar to the majority of recent publications on SSR fingerprinting in plants and animals, where the usage of capillary electrophoresis technology is more prevalent (Kim et al. 2020; Sánchez et al. 2020). This method is a better alternative to the tedious work of scoring SSR alleles on polyacrylamide gel electrophoresis and offers better SSR alleles resolution of up to 1 base pair.

SSR is a category of short tandem repeats with a unit length of one to ten nucleotides. It is estimated to have a mutation rate of 10^3 to 10^6 per cell generation, and the mutation usually involves the addition or deletion of its repeat unit. One model that explains SSR polymorphism is the strand slippage replication model. In this model, it is proposed that DNA replication errors occur when mispairing happens between the template and the nascent strand. The template or nascent strand can loop out, leading to repeat expansion. Following that, unequal crossing over may happen during recombination events, leading to either contraction or expansion of SSR (Gemayel et al. 2012; Verstrepen et al. 2005). Over an extensive evolutionary history, SSR has undergone numerous replication events, and inevitably, expansion or contraction of the SSR repeat motif will happen. This characteristic makes SSR a suitable candidate marker for developing DNA fingerprinting panel, as it increases the likelihood of detecting allelic differences within and between various plant species varieties. Our Sanger sequencing result agrees with the general concept of SSR polymorphism. The size of SSR repeat motifs detected within our cultivars, as compared to the other sources, is the cause of SSR alleles polymorphism detected by the fingerprinting panel. Our data also demonstrate the relatively stable nature of SSR mutations, as individuals

from the same banana variety exhibit similar genetic profiles.

MATERIALS DIVERSITY

Our findings indicate that the intra-variety and inter-variety genetic diversity of our banana collections are at 3% and 97%, respectively. The low intra-variety genetic diversity may be attributed to the limitation of this study, as the cultivars tested were sourced from a single site located in central Peninsular Malaysia. Every banana cultivar in our collection has undergone a severe genetic bottleneck due to its preservation method, where each variety is asexually propagated from the cuttings of mother plants. Essentially, this method of re-generating banana plantlets leads to the creation of clones of the mother plants. While this process may limit genetic diversity, it is commercially advantageous in producing identical high-yielding banana varieties for commercial planting. In contrast, the genetic diversity among different banana varieties is much higher. Various factors could be attributed to this result, such as differences in phenotypic characteristics, selection pressure, history of sexual reproduction, and history of genome hybridisation between the different *Musa* species.

The population structure analysis shows two optimum population structures for the dataset at $K=2$ and $K=7$. At $K=2$, the Cavendish variety forms a distinct structure compared to the rest. The other genetic diversity analyses further support the population structure result as the Cavendish variety exhibits greater genetic distance from other cultivars in the PCoA plot. It also forms a single clade in the phylogenetic analysis. Similar to our result, Sardos et al. (2016) also identified groups of AA/AAA, including the Cavendish variety, that did not cluster with the remaining *M. acuminata* cultivars. This is further supported by Martin et al. (2020), who reported that the triploid Cavendish banana variety

received genome contribution from at least one of an uncharacterised genetic pool, apart from the known *M. acuminata* subspecies such as *banksii*, *malaccaensis*, *zebrina*, and *burmannica*. This helps explain the reason behind its genetic makeup being quite different from the other diploid and triploid *M. acuminata* cultivars.

The occurrence of a smaller DeltaK peak at K=7 shows an additional level of clustering within the Malaysian banana cultivars. This in-depth level of clustering is better demonstrated by the PCoA and dendrogram plots. Our dendrogram analysis identified 3 major clades within the Malaysia banana varieties. The 2 biggest clades show a distinct separation of cultivars based on the presence of the B genome. Varieties such as *Raja*, *Nangka*, *Rastali*, *Tanduk Lang*, *Nipah*, and *Saba* form the same major clade due to each variety having at least one copy of the *M. balbisiana* genome. Our result is consistent with Rotchanapreeda et al. (2016), where different varieties of banana cultivars in Thailand were clustered based on the occurrence of a copy of the B genome. As *Musa* A and B genomes diverge around 20.9 - 27.9 million years ago, there is a long separation time in making the genome quite distinct from one another (Wang et al. 2019). Diploid AA and triploid AAA cultivars do not form distinct, separate clades based on our data, which evidently supports that hybridisation of the AA cultivar that gave rise to the AAA cultivar may have occurred through separate hybridisation events (Perrier et al. 2011).

Anthropogenic factor during the domestication period of bananas might have significantly influenced the genetic structure of Malaysian banana cultivars. Human intervention, particularly in the form of selective breeding for certain traits such as seedless, improved fruit quality, and higher yield has likely led to the creation of the modern cultivars. It is also highly probable that certain intrinsic traits of *M. acuminata*, known for its sweet and pleasant flavour, and *M. balbisiana*, recognized for its starchy nature and high dry matter content, were chosen during the domestication process. Interestingly, the preference for consumption methods also appears to have influenced the genetic clustering of these cultivars. Varieties such as *Saba*, *Nipah*, and *Tanduk Lang*, which are typically cooked before being eaten, are closely clustered together. In contrast, varieties that are usually eaten raw such as *Emas*, *Berangan*, and *Lemak Manis*, form another close cluster. Nayar (2010) has noted that bananas become sturdier and starchier as the levels of the B genome increase. It has since been observed that the B genome's presence, coupled with

intensive selection, has enhanced the starchiness and dry matter content in A-genome bananas, making them more appealing as cooking bananas. Thus, the observed clustering of cultivars in our study not only highlights the impact of human intervention on the genetic structure of modern cultivars in Malaysia, but also emphasizes the interconnection between genetic selection and culinary preferences in shaping the diversity of banana cultivars we see today.

Domesticated banana cultivars typically propagate using part of their underground rhizome to form daughter plantlets called suckers, which eventually grow into mature banana plants. However, this natural propagation method is insufficient in supplying many seedlings for commercial planting. To address this issue, many commercial companies have adopted the plant cloning procedure by utilising suckers for the mass production of banana planting materials. Additional to assessing the cultivars' genetic diversity, this panel can be used to assess sucker fidelity to ensure that it conforms to the correct variety. Occasionally, samples can be tested in the tissue culture facility to ensure that the intermixing of different banana varieties does not occur.

CONCLUSION

In this study, we have demonstrated the usage of polymorphic SSR markers in constructing an effective fingerprinting panel for bananas. The panel shows the genetic diversity and relatedness between 11 economically important banana cultivars in Malaysia. The data from the diversity and relatedness analyses is useful for broadening the cultivars' genetic makeup through future banana breeding programs. Moreover, the application of this panel in future banana conservation work, especially in authenticating similar cultivars from different parts of the country, can serve as a valuable tool for preserving Malaysia's banana genetic diversity. Furthermore, the panel has a potential application to ensure a true-to-type production of banana planting material in a tissue culture facility.

ACKNOWLEDGEMENTS

We would like to thank everyone from the Genomics, Strategic Crops and Tissue Culture Production Units for their contribution and assistance in completing this project. We are also grateful to FGV R&D Sdn Bhd for providing the necessary financial support that made this work possible (FR-0606-80-0221). All authors declare there is no competing interest during the production of this work.

REFERENCES

- Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32(3): 314-331.
- Castellana, S., Ranzino, L., Beritognolo, I., Cherubini, M., Luneia, R., Villani, F. & Mattioni, C. 2020. Genetic characterization and molecular fingerprint of traditional Umbrian tomato (*Solanum lycopersicum* L.) landraces through SSR markers and application for varietal identification. *Genetic Resources and Crop Evolution* 67(7): 1807-1820.
- Cenci, A., Sardos, J., Hueber, Y., Martin, G., Breton, C., Roux, N., Swennen, R., Carpentier, S.C. & Rouard, M. 2021. Unravelling the complex story of intergenomic recombination in ABB allotriploid bananas. *Annals of Botany* 127(1): 7-20.
- Creste, S., Tullmann Neto, A., de Oliveira Silva, S. & Figueira, A. 2003. Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. *Euphytica* 132: 259-268.
- D'Hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., Da Silva, C., Jabbari, K., Cardi, C., Poulain, J., Souquet, M., Labadie, K., Jourda, C., Lenggellé, J., Rodier-Goud, M., Alberti, A., Bernard, M., Correa, M., Ayyampalayam, S., McKain, M.R., Leebens-Mack, J., Burgess, D., Freeling, M., Mbéguié-A-Mbéguié, D., Chabannes, M., Wicker, T., Panaud, O., Barbosa, J., Hribova, E., Heslop-Harrison, P., Habas, R., Rivallan, R., Francois, P., Poirion, C., Kilian, A., Burthia, D., Jenny, C., Bakry, F., Brown, S., Guignon, V., Kema, G., Dita, M., Waalwijk, C., Joseph, S., Dievart, A., Jaillon, O., Leclercq, J., Argout, X., Lyons, E., Almeida, A., Jeridi, M., Dolezel, J., Roux, N., Risterucci, A.M., Weissenbach, J., Ruiz, M., Glaszmann, J.C., Qué-tier, F., Yahiaoui, N. & Wincker, P. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488(7410): 213-217.
- Doyle, J.J. & Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12(1): 13-15.
- Earl, D.A. & VonHoldt, B.M. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources* 4: 359-361.
- FAO. 2023. *Banana Facts and Figures*. <https://www.fao.org/economic/est/est-commodities/oilcrops/bananas/bananafacts/en/#.YvMccXZBy72> Accessed July 13 2023.
- Gemayel, R., Cho, J., Boeynaems, S. & Verstrepen, K.J. 2012. Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes* 3(3): 461-480.
- Hippolyte, I., Bakry, F., Seguin, M., Gardes, L., Rivallan, R., Risterucci, A.M., Jenny, C., Perrier, X., Carreel, F. & Argout, X. 2010. A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biology* 10(1): 1-18.
- Howell, E.C., Newbury, H.J., Swennen, R.L., Withers, L.A. & Ford-Lloyd, B.V. 1994. The use of RAPD for identifying and classifying *Musa* germplasm. *Genome* 37(2): 328-332.
- Kim, M.J., Lee, Y.M., Suh, S.M. & Kim, H.Y. 2020. Species identification of red deer (*Cervus elaphus*), roe deer (*Capreolus capreolus*), and water deer (*Hydropotes inermis*) using capillary electrophoresis-based multiplex PCR. *Foods* 9(8): 982.
- Martin, G., Cardi, C., Sarah, G., Ricci, S., Jenny, C., Fondi, E., Perrier, X., Glaszmann, J.C., D'Hont, A. & Yahiaoui, N. 2020. Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *The Plant Journal* 102(5): 1008-1025.
- Meng, Y.S., Zhao, N., Li, H., Zhai, H., He, S.Z. & Liu, Q.C. 2018. SSR fingerprinting of 203 sweetpotato (*Ipomoea batatas* (L.) Lam.) varieties. *Journal of Integrative Agriculture* 17(1): 86-93.
- Nayar, M.N. 2010. The bananas: Botany, origin, dispersal. *Horticultural Reviews* 36(3): 117-164.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- Ning, S.P., Xu, L.B., Lu, Y., Huang, B.Z. & Ge, X.J. 2007. Genome composition and genetic diversity of *Musa* germplasm from China revealed by PCR-RFLP and SSR markers. *Scientia Horticulturae* 114(4): 281-288.
- Peakall, R. & Smouse, P.E. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6(1): 288-295.
- Perrier, X., De Langhe, E., Donohue, M., Lentfer, C., Vrydaghs, L., Bakry, F., Carreel, F., Hippolyte, I., Horry, J.P., Jenny, C., Lebot, V., Risterucci, A.M., Tomekpe, K., Doutrelepont, H., Ball, T., Manwaring, J., de Maret, P. & Denham, T. 2011. Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proceedings of the National Academy of Sciences* 108(28): 11311-11318.
- Perrier, X. & Jacquemoud-Collet, J.P. 2006. *DARwin software* <http://darwin.cirad.fr/darwin>
- Prasad, M., Varshney, R. K., Roy, J., Balyan, H. & Gupta, P. 2000. The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. *Theoretical Applied Genetics* 100(3-4): 584.
- Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2): 945-959.
- Pruvot-Woehl, S., Krishnan, S., Solano, W., Schilling, T., Toniutti, L., Bertrand, B. & Montagnon, C. 2020. Authentication of *Coffea arabica* varieties through DNA fingerprinting and its significance for the coffee sector. *Journal of AOAC INTERNATIONAL* 103(2): 325-334.
- Ravishankar, K.V., Vidhya, L., Cyriac, A., Rekha, A., Goel, R., Singh, N. & Sharma, T. 2012. Development of SSR markers based on a survey of genomic sequences and their molecular analysis in banana (*Musa* spp.). *The Journal of Horticultural Science and Biotechnology* 87(1): 84-88.

- Rotchanapreeda, T., Wongniam, S., Swangpol, S.C., Chareonsap, P.P., Sukkaewmanee, N. & Somana, J. 2016. Development of SSR markers from *Musa balbisiana* for genetic diversity analysis among Thai bananas. *Plant Systematics and Evolution* 302(7): 739-761.
- Sánchez, E., Solano, W., Gatica-Arias, A., Chavarría, M. & Araya-Valverde, E. 2020. Microsatellite DNA fingerprinting of *Coffea* sp. germplasm conserved in Costa Rica through singleplex and multiplex PCR. *Crop Breeding and Applied Biotechnology* 20(1): e27812013.
- Sardos, J., Perrier, X., Doležel, J., Hřibová, E., Christelová, P., Van Den Houwe, I., Kilian, A. & Roux, N. 2016. DArT whole genome profiling provides insights on the evolution and taxonomy of edible Banana (*Musa* spp.). *Annals of Botany* 118(7): 1269-1278.
- Simmonds, N.W. & Shepherd, K. 1955. The taxonomy and origins of the cultivated bananas. *Botanical Journal of the Linnean Society* 55(359): 302-312.
- Spanoghe, M., Marique, T., Rivière, J., Lanterbecq, D. & Gadenne, M. 2015. Investigation and development of potato parentage analysis methods using multiplexed SSR fingerprinting. *Potato Research* 58: 43-65.
- Suman, S. 2017. Plant tissue culture: A promising tool of quality material production with special reference to micropropagation of banana. *Biochemical and Cellular Archives* 17(1): 1-26.
- Valmayor, R.V., Jamaluddin, S.H., Silayoi, B., Kusumo, S., Danh, L.D., Pascua, O.C. & Espino, R.R.C. 2000. Banana cultivar names and synonyms in Southeast Asia. In *Advancing Banana and Plantain R&D in Asia and the Pacific*, edited by Molina, A.B. & Roa, V.N. *Proceeding of the 9th INIBAP-ASPNET Regional Advisory Committee* held at South China Agricultural University, Guangzhou, China 2-5 November 1999, pp. 55-66.
- Verstrepen, K.J., Jansen, A., Lewitter, F. & Fink, G.R. 2005. Intragenic tandem repeats generate functional variability. *Nature Genetics* 37(9): 986-990.
- Wang, X.L., Chiang, T.Y., Roux, N., Hao, G. & Ge, X.J. 2007. Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by AFLP markers. *Genetic Resources and Crop Evolution* 54(5): 1125-1132.
- Wang, Z., Miao, H., Liu, J., Xu, B., Yao, X., Xu, C., Zhao, S., Fang, X., Jia, C., Wang, J., Zhang, J., Li, J., Xu, Y., Wang, J., Ma, W., Wu, Z., Yu, L., Yang, Y., Liu, C., Guo, Y., Sun, S., Baurens, F. C., Martin, G., Salmon, F., Garsmeur, O., Yahiaoui, N., Hervouet, C., Rouard, M., Laboureau, N., Habas, R., Ricci, S., Peng, M., Guo, A., Xie, J., Li, Y., Ding, Z., Yan, Y., Tie, W., D'Hont, A., Hu, W. & Jin, Z. 2019. *Musa balbisiana* genome reveals subgenome evolution and functional divergence. *Nature Plants* 5(8): 810-821.
- Ying, S.T. & Zaman, F.Q. 2006. DNA extraction from mature oil palm leaves. *Journal of Oil Palm Research* 18: 219.

*Corresponding author; email: amer.iz@fgvholdings.com