Dominant Allele Phylogeny and Constitutive Subgenome Haplotype Inference in Bananas Using Mitochondrial and Nuclear Markers

Kariuki Samwel Muiruri1,2, Anne Britt3, Nelson Onzere Amugune2, Edward Nguu4, Simon Chan3,†, and Leena Tripathi1,*

1International Institute of Tropical Agriculture (IITA), Nairobi, Kenya
2School of Biological Sciences, University of Nairobi, Kenya
3Department of Plant Biology, University of California, Davis
4Department of Biochemistry, University of Nairobi, Kenya

*Corresponding author: E-mail: l.tripathi@cgiar.org.
Accepted: August 27, 2017
†Deceased

Data deposition: The sequences obtained were deposited in the NCBI under accession numbers KP751256 - KP751292 for NAD1 and KP751293 - KP751328 for CENH3.

Abstract

Cultivated bananas (Musa spp.) have undergone domestication patterns involving crosses of wild progenitors followed by long periods of clonal propagation. Majority of cultivated bananas are polyploids with different constitutive subgenomes and knowledge on phylogenies to their progenitors at the species and subspecies levels is essential. Here, the mitochondrial (NAD1) and nuclear (CENH3) markers were used to phylogenetically position cultivated banana genotypes to diploid progenitors. The CENH3 nuclear marker was used to identify minimum representative haplotype number in polyploids and diploid bananas based on single nucleotide polymorphisms. The mitochondrial marker NAD1 was observed to be ideal in differentiating bananas of different genomic constitutions based on size of amplicons as well as sequence. The genotypes phylogenetically segregated based on the dominant genome; AAB genotypes grouped with AA and AAA, and the ABB together with BB. Both markers differentiated banana sections, but could not differentiate subspecies within the A genomic group. On the basis of CENH3 marker, a total of 13 haplotypes (five in both diploids and triploids, three in diploids, and rest unique to triploids) were identified from the genotypes tested. The presence of haplotypes, which were common in diploids and triploids, stipulate possibility of a shared ancestry in the genotypes involved in this study. Furthermore, the presence of multiple haplotypes in some diploid bananas indicates their being heterozygous. The haplotypes identified in this study are of importance because they can be used to check the level of homozygosity in breeding lines as well as to track segregation in progenies.

Key words: banana, haplotype, genotypes, phylogeny, NAD1 and CENH3.

Introduction

Banana including plantain is among the world’s most important staple food crops cultivated with an annual production of ~145 million tonnes of which <17% is for export trade (FAOSTAT 2014). Africa contributes one-third of the global production with East Africa being the largest banana-growing region accounting for ~40% of the total production in Africa. Banana production is hampered by pests and diseases including weevils and nematodes, fusarium wilt caused by the fungi Fusarium oxysporum f. sp. cubense, banana Xanthomonas wilt caused by bacteria Xanthomonas campestris pv. musacearum, black Sigatoka caused by Mycosphaerella fijiensis and viral diseases caused by Banana streak virus (BSV) and Banana bunchy top virus (BBTV; Jones 2000; Surridge et al. 2003; Tripathi et al. 2009). Large-scale producers can afford chemical control for the pests and diseases but the...
smallholder farmers, who produce at least 85% of the global total, cannot afford chemical control. The use of resistant cultivars is considered the most effective, economical and environmentally friendly approach to control pests and diseases. Major banana breeding programs therefore target creation of pests and diseases resistant varieties by introgressing favorable characters from wild species and genotypes (Lorenzen et al. 2010). Breeding is however limited by narrow genetic diversity and the sterility of the polyploid genomes of cultivated banana (Lorenzen et al. 2010). This has resulted in renewed interest in cultivated and noncultivated bananas and led to an increase of interest regarding their evolutionary relationships (Christelov et al. 2011; Häkkinen 2013; Hřibová et al. 2011; Boonruangrod et al. 2008; Baurens et al. 2010).

The genus *Musa* has for a long time been classified into five sections, namely *Eumusa* (*n* = 11), *Rhodochlamys* (*n* = 11), *Callimusa* (*n* = 10 or *n* = 9), *Australimusa* (*n* = 10; Cheesman 1950), and *Ingentimusa* (*n* = 7; Argent 1976), with majority of cultivated bananas being grouped in *Eumusa*. However, recent studies based on molecular phylogenetics have found no support for the five sections and these have now been reappraised and collapsed to only two; section *Musa* (comprising the former sections *Rhodochlamys* and *Eumusa*) and *Callimusa* (with the former *Callimusa*, *Australimusa*, and *Ingentimusa*; Häkkinen 2013). However, because the reappraisal efforts by Häkkinen was a proposition, the old sections will be maintained in this study, but inconsistencies and similarities (to the old and new sections) will be highlighted whenever they arise. There are four progenitors of the domesticated banana *Musa acuminata* Colla (A genome, 2n = 22), *Musa balbisiana* Colla (B genome, 2n = 22), and to a lesser extent, *Musa schizocarpa* (S genome) and *Musa textilis/Musa maclayi* (T genome; Cheesman 1950). On the basis of chloroplast and mitochondrial genomes analysis, the species of *M. acuminata* have been classified into eight subspecies (subsp.), including subsp. *bankii* from Papua New Guinea, subsp. *errans* from Philippines, subsp. *zebrina* from the Indonesia, subsp. *malaccensis*, subsp. *microcarpa*, subsp. *burmannicoides*, subsp. *burmannica* and subsp. *siamea* with a distribution in Thailand, Malaysia, and Philippines (Carreel et al. 2002). No subspecies have been identified in *M. balbisiana* until now. Cultivated *Musa* clones are mostly derived from intersubspecific crossing of *M. acuminata* (A genome) and to a lesser extent interspecific hybrids of *M. acuminata × M. balbisiana* (B genome) although there still exists a few other hybrids from *M. acuminata × M. schizocarpa* (S genome) and *M. acuminata × M. textilis* (T genome) crosses respectively (Hřibová et al. 2011). The haploid genome contributed by *M. acuminata* is referred to as the A genome and that from *M. balbisiana* the B genome. On the basis of morphological characters and genetic composition, cultivated bananas were grouped into the genotypes AA, AB, AAA, AAB, ABB (Simmonds and Shepherd 1955). The majority of economically important banana cultivars belongs to the AAA genotype, which is characterized by low genetic diversity due to clonal propagation of nearly-sterile dominant cultivars (Baurens et al. 2010).

Cultivated banana and plantain commonly exist as autopolyploids and allopolyploids, the later representing different combinations of A and B and sometimes S and T genomes. Mutations may exist in heterozygous allele sequences within individual subgenomes as well as in homoeologous sequences among the polyploid’s constitutive subgenomes (Kaur et al. 2012). Identification of number of gene-alleles present in polyploids is very important as variability in numbers has been observed to have either additive or nonadditive expression patterns. Furthermore, gene redundancy in polyploid bananas covers-up for loss-of-function or even rescue lethal mutations arising in the homeologs and this masks evolutionary changes because they are not morphologically visible. Moreover, allele-dosage can help in identification of haplotypes which can in turn help identify progenitors contributing to the polyploid genome (Baurens et al. 2010; Boonruangrod et al. 2008; Li et al. 2013). The variations due to subgenomic alleles may sometimes be difficult to identify without mapping of the polymorphic loci, clone-based-sequencing or use of the latest approaches of haplotyping like genotyping by sequencing (GBS). Moreover, some of mapping approaches like Simple Sequence Repeats (SSRs) may not clearly identify biases like segregation bias common in mapping populations. Furthermore, most subgeneric allelic discriminatory methods are expensive and out of reach for majority of laboratories especially those in developing countries. Moreover, the approaches used in haplotype identification in banana have included manually comparing and scoring amplified bands on agarose gel (Boonruangrod et al. 2008) which is laborious and could result in bias. Despite the advancement in haplotype identification techniques like next generation sequencing, the prices are still prohibitive for them to be used in some of the laboratories. Low cost approaches that simplify the identification of progenitors from cultivated bananas would contribute more to banana breeding efforts.

Amplified sequences have previously been used to identify homologs, homoeologs, SNP-alleles as well as haplotypes in several crops like maize (Ching et al. 2002; Ding and Cantor 2003). The use of amplified sequences to identify haplotypes has led to development of algorithms that enable phasing (Neigenfind et al. 2008; Browning and Browning 2011; Efros and Halperin 2012; Guo et al. 2012). As DNA in polyploids contains multiple nucleotide calls at regions in which the constitutive genomes differ, these can be used to identify haplotypes of a specific gene within the polyploid genomes. Haplotypes especially from low-copy genes like *CENH3* can be used to identify diploid progenitors in polyploids as well as for screening introgression lines (IL) in breeding. The use of amplified sequence drastically reduces haplotyping cost and especially for direct sequencing of PCR products compared with the long procedure of haplotyping based on cloning.
This study used the partial coding sequence of the markers centromere specific histone 3 (CENH3) and nicotinamide adenine dinucleotide dehydrogenase subunit 1 (NAD1) to establish the relationships between cultivated diploid and triploid as well as wild-type diploid progenitor bananas. On the basis of the nuclear gene CENH3, we further inferred haplotype(s) from amplified sequence of triploid and diploid bananas. The mitochondrial gene (NAD1) coding for NADH dehydrogenase subunit 1, an enzyme in metabolism of fatty acids, sugar and amino acids is generally conserved but variable enough to be used in differentiating taxonomic levels in plant molecular phylogeny (Bremer et al. 2002). The CENH3 gene is functionally conserved but is highly variable within the N-terminal tail region and exists in single or multiple copies in plants (Dunemann et al. 2014; Hirsch et al. 2009; Masonbrink et al. 2014; Wang et al. 2011). In this study, we obtained comparative phylogeny as well as representative haplotypes in bananas using the mitochondrial and nuclear genes NAD1 and CENH3 respectively.

Materials and Methods

Plant Material

In this study, a total of 40 banana genotypes/cultivars with variable genomic compositions were used (table 1). The samples included 10 genotypes of AA, 6 of AAA, 7 of AAB, 2 of AB, 6 of ABB, 1 of AS, and 5 of BB genomic composition and one tetraploid (ABBT cultivar) and the species Musa ornata and M. textilis from the sections Rhodochlamys and Australimusa, respectively. The cultivar “Ngombe” was obtained from Kenyan Agricultural and Livestock Research Organization (KALRO) and initiated as the in vitro culture at IITA-Kenya. The remaining samples were obtained from Bioversity International Musa International Transit Center (ITC) hosted in KU Leuven, Belgium (Rucas et al. 2017).

Genomic DNA Extraction

Genomic DNA was extracted from all the cultivars using CTAB (Porebski et al. 1997) except for the cultivar “Ngombe” that was extracted using Qiagen DNA extraction kit (Qiagen, Hilden, Germany).

Primer Design

Primers were designed to amplify partial coding sequence of CENH3 and NAD1 genes. Primers for CENH3 were designed based on coding sequence of “Doubled Haploid Pahang” (DH Pahang) using PrimerSelect-DNASTAR software (Madison, USA). The CENH3 primers were designed to target a 322 bp amplicon from position 6211 to 6533, which corresponds to the exon 1 and part of intron 1 obtained from sило analysis of the 7 kb genomic sequence from M. acuminata genotype “DH Pahang”.

PCR Analysis, Sequencing and Sequence Assembly

Six primer sets (three for each target region) were first optimized with eight genotypes selected from the 40 samples under study. On the basis of quality of PCR products and variability in the sequences, the best primer pair for each marker was selected and used to amplify the remaining samples. The sequences of the two primer pairs were: CENH3_gene_F2: CTCGTGTAGATGCAGACAC (6530-6548) and CENH3_gene_R2: CTGGTGCCGTGCTTCC (6212–6227) for amplification of CENH3 gene and NAD1_c5F: GTCCCGGGGCAACCAC and NAD1_c5R: GCAGTCGGGGCACAAG for amplification of NAD1 gene. PCR reactions were performed in a final volume of 20 μl reaction using the Bioneer PCR premix (Bioneer, Daejeon, South Korea) containing: 1 U of Top DNA polymerase, 250μM dNTP, 10 mM Tris–HCl (pH 9.0), 30 mM KCl and 1.5 mM MgCl2. The amplifications were done in Applied Biosystem (ABI) GeneAmp PCR machine 9,700 using the profile: Initial denaturation of 94 °C for 5 min, 40 cycles of 94 °C for 30 s, annealing temperature of 59 °C for 1 min, 72 °C extension for 40 s, and a final extension of 72 °C for 10 min for CENH3 gene, while NAD1 was amplified with an initial denaturation of 94 °C for 5 min, 40 cycles of 94 °C for 30 s, annealing temperature of at 64 °C for 30 s and extended at 72 °C for 1.5 min. The PCR products were separated on a 1.5% agarose gel stained with gelred (Biotium, CA) to confirm successful amplifications. Products with successful amplifications were cleaned using Bioneer PCR purification kit (Bioneer, Daejeon, South Korea) according to the manufacturer’s protocol.

Sequencing of PCR products was performed from both ends (Forward and reverse) using automated sequencers ABI 3130 and 3730 prism as per set protocols. Sequences were assembled and edited in Sequencher version 4.1. All ends that had low confidence base calls were trimmed to retain high confidence sequences. Sequence assembly parameters were set at 85 and 25 and assembly done using the “assemble automatically parameter”. The criteria for calling heterozygous secondary bases were set at 30% of the
primary chromatogram. Finally, IUPAC codes were assigned to all the positions that were meeting the criteria for multiple calls and sequences exported in FASTA concatenated format for further analyses.

Sequence Alignment and Phylogenetic Reconstruction

FASTA concatenated sequences were aligned in ClustalW (Thompson et al. 1994) as implemented in MEGAS (Tamura et al. 2011). Alignment of CENH3 sequences was done default Gap Opening Penalty (GOP), Gap Extension Penalty (GEP) and DNA weight matrix for both markers. Phylogeny reconstruction was performed by maximum likelihood (ML) for both markers. The two markers had different parameters in reconstruction except for the bootstrap (BP) statistical support of individual clades and the respective number of BP replicates which was set at 1,000. Model selection was performed in phyML (Guindon et al. 2010). During phylogenetic analysis for CENH3 sequences, the rates among sites were gamma distributed with a gamma parameter of 8.0, gaps with missing sequence sections were treated by

| No. | Accession No. | Accession Name          | Section | Species/Group | Subspecies/Subgroup | Country of Origin |
|-----|---------------|-------------------------|---------|---------------|----------------------|-------------------|
| 1   | ITC0046       | Monthan Eumusa ABB      |         | AB            | subgr. Monthan       | India             |
| 2   | ITC0245       | Safet Velchi Eumusa AB  |         | A             | subgr. Ney Poovan    | India             |
| 3   | ITC0247       | Honduras Eumusa balbisiana |         | n/a           |                       | Honduras          |
| 4   | ITC0420       | Pisang Kayu Eumusa AAA  |         | AB            | subgr. Orotava       | Indonesia (IDN 098) |
| 5   | ITC0472       | Pelipita Eumusa ABB     |         | A             | subgr. Philippines   |                  |
| 6   | ITC0582       | Lady Finger Eumusa AAB  |         | A             | subgr. Nadan         | India             |
| 7   | ITC0637       | Musa ornata Rhodochlamys ornata |      | subsp. ornata |                       |                  |
| 8   | ITC0659       | Khae Phrae Eumusa acuminata |         | AB            | subgr. Thai (THA 15)  |                  |
| 9   | ITC0660       | Doe Eumusa ABB          |         | A             | subgr. Bluggoe       |                  |
| 10  | ITC0679       | Namwa Khom Eumusa AB    |         | A             | subgr. Philippines   |                  |
| 11  | ITC0852       | Lady Finger Eumusa AAB  |         | A             | subgr. Nadan         | India             |
| 12  | ITC1034       | Musa ornata Rhodochlamys ornata |      | subsp. ornata |                       |                  |
| 13  | ITC1064       | Pisang Bakar Eumusa AAA |         | AB            | subgr. Amnon         | Indonesia (IDN 106) |
| 14  | ITC1102       | Tani Eumusa balbisiana  |         | n/a           |                       |                  |
| 15  | ITC1138       | Saba Eumusa AB          |         | A             | subgr. Saba          | Philippines       |
| 16  | ITC1140       | Red Yade Eumusa AAB     |         | A             | subgr. Plantain      | Cameroon          |
| 17  | ITC1152       | Wompa Eumusa ABB        |         | A             | subgr. Plantain      | Papua New Guinea (PNG 063) |
| 18  | ITC1156       | Pisang Batu Eumusa balbisiana |      | n/a           |                       | Indonesia (IDN 080) |
| 19  | ITC1177       | Zebrina Eumusa acuminata |         | AB            | subgr. zebrina       | Indonesia         |
| 20  | ITC1187       | Tomolo Eumusa AA        |         | A             | subgr. Philippines   |                  |
| 21  | ITC0393       | Truncatea Eumusa acuminata |        | subsp. truncata |                       |                  |
| 22  | ITC0804       | Mbwaizume Eumusa AAA    |         | A             | subgr. Mutika/Lujugira | Burundi          |
| 23  | ITC0652       | Klui Tparator Eumusa ABB |         | A             | n/a                  | Thailand (THA 020) |
| 24  | ITC0769       | Figue Pomme Géante Eumusa AAB |      | subgr. Silk |                       | Guadeloupe        |
| 25  | ITC1325       | Orisehe Eumusa AAA      |         | A             | subgr. Plantain      | Nigeria           |
| 26  | ITC0623       | Bankisi Eumusa acuminata |         | subsp. banksi |                       | Papua New Guinea |
| 27  | ITC0249       | Calcutta 4 Eumusa acuminata |        | subsp. burmannicoides | India             |
| 28  | ITC0653       | Pisang Mas Eumusa AA    |         | A             | subgr. Sucrier       | Malaysia          |
| 29  | ITC0962       | Prata Ana Eumusa AAB    |         | A             | subgr. Pome          | Brazil            |
| 30  | ITC1441       | Pisang Ceylan Eumusa AAB |         | A             | subgr. Mysore        | Malaysia          |
| 31  | ITC0250       | Malaccensis Eumusa acuminata |      | subsp. malaccensis |                       |                  |
| 32  | ITC0393       | Long Tavoy Eumusa acuminata |       | subsp. burmannica |                       |                  |
| 33  | ITC1511       | DH Pahang (CIRAD 930) Eumusa acuminata |     | subsp. malaccensis | Guadeloupe          |
| 34  | ITC0649       | Foonah Eumusa AAB       |         | A             | –                    | Pome/Prata        |
| 35  | ITC0769       | Ngombe Eumusa AAB       |         | A             | subgr. East African Highland Banana | Kenya |
| 36  | ITC0180       | Grande Naine Eumusa AAA |         | A             | Cavendish            | ITC               |
| 37  | ITC1588       | Lal Velchi Eumusa balbisiana |     | type 3        |                       | ITC               |
| 38  | ITC1063       | Pisang Klutuk Wulung Eumusa balbisiana |    | type 4        |                       | Indonesia (IDN 056) |
| 39  | ITC0702       | Textilis Australimusa textilis |       | subsp. textilis |                       |                  |
| 40  | ITC1238       | Yawa 2 Eumusa ABT       |         | A             | n/a                  | Papua New Guinea  |
pairwise deletion for \textit{CENH3}. Phylogenetic reconstruction for \textit{NAD1} used the model Kimura 2 parameter (KC) and the rates among sites were uniform while gaps and missing data were treated by complete deletion. Equally for \textit{NAD1} gaps in the alignment gaps were treated as missing data.

Haplotype Inference

Single nucleotide polymorphisms (SNPs) for \textit{CENH3} to which had been assigned IUPAC codes were exported to excel from MEGA5 and saved as a .csv file. The SNP positions were grouped as per the genotype’s ploidy (triploids and diploids). Triploid genotype that only had two nucleotide calls (a primary and secondary) at the target SNP site (missing a tertiary chromatogram) the IUPAC code N (A or T or G or C) was added in place of the third, indicating that the third base call may have been any of the four at that position (including homozygous type). SNPs within the diploid genotypes that only had one base call equally had an N added to complete the two base calls associated with diploidy (supplementary table 2, Supplementary Material online). The species \textit{M. ornata} and \textit{M. textilis} plus any other genotype that did not have multiple calls were assumed to be homozygous and therefore omitted from the analysis.

The software SATlotyper (Neigenfind et al. 2008) was used for haplotype inference. Both triploid and diploid SNP positions were run with the SATsolvers set as MiniSat_v1.14_cygwin. All the SNP positions having multiple calls were used for analysis with a BP value of 100. Haplotypes identified in haploids were physically checked for duplication in triploids. Haplotypes that were duplicated in diploid and triploid genotypes from the analysis.

Results

PCR Amplification of \textit{CENH3} and \textit{NAD1} Genes in Banana Genotypes

Amplification was performed on banana genotypes with different genomic compositions and different sizes of amplicons were obtained using \textit{NAD1} marker (supplementary fig. 1, Supplementary Material online). However, a single amplicon of \(\sim 322\) bp was obtained with the marker \textit{CENH3}. On the basis of \textit{NAD1} specific primers, bananas of genomic groups AAA, AA, AAB, and AB were observed to have a band size of \(\sim 1,000\) bp, whereas ABB and BB genomic groups had a band size of \(1,500\) bp. The species \textit{M. textilis} (T-genome) was observed to have the smallest band size of \(\sim 900\) bp. None of the genotype used in this study showed multiple bands on the gel. PCR products were sequenced in order to identify the dominant allele and also to check reasons behind the differences in size of \textit{NAD1} amplicons. To ensure sequence authenticity, sequencing was done bidirectionally using the same primers used for amplification (table 2). The differences in sizes of bands in amplicons from different genomic groups observed in PCR were identified to be due to insertions and deletions (indels; fig. 1).

Sectional and Genomic Group Congruence

Bidirectional sequencing and clean-up resulted in a total of 59 sequences, 36 for \textit{NAD1} and 23 for \textit{CENH3}. The data for the genotypes Pelipita, Calcutta 4, Tani, Red Yade, and Yawa 2 were treated as missing data for the \textit{NAD1} marker due to consistently poor quality of sequence. The sequences obtained were deposited in the NCBI under accession numbers KP751256–KP751292 for \textit{NAD1} and KP751293–KP751328 for \textit{CENH3}. Because \textit{CENH3} was used in haplotype inference, it was paramount to map it in the context of the available banana whole genome (fig. 2). The \textit{CENH3} gene was observed to be located within the contig \textit{NW_008990373.1} of the \textit{M. acuminata subsp. malaccensis} chromosome 8 genomic scaffold (Ma08_p09050.1; banana genome, annotation version 2) specifically within scaffold ASM31385v2_scaffold_4 (D’Hont et al. 2012) of the NCBI annotation release 101 (fig. 2).

Multiple alignments of clean sequences of \textit{NAD1} and \textit{CENH3} for all the 40 genotypes resulted in an aligned length of 1,448 bp (including indels in some cultivars) and 278 bp for the two markers respectively. Among all sequences, 196 and four SNPs were observed to be variable for \textit{NAD1} and \textit{CENH3}, respectively. In order to ensure that the observed SNP variations were genuine and not due to PCR errors, they were
further confirmed by resequencing (of independent PCR products) in the genotypes involved. The best ML model selected for the markers \textit{CENH3} and \textit{NAD1} were TN93 (supplementary table 3, Supplementary Material online) and GTR (supplementary table 4, Supplementary Material online), respectively. During model selection, the two markers were observed to have different evolutionary rates and hence alignments were conducted separately. Twenty SNP positions in \textit{NAD1} alignment were observed to differentiate A and B genomes of banana (table 3).

### Genotype Congruence Based on the Marker \textit{NAD1}

Three clades were obtained on phylogenetic reconstruction with \textit{NAD1}, and denoted as clades A, B and C (fig. 3 and supplementary data 1, Supplementary Material online). The clades were strongly supported at 75 and 100 BP values for clades A and B, respectively. The clade A was represented by triploid and diploid genotypes from different genomic groups. The triploid genotypes clustered in this clade were either AAA or AAB genomic groups (fig. 3). Diploids in this clade have AA genomic composition, AS genomic constitution represented by the genotype Wompa, and AB genotypes represented by Safet Velchi and Kunnan (fig. 3). \textit{NAD1} marker was unable to

---

**Fig. 1.**—Multiple sequence alignment of \textit{NAD1} sequences representing various genotypes and species from the different genomic groups. The genotypes included are Wompa (AS), Yawa 2 (ABBT), Namwa Khom (ABB), Pelipita (ABB), Lady Finger (AAA), Zebrina (AA), Pisang Klutuk Wulung (BB) and Ngombe (AAA), and species \textit{Musa textilis} (T-genome) and \textit{Musa omata}. The reference genomic sequence represents the partial genome region from the \textit{Musa acuminata} genome GenBank number GCAIC01015283.

**Fig. 2.**—Banana \textit{CENH3} sequence mapped in context of the \textit{Musa acuminata} reference whole genome contig sequences (WGS), reference sequences (RefSeq), and gene sequences (Gene_Seq). A-RefSeq RNA map, B-Genes_seq map, and C-WGS contig map.

---

**Table 1.**—Summary of the \textit{NAD1} allele frequency.

| Genomic | A | B | C |
|---------|---|---|---|
| AAA     | 0 | 10| 0 |
| AAB     | 10| 0 | 10|
| AA      | 0 | 0 | 0 |
| AS      | 1 | 0 | 0 |
| AB      | 0 | 1 | 0 |
| BB      | 0 | 0 | 0 |
| AAA     | 0 | 10| 0 |
| AAB     | 10| 0 | 10|
| AA      | 0 | 0 | 0 |
| AS      | 1 | 0 | 0 |
| AB      | 0 | 1 | 0 |
| BB      | 0 | 0 | 0 |
| AAA     | 0 | 10| 0 |
| AAB     | 10| 0 | 10|
| AA      | 0 | 0 | 0 |
| AS      | 1 | 0 | 0 |
| AB      | 0 | 1 | 0 |
| BB      | 0 | 0 | 0 |

---

**Table 2.**—Summary of the \textit{CENH3} allele frequency.

| Genomic | A | B | C |
|---------|---|---|---|
| AAA     | 0 | 10| 0 |
| AAB     | 10| 0 | 10|
| AA      | 0 | 0 | 0 |
| AS      | 1 | 0 | 0 |
| AB      | 0 | 1 | 0 |
| BB      | 0 | 0 | 0 |
| AAA     | 0 | 10| 0 |
| AAB     | 10| 0 | 10|
| AA      | 0 | 0 | 0 |
| AS      | 1 | 0 | 0 |
| AB      | 0 | 1 | 0 |
| BB      | 0 | 0 | 0 |
| AAA     | 0 | 10| 0 |
| AAB     | 10| 0 | 10|
| AA      | 0 | 0 | 0 |
| AS      | 1 | 0 | 0 |
| AB      | 0 | 1 | 0 |
| BB      | 0 | 0 | 0 |

---

**Table 3.**—Summary of the \textit{NAD1} SNP positions.

| Genomic | SNP Position | A | B | C |
|---------|--------------|---|---|---|
| AAA     | 022           | 0 | 10| 0 |
| AAB     | 032           | 10| 0 | 10|
| AA      | 042           | 0 | 0 | 0 |
| AS      | 052           | 1 | 0 | 0 |
| AB      | 062           | 0 | 1 | 0 |
| BB      | 072           | 0 | 0 | 0 |
| AAA     | 082           | 0 | 10| 0 |
| AAB     | 092           | 10| 0 | 10|
| AA      | 102           | 0 | 0 | 0 |
| AS      | 112           | 1 | 0 | 0 |
| AB      | 122           | 0 | 1 | 0 |
| BB      | 132           | 0 | 0 | 0 |
| AAA     | 142           | 0 | 10| 0 |
| AAB     | 152           | 10| 0 | 10|
| AA      | 162           | 0 | 0 | 0 |
| AS      | 172           | 1 | 0 | 0 |
| AB      | 182           | 0 | 1 | 0 |
| BB      | 192           | 0 | 0 | 0 |
| AAA     | 202           | 0 | 10| 0 |
| AAB     | 212           | 10| 0 | 10|
| AA      | 222           | 0 | 0 | 0 |
| AS      | 232           | 1 | 0 | 0 |
| AB      | 242           | 0 | 1 | 0 |
| BB      | 252           | 0 | 0 | 0 |

---

**Table 4.**—Summary of the \textit{CENH3} SNP positions.

| Genomic | SNP Position | A | B | C |
|---------|--------------|---|---|---|
| AAA     | 022           | 0 | 10| 0 |
| AAB     | 032           | 10| 0 | 10|
| AA      | 042           | 0 | 0 | 0 |
| AS      | 052           | 1 | 0 | 0 |
| AB      | 062           | 0 | 1 | 0 |
| BB      | 072           | 0 | 0 | 0 |
| AAA     | 082           | 0 | 10| 0 |
| AAB     | 092           | 10| 0 | 10|
| AA      | 102           | 0 | 0 | 0 |
| AS      | 112           | 1 | 0 | 0 |
| AB      | 122           | 0 | 1 | 0 |
| BB      | 132           | 0 | 0 | 0 |
| AAA     | 142           | 0 | 10| 0 |
| AAB     | 152           | 10| 0 | 10|
| AA      | 162           | 0 | 0 | 0 |
| AS      | 172           | 1 | 0 | 0 |
| AB      | 182           | 0 | 1 | 0 |
| BB      | 192           | 0 | 0 | 0 |
| AAA     | 202           | 0 | 10| 0 |
| AAB     | 212           | 10| 0 | 10|
| AA      | 222           | 0 | 0 | 0 |
| AS      | 232           | 1 | 0 | 0 |
| AB      | 242           | 0 | 1 | 0 |
| BB      | 252           | 0 | 0 | 0 |
differentiate diploids within AA genome represented by the subspecies *banksii*, *zebrina*, *malaccensis*, and *siamea*. However, the diploid AA in the subspecies *truncata* segregated within clade A albeit as a subclade.

The clade B in the NAD1 phylogeny had representation from BB and ABB genomic groups and was strongly supported at a BP support of 100. This clade had nine subclades represented by individual genotypes of both BB and ABB genomic constitution. The genotypes with ABB genome within this clade were represented by Saba, Namwa Khom, Kluai Tiparot, Monthan, and Dole (fig. 3). The genotypes with BB genomes within this clade were represented by Pisang Klutuk Wulung, Pisang Batu, Lal Velchi, and Honduras.

In the third clade C had the species *M. textilis* as the only representative of the section *Australimusa*. The species *M. ornata* of the section *Rhodochlamys* did not cluster to any of the three clades (A, B or C).

### Genotype Congruence Based on the Marker CENH3

The phylogeny based on the marker CENH3 had two major clades and associated subclades (fig. 4 and supplementary data 2, Supplementary Material online). The first clade the “A genome genotypes” had representative genotypes with A genome (AA, AB, AAA, and AAB) except for the genotype Wompa, which is an interspecific hybrid between *M. acuminata* and *M. schizocarpa* (AS genomic composition) and the species *M. ornata*. This marker was able to differentiate subspecies and genome groups within the A genome clade and these appeared in the phylogenetic tree as subclades, each having one or more genotype(s). Two subclades with more than one genotype were observed to segregate from the A genome clade. These two multigenotype subclades included the one with the genotypes Zebrina (AA), Mbwazirume (AAA), and Ngombe (AAA), which were supported at a 57 BP. The

---

### Table 3

| Genotype | Genomic Group |
|----------|---------------|
| Dole     | ABB           |
| Honduras | BB            |
| Klui Tiparot | ABB       |
| Monthan | ABB           |
| Pelipita | ABB           |
| Pisang Satu | BB          |
| Pisang Klutuk Wulung | BB |
| Tani     | BB            |
| Figue Pomme Geante | AAB |
| Lady Finger | AAB        |
| *Musa acuminata* spp. *banksii* | AA |
| *Muscotia* 4 | A |
| *Musa acuminata* spp. *malaccensis* | AA |
| *Musa acuminata* spp. *truncata* | AA |
| Wompa    | AS            |
| Zebrina  | AA            |

A, adenine; C, cytosine; G, guanine; T, thymine; greyed cells represent AA, AS, AAA, and AAB genotypes while white ones represent ABB and BB genotypes.
second subclade within this clade that had multiple genotypes was supported at 68 BP value and had the genotypes Figue Pomme Géante (AAB), Safet Velchi (AB), and Namwa Khom (ABB). This clade also had multiple single-genotype subclades with different genomic groups (fig. 4). The single genotype clades included those with the genotypes, Banksii, Pelipita, Khae Phrae, DH Pahang, Tomolo, Musa acuminate subsp. malaccensis, Pisang Ceylan, Lady Finger, Pisang Mas, Pisang Kayu, Wompa, M. ornata, Prata Ana, Orishele, Kunnan, Musa acuminate subsp. truncata, Focona, Grande Naïne, Calcutta 4, Agutay, and Long Tavoy (fig. 4).

The second clade based on CENH3 marker, “B genome clade” had genotypes from the BB, ABB, and ABBT genomic groups (fig. 4). This clade had a total of eight genotypes, out of which four were from BB, three from ABB and one with ABBT genomic composition. The BB genotypes in this subclade included the genotypes Pisang Batu, Lal Velchi, Pisang Klutuk Wulung, and Tani. The ABB genotypes in this B genome clade included Saba, Dole and Kluai Tirapot. The only tetraploid within this clade was represented by Yawa 2 (ABBT). The genotypes Honduras, Monthan, Pisang Bakar and M. textilis were treated as missing data resulting from failure to PCR amplify.

Haplotypes in Diploids and Triploids

Haplotypes were obtained by analyzing SNP positions with multiple base calls within the CENH3 gene. Seven SNP positions in 34 genotypes were used to infer haplotypes (table 4). A total of 13 haplotypes were observed based on the seven SNPs and out of these five were found to be present in diploids and triploids, three SNPs were observed in diploids only while the rest were unique to triploids. Haplotype numbers 1, 2, 3, 5 were observed to have the highest genotype representation (table 4).

The genotypes [numbers represent the order of names as listed in table 1 were Pisang Batu (18), Honduras (3), Pisang Klutuk Wulung (38), and Lal Velchi (37)] with B genome were represented within haplotype number 1 (table 4). The genotype Lady Finger (6) with the genomic composition AAB, the genotypes Dole (10), Kluai Tiparot (23), and Saba (15) from ABB as well as Kunnan (12) from AB genomic group were also observed to be associated with this haplotype. Haplotype number 2 has representative from diploids and triploids with A genome as well as interspecific triploids with A and B genome. The diploids with AA genomic group associating with haplotype 2 were represented by the genotypes Agutay (11), Banksii (26), Khae Phrae (9), DH Pahang (33), Zebrina (19), and Long Tavoy (32). The triploid AAA linking to this haplotype was represented by the genotype Pisang Kayu (4). Moreover, haplotype 2 had genotypes with AABB and ABBBB genomic composition closely linking to it at these seven SNP positions. The genotypes Figue Pomme Géante (24), Prata Ana (29), Orishele (25), Focona (34), and Lady Finger (6) were the AABB genomic group genotypes that were closely linked to haplotype 2. The genotype Pelipita (5) was the only genotype with the ABBB genome composition that had close sequence similarity to this haplotype. Haplotype number 3 had mainly A genome containing genotypes Mbwazirume (22), Tomolo (20), Wompa (17), and Ngombe (35). Haplotype number 4 was closely associated with two

---

Table 4

| Haplotype number | Haplotype Genotypes² | Genotypes⁴ |
|------------------|----------------------|-------------|
| 1                | CGAAACT⁵             | 3, 6, 10, 12, 15, 18, 23, 37, 38 |
| 2                | CGAAATC⁵             | 4, 5, 6, 9, 11, 19, 24, 25, 26, 29, 32, 33, 34 |
| 3                | CGAAATG⁵             | 17, 20, 22, 35 |
| 4                | CGATACT⁵             | 15, 38 |
| 5                | CGGAATC⁵             | 2, 20, 29, 30 |
| 6                | CGAGTT              | 17 |
| 7                | CGAGGCC             | 19 |
| 8                | GCAATC               | 2 |
| 9                | CGAGTCT              | 30 |
| 10               | GCAATT               | 29 |
| 11               | GCAATCG              | 16 |
| 12               | GGGAATC              | 8 |
| 13               | GGGAATT              | 8 |

The column genotypes indicate the genotypes that had representation within that haplotype. The genotype numbers are as listed in the last column.

²Haplotypes that were present in both diploids and triploids.

⁴Numbers in the genotype column represent the names as listed in table 1.
genotypes Pisang Klutuk Wulung (38) and Saba (15). Pisang Klutuk Wulung is a BB genomic group genotype, while the genotype Saba is of ABB genomic composition. The genotypes Pisang Ceylan (30), Tomolo (20), Prata Ana (29), and Safet Velchi (2) were the genotypes that were within haplotype 5. Haplotypes 6–13 had only one representative genotype, Wompa (17), Zebrina (19), Safet Velchi (2), Pisang Ceylan (30), Prata Ana (29), Red Yade (16), and Namwa Khom (8) for haplotypes 6–13 respectively (table 4).

Discussion

Several studies have been conducted using molecular approaches in banana to identify their origins and domestication patterns especially cultivated polyploids from their diploid progenitors (De Langhe et al. 2010; Hippolyte et al. 2012; Lescot et al. 2008). In this study, a mitochondrial and a nuclear marker were used to establish the phylogeny and specific haplotypes of cultivated triploids and a tetraploid banana to the potential diploid progenitors. The phylogenies using the dominant allele with both markers resulted in clear segregation of genotypes based on the respective genomic constitution. The major genome was observed to dominate with AAB genotypes segregating together with AA and AAA genotypes and ABB with BB genotypes. Moreover, the A genome was observed to dominate in the genotypes with AB and AS genomic composition as they were observed to segregate with the AA and AAA. It is therefore important to consider the subgenomic alleles through haplotyping.

The sizes of NAD1 amplicons were dependent on the genomic group, which was attributable to insertions and deletions (indels) within the sequences. The genotypes with BB and ABB genomes were observed to have a larger amplicon in comparison with genotypes with AA, AAB, and AAA genomic composition. This marker was able to differentiate the B cytoplasm in ABB genotypes and A cytoplasm in AAB heterogenic triploids, as clearly reflected through the band sizes of amplicons (supplementary fig. 1, Supplementary Material online).

Three different sections (Eumusa, Rhodochlamys, and Australimusa) represented in this work were differentiated using the two markers. The result that M. ornata representative of the section Rhodochlamys clustered with the A genome genotypes of the section Eumusa using the CENH3 marker is consistent with other studies (Christelová et al. 2011; Hřibová et al. 2011; Wong 2002). This observation corroborates Wong’s (2002) and Christelová et al.’s (2011) findings and suggestions that these two sections should be merged, as there are no great genetic differences that warrant their positioning into different sections.

The only representative of the AS diploid genotype “Wompa” belonging to M. schizocarpa was observed to cluster with the M. acuminata clade within the section Eumusa hence supporting the point made by Hřibová et al. (2011) that M. schizocarpa is closely related to M. acuminata. Alternatively, this clustering could also be due to the A-allele in this AS genotype being the dominant allele and therefore easily amplifiable and detectable.

It was observed that the B genome clade and the respective genotypes are distant to that of M. acuminata despite being in the same section (figs. 3 and 4). Despite being in a distinct clade, the monophyly of the B genome clade cannot be assertively concluded due to the limited number of markers used in this study. However, the species M. ornata of the section Rhodochlamys was observed to be closer to M. acuminata than it was to M. balbisiana. This observation is consistent with the observation by Christelová et al. (2011) on Musa species where B genome is shown to have diverged 27.9 Ma while M. ornata of the section Rhodochlamys diverged only 8.8 Ma. Furthermore, the results that M. ornata of the section Rhodochlamys and M. acuminata of Eumusa being similar confirm Simmonds (1962) work that showed that hybridization between species of section Eumusa and Rhodochlamys are successful and attributed to lack of any reproductive barriers between the two sections.

Surprisingly in this work, the clustering of the ABB genotypes Pelipita, Monthan, Dole, and Saba to the B genome clade is contrary to what was observed by Boonruangrod et al. (2008). Indeed, clustering of the genotypes with ABB genome was not limited to the four genotypes but to all the ABBs; genotypes Klui Tiparo and Namwa Khom were also observed to cluster to the B genome clade (figs. 3 and 4). The segregation of AAB genotypes is in accordance with the results obtained by Carrel et al. (2002), where all the AAB apart from Pisang Rajah and Pisang Kelat subgroups were found to have A genome cytoplasmic constitution. The segregation of this genomic group with NAD1 was not only observed in the sequences but also in the sizes of the amplicons. This suggests that the mtDNA region used in this study may not solely and conclusively identify the mitochondrial types hence more markers may be needed.

The segregation based on the marker CENH3 was similar in many ways to that of NAD1. The two major clades representing the A and B genomes observed in NAD1 phylogeny were also observed with CENH3, which shows consistency. The identification of A and B was derived by the observation that no diploid BB or AA genotypes clustered together, they were in either one of the two main clades for both markers (fig. 4). The major deviation to the NAD1 phylogeny was mainly in the A genome clade where this marker was unable to differentiate the subspecies (fig. 4). The marker CENH3 is a nuclear based marker and hence the clustering of the A genome subspecies was expected.

The segregation of the A genome in subspecies malaccensis and truncata in this study compares favorably with the one observed by Wong (2002) where the two subspecies clustered separately. However, this observation deviates from the placement of the subspecies truncata as a synonym of the subspecies malaccensis by Hotta (1989). This study clearly
differentiated the two subspecies, with malaccensis being supported as a separate clade from truncata (fig. 4). Despite the limited number of markers used in this study, our study confirms the closeness of the subspecies burmannica (Long Tavoy) and burmannicoides (Calcutta 4) similar to observations made by Ude et al. (2002) and later confirmed and proposed to be joined as one (Sardos et al. 2016; fig. 4).

The segregation within the B genome clade was consistent in both markers used in this study and is similar in some aspect to observations made by Ude et al. (2002), which observed that ABB genotypes were closer to BB diploids. In this study using both markers, the AAB and ABB genotypes were observed to segregate into either A or B genome clades. Similarly, Ude et al. (2002) observed a close clustering between the triploid AAB and ABB with the diploid AA and BB, respectively. However, AAB and ABB genomic groups were observed to have A and B as well as B and A haplotypes respectively (table 3).

Different studies have indicated that only M. acuminata subspecies that originated from the islands of Southeast Asia (ISEA) have had genomic contribution to cultivated bananas (Li et al. 2013; Perrier et al. 2011). In this study, both phylogeny and haplotype inference indicated that ISEA subspecies contributed to triploid cultivated bananas (figs. 3 and 4). Phylogeny using both NAD1 and CENH3 was able to establish one major clade (Clade A) that had ISEA subspecies zebrina, errans, banksii, truncata, and malaccensis clustering. Moreover, the EAHB have been shown to have had genomic contribution from ISEA subspecies zebrina and banksii (Perrier et al. 2011). In this study, based on CENH3 marker, EAHB genotypes Ngombe and Mbwazirmwe were observed to cluster within clade A to major ISEA subspecies (figs. 3 and 4). These observations further attest to the ISEA subspecies contributing to EAHB. Observation of the diploid genotype Tomolo clustering with the wild type genotype M. acuminata subsp. malaccensis (fig. 4) was contrary to other observations made that have linked this cooking diploid cultivar to wild type subspecies banksii (Carreel et al. 2002; De Langhe et al. 2010; Li et al. 2013). However, this deviation was observed only with the phylogeny based on CENH3 marker. This deviation can be explained by the fact that CENH3 marker is a nuclear gene, while NAD1 marker is of mitochondrial nature and the latter has uniparental inheritance.

The maternal and paternal nature of inheritance of chloroplast (cpDNA) and mitochondrial (mtDNA) genetic material in bananas (Fauré et al. 1993; Heslop-Harrison and Schwarzacher 2007) can facilitate the identification of the wild genotypes that contributed to banana cultivars. To infer haplotypes from amplified sequences, only the marker CENH3 was used, due to the fact that extranuclear DNA (NAD1) in bananas is monoparentally inherited and no crossing over in the case of NAD1 may have taken place during breeding. Furthermore, no multiple peaks were observed in the sequences obtained from triploid and diploid genotypes based on NAD1.

A total of 18 haplotypes (denoted by the number 1–18) were observed in this study. Haplotypes represented in a cultivar have been used as an indicator of the species that contributed to the cultivar and thus indicating the domestication patterns (Boonruangrod et al. 2008; Li et al. 2013). In our study, Haplotype 1 mainly consisted of B genome derived haplotypes (table 4). The two haplotypes for the genotype Kunnan (AB) were found to closely associate with haplotype 1, suggesting that indeed the B genome diploids contributed to these genotypes. Our results do not support the results obtained by Li et al. (2013), which suggested that this genotype should be renamed to AA. This observation however raises the question of why the A genome in the AB genotype was not identified in the analysis. The triploid genotypes Lady Finger (AABB), Saba (ABBB), and Dole (ABB) were observed to have a haplotype similar to haplotype 1 which is a B genome haplotype (table 4). This indicates that B genome diploids contributed to the genomes of triploids and is in line with observations made from other studies (Li et al. 2013; Perrier et al. 2011). However, this study could not identify which BB diploid specifically contributed to the B genome in triploids. The clustering of two Lady Finger haplotypes to haplotype 2, which is essentially an A genome haplotype group, indicates that both AA haplotypes in the AABB haplotype of Lady Finger may have originated from the same M. acuminata subspecies. The two haplotypes of the genotype Wompa were related to Haplotypes 3 and 6. Haplotypes 3 had other genotypes of A genomic composition but no diploid wild type genotypes were in this haplotype. Two of the four genotypes that were together with Wompa in this haplotype were EAHB genotypes Ngombe and Mbwazirmwe. The other genotype with a haplotype represented in this group was the diploid genotype Tomolo (table 4). This indicates that Tomolo may have contributed its genome to EAHB genotypes. The approach used here provides a quick and easy approach to identify haplotypes and constitutive subgenomic polymorphisms.

In this study, no single genotype (either homogenomic or heterogenomic triploid) was observed to have mtDNA haplotypes that were obtained with the marker NAD1. This observation can be explained by the uniparental nature of mitochondrial inheritance in banana where both mitochondria and chloroplast DNA have previously been used to identify the origins of genomes in diploid, homo- and heterogenomic banana genotypes (Boonruangrod et al. 2008; Carreel et al. 2002). In the present study, direct sequencing of PCR products allowed to differentiate easily the maternal mtDNA types. We observed that multiple CENH3 alleles may exist in bananas similar to observations made in other crops.

Acknowledgments

This work was supported by the National Science Foundation (NSF) and the Bill & Melinda Gates Foundation (BMGF) under
the program Basic Research Enabling Agricultural Development (BREAD) as part of the project “Fast Breeding for Slow Cycling Crops: Doubled Haploids in Cassava and Banana/Plantain” [award # 1109882]. The authors wish to thank Dr Jim Lorenzen (BMGF, USA) and Anna Oliveira (ITIA, Nigeria) for critically reviewing the manuscript and Dr Neigenfind (Max Planck Institute of Molecular Plant Physiology, Germany) for invaluable suggestions with the software SATlOTyper and haplotype inference.

**Supplementary Material**

Supplementary data are available at Genome Biology and Evolution online.

**Literature Cited**

Argent GCG. 1976. The wild bananas of Papua New-Guinea. Notes R Botan Gard Edinburg. 35:77–114.

Baumfeil F, et al. 2010. Mechanisms of haplotype divergence at the RGA09 nucleotide-binding leucine-rich repeat gene locus in wild banana (Musa balbisiana). BMC Plant Biol. 10:149.

Boonruangrod R, Desai D, Fluch S, Berenyi M, Burg K. 2008. Identification of cytoplasmic ancestor gene-pools of Musa acuminata Colla and Musa balbisiana Colla and their hybrids by chloroplast and mitochondrial haplotyping. Theor Appl Genet. 118(1):43–55.

Bremer K, Anderberg AA, Mari K. 2002. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. Mol Phylogenet Evol. 24(2):274–301.

Browning SR, Browning BL. 2011. Haplotype phasing: existing methods and new developments. Nat Rev Genet. 12(10):703–714.

Cameel F, et al. 2002. Ascertaining maternal and paternal lineage within Musa by chloroplast and mitochondrial DNA RFLP analyses. Genome 45(4):679–692.

Cheesman EE. 1950. The classification of the bananas. Kew Bull. 5(27-31):51–155.

Ching A, et al. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet. 3:19.

De Langhe E, Hribová E, Carpentier S, Dolezel J, Swennen R. 2010. Did Colla and their hybrids by chloroplast and mitochondrial DNA RFLP analyses. Genome 109(5):937–951.

Ding C, Cantor CR. 2003. Direct molecular haplotyping of long-range genomic DNA with M1-PCR. Proc Natl Acad Sci U S A. 100(13):7449–7453.

Dunemann F, Schrader O, Budahn H, Houben A. 2014. Characterization of centromeric histone H3 (CENH3) variants in cultivated and wild carrots (Daucus carota). PLoS ONE. 9(6):e98504.

Efros A, Halperin E. 2012. Haplotype reconstruction using perfect phylogeny and sequence data. BMC Bioinformatics. 13(Suppl 6):53.

Fauré S, et al. 1993. A molecular marker-based linkage map of diploid bananas (Musa acuminata). Theor Appl Genet. 87:517–526.

Guo Y-P, Wang S-Z, Vogl C, Ehrendorfer F. 2012. Nuclear and plastid haplotypes suggest rapid diploid and polyploid speciation in the N Hemisphere Achillia millefolium complex (Asteraceae). BMC Evol Biol. 12(1):2.

Häkkinen M. 2013. Reappraisal of sectional taxonomy in Musa (Musaceae). Taxon 62(4):809–813.

Heslop-Harrison JS, Schwarzacher T. 2007. Domestication, genomics and the future for banana. Ann Bot. 100(5):1073–1084.

Hippolyte J, et al. 2012. Foundation characteristics of edible Musa triploids revealed from allelic distribution of SSR. Ann Bot. 109(5):937–951.

Hirsch CD, Wu Y, Yan H, Jiang J. 2009. Lineage-specific adaptive evolution of the centromeric protein CENH3 in diploid and allotetraploid Oryza species. Mol Biol Evol. 26(12):2877–2885.

D’Hont A, et al. 2012. ‘The banana (Musa acuminata) genome and the evolution of monocotyledonous plants’. Nature 488:213–217.

Hotta M. 1989. Identification list of Ensete and Musa (Musaceae) in SE Asia and West Malesia. In: Hotta M, editor. Diversity and plant-animal interaction in equatorial rainforests. Japan: Kagoshima University, 67–75.

Hřibová E, et al. 2011. The ITS1-5.8S-ITS2 sequence region in the Musaceae: structure, diversity and use in molecular phylogeny. PLoS One. 6(1):e17863.

Kaur S, Franki MG, Forster JW. 2012. Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. Plant Biotechnol J. 10(2):125–138.

Lescot M, et al. 2008. Insights into the Musa genome: syntenic relationships to rice and between Musa species. BMC Genomics. 9:58.

Li LF, et al. 2013. Origins and domestication of cultivated banana inferred from chloroplast and nuclear genes. PLoS ONE. 8(11):e80502.

Lorenzen J, et al. 2010. Overview of banana and plantain (Musa spp.) improvement in Africa: past and future. Acta Hortic. 879:595–603.

Masonbrink RE, et al. 2014. CentH3 evolution in diploids and polyploids of three angiosperm genera. BMC Plant Biol. 14:383.

Neigenfind J, et al. 2008. Haplotype inference from unphased SNP data in heterozygous polyploids based on SAT. BMC Genomics. 9:356.

Perrier X, et al. 2011. Multidisciplinary perspectives on banana (Musa spp.) domestication. Proc Natl Acad Sci U S A. 108:11311–11318.

Porebski S, Bailey LG, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Rep. 15(1):9–15.

Ruas M, Guignon V, Sempere G, Sardos J, Hueber Y. 2017. ‘Database tool MGIS: managing banana (Musa spp.) genetic resources information and high-throughput genotyping data’. Database (Oxford) 2017:1–12.

Sardos J, et al. 2016. DAriT whole genome profiling provides insights on the evolution and taxonomy of edible Banana (Musa spp.’). Adv Genet 118(7):1269–1278.

Simmonds NW. 1962. The Evolution of Bananas. London, UK: Longman. Simmonds NW, Shepherd K. 1955. The taxonomy and origins of the cultivated bananas. J Linn Soc Lond Bot. 55:302–312.

Surridge AKJ, Vlijoen A, Crous PW, Wehner FC. 2003. Identification of the pathogen associated with Sigatoka disease of banana in South Africa. Aust Plant Pathol. 32(1):27–31.

Tamura K, et al. 2011. MEAGAS: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Tripathi L, et al. 2009. Xanthomonas wilt: a threat to banana production in East Africa. Plant Dis. 93(5):440–451.
Ude G, Pillay M, Nwakanma D, Tenkouano A. 2002. Genetic diversity in Musa acuminata Colla and Musa balbisiana Colla and some of their natural hybrids using AFLP Markers. Theor Appl Genet. 104(8):1246–1252.

Wang G, et al. 2011. Characterization of CENH3 proteins and centromere-associated DNA sequences in diploid and allotetraploid Brassica species. Chromosoma 120(4):353–365.

Wong C. 2002. Assessment of the Validity of the Sections in Musa (Musaceae) using AFLP. Ann Bot. 90(2):231–238.

Associate editor: Aoife McLysaght