Author’s post-print of the paper published in the Proceedings of the International ISHS-ProMusa Symposium on Global Perspectives on Asian Challenges, Acta Horticulturae 897, ISHS 2011. Available from www.musalit.org

Molecular Analysis Reveals Multiple Domestications of Edible Bananas

H. Volkaert

BIOTEC - National Science and Technology Development Agency, Thailand Science Park, KhlongLuang, PathumThanee 12120, Center of Excellence in Agricultural Biotechnology, Postgraduate Education and Research Development Office, Commission on Higher Education, Ministry of Education, and Center for Agricultural Biotechnology, Kasetsart University Kamphaengsaen Campus, NakornPathom 73140 Thailand

Keywords: *Musa acuminata*, *Musa balbisiana*, phylogeography, single-copy gene, SNP.

Abstract

Molecular sequence analysis of four genes in a set of 100 cultivated and wild (*Musa acuminata* and *M. balbisiana*) banana accessions has been done to determine the origin of edible bananas and plantains. A few *Musa schizocarpa* and Australimusa accessions were included as outgroups. The domestication of edible bananas involves three taxonomic groups: *M. acuminata* ssp. *banksii* / *errans*, *M. acuminata* ssp. *malaccensis* / *microcarpa* / *zebrina* / *burmanica* / *siamea* group and *M. balbisiana*. Plantains (AAB) and several ABB bananas most likely originated through a hybridisation event between *M. acuminata* ssp. *errans* (or more generally the *M. acuminata* ssp. *banksii* group) and *M. balbisiana*. Most AAB bananas such as ‘Mysore’ and some BBA bananas such as ‘Pisang Awak’ are the result of hybridisation between *M. acuminata* ssp. “non-banksii” and *M. balbisiana*. Most edible sweet bananas (AA and AAA dessert) are derived from hybridisations between subspecies within the *M. acuminata* “non-banksii” group. Several unique SNPs have been identified in these bananas and East African highland bananas (AAA) that so far have not been found in any of the wild *M. acuminata* accessions in the International Transit Center (ITC) collection. A search for these SNPs in the wild *Musa* populations would shed light on the original location of the hybridising populations that gave rise to the edible bananas. The variable contribution of parents at different genetic loci indicates that most edible bananas are not direct hybrids, but have gone through a few or several generations of backcrossing. Some diversity has been found in *M. balbisiana*. The *M. balbisiana* involved in the origin of plantains, other AAB bananas and ABB ‘Monthan’ group is distinct from the *M. balbisiana* involved in the origin of BBA ‘Pisang Awak’ group. A thorough study of the genetic diversity within *M. balbisiana* throughout its area of distribution is deemed necessary. The implications for breeding of edible bananas are discussed.

INTRODUCTION

Bananas and plantains are a staple food for millions of people in Africa and South Asia and a very important source of income for many tropical developing countries. Several diseases have severely impacted banana cultivation in the past, and Mycosphaerella leaf spot diseases, nematodes and new strains of Fusarium are a constant threat leading to intensive use of fungicides and pesticides. Banana breeding aims to provide new cultivars that show better yields or resistance traits but it is complicated because of sterility or low fertility, triploidy and the unknown origin of the cultivars. Banana and plantain cultivars are diploid or triploid unseeded clones derived from wild

diploid fertile ancestors originating from South and Southeast Asia, mostly belonging to two species of the *Musa* genus, *Musa acuminata* and *Musa balbisiana*.

Bananas have been classified into broad genomic groups on the basis of agromorphological characters (Simmonds and Shepherd, 1955). The use of molecular markers such as isozymes (Jarret and Litz, 1986), chloroplast DNA variants (Carreel et al., 2002), AFLPs and RAPDs (Howell et al., 1994; Pillay et al., 2001; Wong et al., 2001; 2002), RFLPs (Jarret et al., 1992; Raboin et al., 2005), microsatellites (Lagoda et al., 1998a, b; Grapin et al., 1998; Creste et al., 2003; 2004) and chromosome hybridisation techniques (Osuji et al., 1997; 1998; Dolezelova et al., 1998; D'Hont et al., 2000; Valarik et al., 2002) have confirmed the general groupings, but have not been able to resolve the origin of cultivated bananas in detail. The aim of the present study was to trace the wild diploid ancestors of the cultivated bananas, especially the A-B hybrid starchy bananas. This was performed through the systematic comparison of the DNA sequence of nuclear gene haplotypes of the studied cultivars with that of a sample of diploid wild accessions present in the international *Musa* germplasm collection supplemented with some recently collected wild plants from Thailand for which the geographical origin is known.

MATERIALS AND METHODS

Dried leaf tissues were obtained from the *Musa* germplasm collection of Bioversity's International Transit Center hosted by the Katholieke Universiteit Leuven, Belgium. The obtained samples included several accessions of wild *M. acuminata* ssp. *banksii*, *malaccensis*, *zebrina*, *burmanica*, *M. balbisiana* and *M. schizocarpa*, as well as several cultivars of the AAB, ABB, BBA, AAA and AA genome groups. Some samples from local wild and cultivated accessions were added (Table 1).

Initially, fragments corresponding to genes were amplified from *M. acuminata* and *M. balbisiana* using PCR primers targeting conserved regions as revealed by multiple sequence alignments of plant alcohol dehydrogenase (ADH), glyceraldehyde-3-phosphate dehydrogenase (G3PDH), isocitrate dehydrogenase (IDH), catalase (CAT), granule-bound starch synthase (GBSS) and auxin response factors 6,8 (ARF). The fragments were cloned and sequenced. Based on the obtained sequence information, primers for the PCR amplification of single loci and one or two internal primers for sequencing were designed (Table 2). The primers for PCR were located in exons, and the obtained fragments contained one or more introns.

For the survey of *Musa* diversity, genomic DNA was amplified using the specific primers in a PCR reaction containing low nucleotide (80 pM) and primer (150 pM) concentrations. The PCR amplifications were checked on agarose gel, and if a clear single band was obtained, the fragments were sent for sequencing (1stBase, Malaysia, http://www.base-asia.com) without prior purification.

Haplotypes were reconstructed from the diploid genotypic sequences using the PHASE algorithm (Stephens et al., 2001) as implemented in DnaSP v. 5 (Librado and Rozas, 2009). For sequences containing indels, haplotypes beyond the indel site were inferred using Indelligent v.1.2 (http://ctap.inhs.uiuc.edu/dmitriev/indel.asp). The haplotypes for triploid genotypic sequences were manually inferred using previously identified haplotypes as a guide, minimising inference of additional haplotypes. From some accessions, PCR-amplified products were cloned and sequenced to confirm the ordering of alleles at SNP sites into full-length haplotypes. For these PCRs, Phusion® (Finnzyme, Finland) enzyme and reaction conditions were used to minimise PCR recombination (Cronn et al., 2002) and substitution errors.

Haplotype networks were reconstructed using the maximum parsimony approach implemented in TCS v.2.1 (Clement et al., 2000) and the median joining network approach implemented in NETWORK v.4.5 (Bandelt et al., 1999).

RESULTS AND DISCUSSION

The primers amplified single fragments as confirmed by SSCP and sequencing analysis. For each locus, two to eight alleles were found within *M. balbisiana*, while around 30 were found in *M. acuminata* accessions.

Within cultivated bananas, the A, B and S genomes could readily be distinguished. However, the A-B genome composition was found to be inconsistent between loci for several cultivars. The IDH1 locus showed a clear deficit of *M. acuminata* type genomes, with a smaller deficit for the CAT1 locus. For the GBSS1 and ARF locus, slightly fewer B-genome copies were found than anticipated, such as the absence of a B haplotype for GBSS1 in the Pacific plantain 'Luba'.

From the inferred haplotype sequences, a phylogenetic network could be reconstructed, as shown for the *M. acuminata* CAT1 locus (Fig 1a) and the *M. balbisiana* IDH1 sequences (Fig. 1b). The two *M. acuminata* ssp. *truncata* accessions were readily distinguished from all others while only a single SNP separated all *M. acuminata* ssp. *banksii* and *M. acuminata* ssp. *errans* accessions. However, several SNP positions revealed clear allele frequency differences between the groups. *Musa acuminata* ssp. *banksii* / *errans*, *M. acuminata* ssp. *malaccensis* and *M. acuminata* ssp. *zebrina* contributions were found in the diploid and triploid hybrids in various combinations.

The present analysis shows that it is possible to identify the genotype of the original wild *Musa* ancestors that have contributed to the genesis of cultivated bananas. As indicated by several other studies, there is a much larger genetic diversity within the *M. acuminata* complex compared to *M. balbisiana*, though there is sufficient diversity in *M. balbisiana* to contribute to the phylogeographic study of edible bananas and plantains.

The fact that the genome contributions are not in equal amounts for the different loci indicates that, after an initial hybridisation, one or more generations of backcrossing or interbreeding have happened before sterility and parthenocarpy fixed the genotypes that survived through vegetative propagation. This implies that it would be possible to obtain new edible bananas by crossing wild ancestors for a number of generations. Molecular markers for the genes involved in the sterility and parthenocarpy could increase the efficiency of selection by selecting for progenies that contain the desirable alleles. To identify these markers, populations segregating for them need to be found in the wild or established from the wild populations closest to the ancestors. To find these relatives, access to more germplasm from wild populations is needed together with better information about their geographical origin. Extensive surveys of the wild *Musa* populations needs to be done to locate the progenitors of the cultivated bananas.

In this research, an attempt was made to identify the geographical regions based on the limited information available on the origin of the accessions in the ITC collection. To somehow try to confirm this limited information, some wild accessions from Thailand were included. 'Kluai Namwa' (Pisang Awak group) has a combination of *M. balbisiana* chloroplast and nuclear genes similar to 'Tani' or 'Singapuri' and a *M. acuminata* ssp. *malaccensis* genome contribution. The nuclear genotype of the two Thai wild *M. balbisiana* accessions is different, and thus the group of Pisang Awak cultivars most likely did not originate from Thailand. 'Monthan' (ABB) has a unique *M. balbisiana* genotype with *M. acuminata* ssp. *banksii* group A genome.

African and Pacific plantains (both AAB) have a *M. acuminata* ssp. *banksii* A genome contribution together with B genome similar to the ITC accession 'Cameroun'. The single *M. acuminata* ssp. *errans* accession included in the present study matches the A genome of the plantains closest, leading to a probable origin of these bananas in the Philippines, though information on possibly related wild ancestors from other regions such as the Indonesian Island of Sulawesi is completely lacking.

Some of the edible AA and AAA bananas, notably 'Pisang Mas', 'Petite Naine' and 'Leite' show combinations of related *M. acuminata* ssp. *malaccensis* and *M. acuminata* ssp. *zebrina* genomes. The areas where these two subspecies presumably occur together would be Sumatra, Java and other islands of Western Indonesia, though an origin from peninsular Malaysia cannot be ruled out.

The four East African highland banana accessions (EAHB, AAA genome) studied had identical haplotype combinations for all the loci studied, except one single nucleotide substitution, confirming their very close relationships. However, the particular haplotypes found in the EAHBs were not found in any of the accessions studied, though they are closely related to the *M. acuminata* ssp. *zebrina* haplotypes.

ACKNOWLEDGMENTS

I wish to thank my mentors in banana diversity, Dr. E. De Langhe, Dr. R. Swennen and Dr. I. Buddenhagen for their moral support. My colleagues Dr. Sasivimon Swaengpol, Dr. Det Wattanachaiyingcharoen, Dr. Tosak Selanan contributed plant materials and suggestions. Thanks to my students who helped with the lab work at various points: Piyaporn Phansak, Poom Predakul, Siriporn ChuayJaeng, and Sukanya Jeennor. Part of this research was funded through a grant from the Biodiversity Research and Training programme.

Cited Literature

Bandelt, H.J., Forster, P. and Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37-48.

Carreel, F., Gonzalez de Leon, D., Lagoda, P.J.L., Lanaud, C., Jenny, C., Horry, J.P. and Tezenas du Montcel, H. 2002. Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. Genome 45:679-692.

Clement, M., Posada, D. and Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657-1660.

Creste, S., Tulmann-Neto, A., de Olivera Silva, S. and Figueira, A. 2003. Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. Euphytica 132:259-268.

Creste, S., Tulmann-Neto, A., Vencovsky, R., de Oliveira Silva, S. and Figueira, A. 2004. Genetic diversity of *Musa* diploid and triploid accessions from the Brazilian banana breeding program estimated by microsatellite markers. Genet. Res. Crop Evol. 51:723-733.

Cronn, R., Cedroni, M., Haselkorn, T., Grover, C. and Wendel, J.F. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. 104:482-489.

D'Hont, A., Paget-Goy, A., Escoute, J. and Carreel, F. 2000. The interspecific genomic structure of cultivated banana, *Musa* spp., revealed by genomic DNA in situ hybridization. Theor. Appl. Genet. 100:177-183.

Dolezelova, M., Valarik, M., Swennen, R., Horry, J.P. and Dolezel, J. 1998. Physical mapping of the 18S–25S and 5S ribosomal RNA genes in diploid bananas. Biologia Plantarum 41:497-505.

Grapin, A., Noyer, J.L., Carreel, F., Dambier, D., Baurens, F.C., Lanaud, C. and Lagoda, P. 1998. Diploid *Musa acuminata* genetic diversity assayed with sequenced-tagged microsatellite sites. Electrophoresis 19:1374-1380.

Howell, E., Newbury, J., Swennen, R., Withers, L. and Ford-Lloyd, B. 1994. The use of RAPD for identifying and classifying *Musa* germplasm. Genome 37:328-332.

Jarret, R.L., Gawel, N., Whittemore, A. and Sharrock, S. 1992. RFLP based phylogeny of *Musa* species in Papua New Guinea. Theor. Appl. Genet. 84:579-584.

Jarret, R.L. and Litz, R.E. 1986. Enzyme polymorphism in *Musa acuminata* Colla. J. Heredity 77:183-186.

Lagoda, P.J.L., Dambier, D., Grapin, A., Baurens, F.C., Lanaud, C. and Noyer, J.L. 1998a. Nonradioactive sequence-tagged microsatellite site analyses: a method transferable to the tropics. Electrophoresis 19:152-157.

Lagoda, P.J.L., Noyer, J.L., Dambier, D., Baurens, F.C., Grapin, A. and Lanaud, C. 1998b. Sequence tagged microsatellite site (STMS) markers in the Musaceae. Mol. Ecol. 7:657-666.

Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Osuji, J., Crouch, J., Harrisson, G. and Heslop-Harrisson, J. 1998. Molecular cytogenetics of *Musa* species, cultivars and hybrids: location of 18S–5.8S and 5S rDNA and telomere-like sequences. Ann. Bot. 82:243-248.

Osuji, J., Harrisson, G., Crouch, J. and Heslop-Harrison, J.S. 1997. Identification of the genomic constitution of *Musa* L. lines (bananas, plantains and hybrids) using molecular cytogenetics. Ann. Bot. 80:787-793.

Pillay, M., Ogundiwin, E., Nwakanma, D.C., Ude, G. and Tenkouanao, A. 2001. Analysis of genetic diversity and relationships in East African banana germplasm. Theor. Appl. Genet. 102:965-970.

Raboin, L., Carreel. F., Noyer, J., Baurens, F., Horry, J., Bakry, F., Tezenas Du Montcel, H., Ganry, J., Lanaud, C. and Lagoda, P. 2005. Diploid ancestors of triploid export banana cultivars: molecular identification of 2n restitution gamete donors and n gamete donors. Mol. Breeding 16:333-341.

Simmonds, N.W. and Shepherd, K. 1955. The taxonomy and origins of the cultivated bananas. J. Linn. Soc. London (Bot) 5:302-312.

Stephens, M., Smith, N. and Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. Amer. J. Human Genet. 68:978-989.

Valarik, M., Simkova, H., Safar, J., Dolezelova, M. and Dolezel, J. 2002. Isolation, characterization, localization of repetitive DNA sequences in bananas (*Musa* spp.). Chromosome Res. 10:89-100.

Wong, C., Kiew, R., Argent, G., Set, O., Kong Lee, S. and Yuen Gan, Y. 2002. Assessment of the validity of the sections in *Musa* (Musaceae) using AFLP. Ann. Bot. 90:231-238.

Wong, C., Kiew, R., Phang Loh, J., Huat Gan, L., Set, O., Kong Lee, S., Lum, S. and Yuen Gan, Y. 2001. Genetic diversity of the wild banana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. Ann. Bot. 88:1017-1025.

Tables

Table 1. List of *Musa* accessions analysed and tentative haplotype assignments.

ITC#	Name	Genome / Species	Subspecies / CV Group	CAT1	GBSS1	IDH1	ARF
0649	Foconah	AAB	Pome	A_z1+A_m15+B	A1+A2+B2	A1+A2+B5	A2+A13+B
1441	Pisang Ceylan	AAB	Mysore	A_z11+A_m14+B	A1+A2+B2	A6+A2+B4	A2+A9+B
0450	Pisang Palembang	AAB	Pisang Kelat	A_m14+A_m14+B	A1+A2+B2	A1+A1+B1	A2+A11+B
0769	Figue Pomme Geante	AAB	Silk	A_m14+A_m14+B	A1+A2+B2	A1+A1+B1	A2+A11+B
0186	Lysoka	AAB	plantain	A23+A23+B	A1+A2+B2	B1+B1+B1	A13+A13+B
0109	Obino L'ewai	AAB	plantain	A22+A22+B	A1+A2+B2	B1+B1+B1	A13+A13+B
1325	Orishele	AAB	plantain	A23+A23+B	A1+A2+B2	B1+B1+B1	A13+A13+B
0098	Baka	AAB	plantain	A23+A23+B	A1+A2+B2	B1+B1+B1	A13+A13+B
0335	Popo-ulu	AAB	Pacific plantain	A_b24+A_b24+B	A1+A2+B2	B1+B1+B1	A11+A11+B
0802	Luba	AAB	Pacific plantain	A_b25+A_b25+B	A1+A1+A1	A1+A1+B3	A11+A13+B
0843	Pisang Raja Bulu	AAB	Pisang Raja	A_z4+A_b25+B	A1+A2+B2	A2+A2+B3	A5+A11+B
0767	Dole	ABB	Bluggoe	B1+B1+B1	A+B2+B3	B6+B7	A11+B+B
1483	Monthan	ABB	Monthan	A_b24+B1+B2	A+B2+B3	B6+B7	A11+B1+B1
0472	Pelipita	ABB	Pelipita	A_b24+B1+B2	A+B1+B2	B2+B2+B7	A11+B1+B1
0123	Simili Rajah	ABB	Peyan	A_b24+B1+B2	A+B1+B1	B1+B5	ND
0020	Icecream	ABB	Ney Mannan	A_b24+B1+B2	A+B1+B2	B5+B7	A11+B1+B1
0659	NamWaKhom	BBA	Pisang Awak	A_m14+B1+B2	A+B1+B1	A1+B7+B7	A2+B2+B3
local	Teeb-Loei	BBA		B1+B2+B2	A+B2+B2	B3+B3+B3	A+B1+B3
0987	Auko	AB		A_b24+B	A+B2	B3+B3	A11+B1
0245	Safet Velchi	AB		A_m14+B	A+B1	A1+B7	A11+B1
local	Loei	*M. balbisiana*	-	B1+B1	B1+B1	B2+B7	B1+B1
local	MaeRamat	*M. balbisiana*	-	B1+B2	B2+B2	B1+B6	B1+B3
1120	Tani	*M. balbisiana*	-	B1+B1	B1+B1	B2+B7	B1+B1
0271	Eti Kehel	*M. balbisiana*	-	B1+B1	B1+B1	B7+B7	B2+B3
0248	Singapuri	*M. balbisiana*	-	B1+B2	B1+B1	B6+B7	B2+B3
0247	Honduras	*M. balbisiana*	-	B1+B1	B2+B2	B6+B6	B1+B1
0246	Cameroun	*M. balbisiana*	-	B1+B1	B1+B2	B2+B6	B1+B1
1074	Butuhan	*M. balbisiana*	-	B1+B2	B1+B1	B6+B6	B1+B1
local	Pongla	BB soft seed	-	B1+B1	B1+B1	B2+B7	B1+B1

ITC#	Name	Genome / Species	Subspecies / CV Group	CAT1	GBSS1	IDH1	ARF
0084	Mbwazirume	AAA	EAHB	$A_z8+A_z8+A_z8$	A1+A2+A3	A1+A2	A3+A11+A13
0082	Intokatoke	AAA	EAHB	$A_z8+A_z8+A_z8$	A1+A2+A3	A1+A2	A3+A11+A13
0081	Igisahira Gisanzwe	AAA	EAHB	$A_z8+A_z8+A_z8$	A1+A2+A3	A1+A2	A3+A11+A13
0083	Igitsiri	AAA	EAHB	$A_z8+A_z8+A_z8$	A1+A2+A3	A1+A2	A3+A11+A13
0673	Kluai Sa	AAcv		A_m14+A_z4	A1+A2	A2+A3	A5+A11
0663	KhaiNaiOn	AAcv		A_m14+A_z4	A1+A2	A2+A3	A5+A11
local	LebMuNang	AAcv		A_m14+A_z4	A1+A2	A1+A2	A5+A11
0689	Pisang Bangkahulu	AAcv		A_m14+A_z4	A1+A2	A2+A3	A5+A11
0653	Pisang mas	AAcv	Pisang Mas	A_z7+A22	A1+A2	A1+A2	A5+A5
0575	Red Dacca	AAA	Red	$A_z3+A_z6+A_m19$	A1+A2+A3	A1+A3+A3	ND
0277	Leite	AAA	Rio	$A_z3+A_z6+A_m18$	A1+A2+A3	A1+A3+A3	A2+A11+A11
0654	Petite Naine	AAA	Cavendish	$A_z3+A_z6+A_m14$	A1+A2+A3	A1+A2+A2	A2+A5+A13
1122	Gros Michel	AAA	Gros Michel	$A_z3+A_z6+A_m14$	A1+A2+A3	A1+A2+A2	A2+A5+A13
0345	Pisang Berangan	AAA		Az5 + Ab23 + Ab23’	A1+A2+A3	A1+A2+A2	A4+A5+A13
1123	Yangambi km5	AAA		$A_z4+A_m17+A_m21$	A1+A2+A3	A2+A3+A3	ND
0312	Pisang Yari Buaya	AA / AAA	Pisang Yari Buaya	$A_b26+A27$	A1+A1+A2	A2+A3	A_z3+A15
0279	Bie Yeng	AA		A_m14+A_b25	A1+A2	A2+A2	A_z5+A_b11
0249	Calcutta4	*M. acuminata*	*burmanica*	A_z13+A_z13	A1+A2	A3+A3	A7+A7
0072	Tavoy	*M. acuminata*	*burmanica*	A_z1+A_z2	A1+A2	A1+A1	A3+A6
0093	Long Tavoy	*M. acuminata*	*burmanica*	A_z13+A_z13	A1+A2	A3+A3	A7+A7
0660	KhaePhrae	*M. acuminata*	*siamea*	A_z12+A_z12	A1+A1	A3+A3	A6+A6
local	Khlung	*M. acuminata*	*siamea*	A_z10+A_z10	A1+A1	A3+A4	A6+A6
0629	Selangor2	*M. acuminata*	*malaccensis*	A_m14+A_m17	A1+A2	A1+A2	
0609	Pahang	*M. acuminata*	*malaccensis*	A21+A21	A1+A2	A3+A3	A1+A1
1345	Pisang Kra	*M. acuminata*	*malaccensis*	$A_m14+A22$	A1+A2	A1+A2	A3+Ax
0253	Borneo	*M. acuminata*	*malaccensis*	A22+A23	A1+A2	A1+A1	A10+A10
0966	zebrina	*M. acuminata*	*zebrina*	A_z3+A_z6	A1+A2	A3+A4	A8+A9
1177	zebrina	*M. acuminata*	*zebrina*	A_z1+A_z1	A1+A1	A1+A1	A9+A9
1178	Buitenzorg	*M. acuminata*	*zebrina*	A_z1+A23	A1+A2	A1+A1	
0728	Maia Oa	*M. acuminata*	*zebrina*	A_z1+A_z1	A1+A1	A1+A1	A9+A9
1179	Monyet	*M. acuminata*	*zebrina*	A_z1+A_z1	A1+A1	A1+A1	A2+A9
0393	truncata	*M. acuminata*	*truncata*	A_t30+A_t30	A1+A1	A5+A6	A14+A14b
local	Yala	*M. acuminata*	*truncata*	A_t28+A_t29	A1+A2	A2+A2	A8+A14c

ITC#	Name	Genome / Species	Subspecies / CV Group	CAT1	GBSS1	IDH1	ARF
1028	Agutay	*M. acuminata*	*errans*	A23+A23	A1+A1	A6+A6	A12+A12
0619	Banksii	*M. acuminata*	*banksii*	A_b24+A_b24	A1+A1	A1+A1	A11+A11
0853	Banksii	*M. acuminata*	*banksii*	A_b24+A_b24	A1+A1	A1+A1	A11+A11
0879	Banksii	*M. acuminata*	*banksii*	A_b25+A_b25	A1+A1	A1+A1	A11+A11
0896	Banksii	*M. acuminata*	*banksii*	A_b24+A_b24	A1+A1	A1+A1	A11+A13
0464	Higa	*M. acuminata*	*banksii*	A_b24+A_b24	A1+A1	A1+A1	A11+A11
0465	Waigu	*M. acuminata*	*banksii*	A_b24+A_b24	A1+A1	A1+A1	A11+A11
0766	Paliama	*M. acuminata*	*banksii*	A_b24+A_b24	A1+A1	A1+A1	A11+A11
1187	Tomolo	*M. acuminata*	*banksii*	A_b24+A_z8	A1+A1	A1+A1+A3	A11+A11
0806	GodMun	*M. acuminata*	*banksii*	A_b24+A_b24	A1+A1	A1+A1	A11+A11
0846	schizocarpa	*M. schizocarpa*	-	S+S	S+S	S+S	S+S
0926	schizocarpa	*M. schizocarpa*	-	S+S	S+S	S+S	S+S
1002	schizocarpa	*M. schizocarpa*	-	S+S	S+S	S+S	S+S
1152	Wompa	AS	-	A20+S	A+S	S+S	A13+S
0935	Sosi	AS	-	A_b24+S	A+S	S+S	A11+S

Allele numbering for the CAT1 A-genome haplotypes refers to alleles as identified in Fig. 1a. For the A genome at other loci and for the B genome at CAT1 and ARF, the numbering indicates the presence of more than one allele in an individual. For the B genome at IDH1, the numbering refers to the alleles as identified in Fig. 1b while for GBSS, the B-genome numbering is for alleles at a clearly distinguishable length polymorphism: $B_1 = A_7G$, $B2 = A_8CA_3$ and $B3 = A_{12}CA_3$.

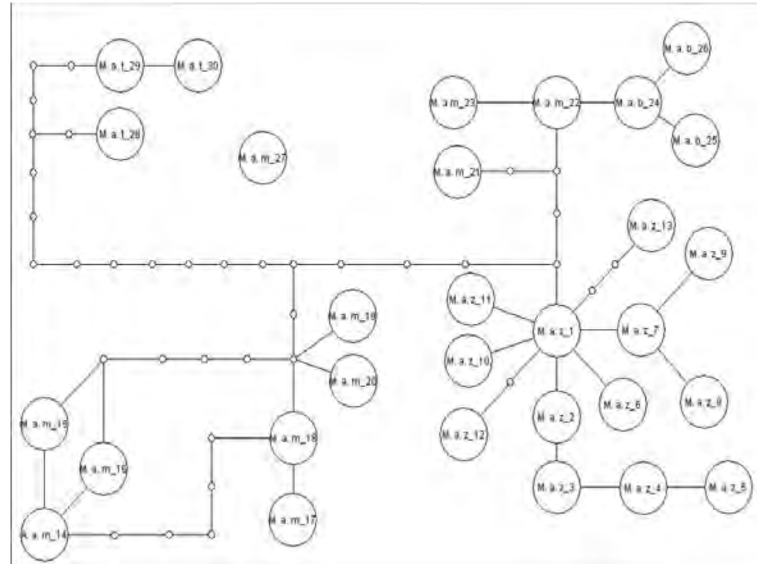
Table 2. List of primers used for PCR amplification and sequencing.

gene	PCR amplification	Sequencing	PCR product length	Aligned sequence length
ADH	ADH1-F TTTTgggAAgCCAAggTAggTg	ADH1-seqF CCATgAggCAgCAgggTA	1000	650
	ADH1-R AACgCAgATATCATggCgTCgAT			
CAT	CAT1-F ACCAggAgTACAAACCCCTA	CAT-seqF ACTgTTATTCATgAgCg	1250	880
	CAT1-R CggAATAAgAgAAAATTCTggT	CAT-seqR CCACTCCggCTAgTTTCC		
G3PDH	G3PDHcp-F gCTCAgAACATCATCCCATCTTC	G3PDHcp-seqR gTCgTTCAgTgCgATACC	1000	ND
	G3PDHcp-R gATACCAgTTTAACgAAgTTgTTC			
	G3PDHcy-F AgCTTCAACATCATTCCTAgCAg	G3PDHcy-seqR AggATTCAAggCAATTCC	1000	ND
	G3PDHcy-R gACACAAgCTTggCAAAgTTAgg			
GBSS	GBSS1-F AgCgTgCAggTTgAggTATTgC	GBSS-seqF gTTCgCTTCTTCCACTgCTA	1020	800

gene	PCR amplification	Sequencing	PCR product length	Aligned sequence length
	GBSS1-R gTCgTACCTgCATggAACACATC			
	GBSS2-F AATgTCCTgCATggAACACATC	GBSS-seqR CgATAgCATACCCgTCgg	1000	180
	GBSS2-R CTTCCTTTCACAggCCTgTTAg			
IDH	IDH1-F gATTTCTTAgCTCAAggTgAg	IDH-seqF ATCTCgAgTAACCCTgAgT	1170	780-800
	IDH1-R TTCTTCggTgTTCAgATACTT			
ARF6/8	ARF6-F ggTCATAgTgAACAggTCAgTC	ARF-seq ACTCACAAAAACACTCCA	1300	800
	ARF6-R ACTgATTgTTCTCATTCCTgAgg			

Figures

(a)



(b)

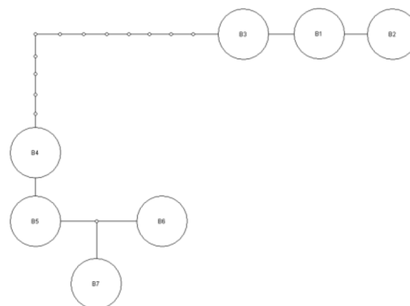


Fig. 1. Parsimony networks for (a) the Musa acuminata haplotypes found at the CAT1 locus and (b) the Musa balbisiana haplotypes identified at the IDH locus. Each edge represents a mutation (nucleotide substitution or insertion deletion) and the small circles represent unsampled haplotypes. Haplotypes A_z1-A_z13 were found in accessions belonging to the M. acuminata *zebrine/burmanica/siamea* group, A_m14-A_m20 t were found in the *malaccensis/micocarpa* group, while the *banksii* group contained the alleles A_b24-A_b26 and the *truncata* group A_t28-A_t30. Alleles 20-23 have not been assigned to a particular subspecies as they are rare but found in one or two accessions across different groups. Allele 27 from 'Pisang Yari Buaya' did not join with the rest of the network at 95% confidence because of several unique substitutions.