



Genetic diversity of wild *Musa acuminata* in Sumatra

Fajarudin Ahmad¹ · Yuyu S. Poerba¹ · Lulut D. Sulistyaningsih² · Siti Subandiyah³ · Hans de Jong⁴ · Gert H. J. Kema⁵ · Hugo A. Volkaert^{6,7}

Received: 4 October 2024 / Accepted: 11 July 2025 / Published online: 9 October 2025
© The Author(s) 2025, modified publication 2026

Abstract

The islands of Indonesia are a major part of the centre of origin of bananas. Here, we evaluated the diversity at 164 locations of Sumatra and added 20 wild banana accessions originating from other islands of Indonesia. We analysed morphological characteristics in a principal components analysis (PCA) and genotyped accessions at four nuclear gene regions, i.e. ADH3, CAT2, GBSS1 and IDH1. By combining the geographical coordinates with the morphological variation, we developed a spatial multivariate analysis. The PCA of the morphological traits clearly separated three major clusters. Based on the collected information, we propose to merge the previously described var. *halabanensis* and var. *alasensis* Nasution into the subsp. *halabanensis*. Moreover, we argue that the var. *sumatrana* should be considered as a subspecies taking into account its genetic and morphological unique characters in addition to its geographical distribution. The four nuclear genes genotyping showed a high genetic diversity of *M. acuminata* often in admixed stands of different haplotypes. Based on the allele frequencies, there was a significant deviation from a Hardy–Weinberg equilibrium with a high level of homozygosity. Multi-locus genetic population structure analysis confirmed the existence of three major clusters with further possible subdivisions. The discovered genetic diversity in the wild *M. acuminata* population across Sumatra expands the current knowledge and awaits its deployment in ongoing breeding efforts for this important crop.

Keywords Banana · Haplotype · Indonesia · Population · Sumatra · Taxonomy

Introduction

Bananas belong to the genus *Musa*, the largest in the family Musaceae in the order Zingiberales (Kress and Specht 2006). The genus comprises 69 species classified into two sections, *Musa* ($2x = 2n = 22$) and *Callimusa* ($2x = 2n = 20$ rarely 14 or 18) (Häkkinen 2013). The centre of diversity of *Musa* is Southeast Asia, particularly on the islands of the

Indonesian archipelago (Simmonds 1962; Nasution 1991). Two species in the section *Musa*, *Musa acuminata* Colla (A genome) and *M. balbisiana* Colla (B genome) are key species in the origin of most cultivated bananas (Stover and Simmonds 1987). Intra- and interspecific hybridization has led to a multitude of edible bananas as diploids (AA, AB), triploids (AAA, AAB, ABB), or tetraploids (ABBB). Edible bananas were brought into cultivation and carried across the tropics along with major human migration events, starting

Handling Editor: Julien Boutte.

✉ Fajarudin Ahmad
fajarudin.ahmad@brin.go.id

✉ Gert H. J. Kema
gert.kema@wur.nl

✉ Hugo A. Volkaert
hugo.a.volkaert@gmail.com

¹ Research Center for Applied Botany, National Research and Innovation Agency (BRIN), Bogor 16911, Indonesia

² Research Center for Biosystematic and Evolution, National Research and Innovation Agency (BRIN), Bogor 16911, Indonesia

³ Faculty of Agriculture, Gadjah Mada University, Yogyakarta 55281, Indonesia

⁴ Laboratory of Genetics, Wageningen University and Research, P.O. Box 16, 6700 AA Wageningen, The Netherlands

⁵ Laboratory of Phytopathology, Wageningen University and Research, P.O. Box 16, 6700 AA Wageningen, The Netherlands

⁶ Center for Agricultural Biotechnology, Kasetsart University Kamphaengsaen Campus, Nakhon Pathom, Thailand

⁷ Center of Excellence on Agricultural Biotechnology (AG-BIO/MHESI), Bangkok, Thailand

thousands of years ago (Perrier et al. 2009). In contrast to the cultivated bananas, wild banana fruits are filled with seeds and have little edible pulp. They grow in the forests, along riversides, while some of them survive among plantations and abandoned fields. *Musa acuminata* is a morphologically highly diverse species. Currently, nine sub-species are recognised within *M. acuminata*, i.e., subsp. *acuminata*, subsp. *banksii* (F.Muell.) N.W.Simmonds, subsp. *burmanica* N.W.Simmonds, subsp. *errans* (Blanco) R.V.Valmayor, subsp. *halabanensis* (Meijer) M.Hotta, subsp. *malaccensis* (Ridl.) N.W.Simmonds, subsp. *microcarpa* (Becc.) N.W.Simmonds, subsp. *siamea* N.W.Simmonds, subsp. *truncata* (Ridl.) N.W.Simmonds (Häkkinen and Väre 2008; Hotta 1989; Simmonds 1956; Valmayor 1998; WFO 2024). In addition, Nasution (1991) described 15 varieties of *M. acuminata* in Indonesia alone, though some are synonymous with described subspecies (Häkkinen and Väre 2008; WFO 2024) (see also Online Resource 1).

Presently, approximately 1900 named varieties (cultivars) of banana are grown by farmers in backyard gardens, small fields, or large-scale plantations (Crichton et al. 2016). Global banana production is estimated at 155 million tons/year, ranking fourth in production after rice, wheat and corn (FAO 2020). In spite of their importance for food security, banana breeding, research, and development programs are very limited in the private and public sectors. However, some international institutes and consortia have initiated research for banana improvement (Ruas et al. 2017; Smith and Pillay 2020).

The study of genetic diversity of cultivated bananas and their wild ancestors is important for conservation and breeding. Swangpol et al. (2007) used four chloroplast regions to study the contribution of various wild bananas to the cultivars. Later, Li et al. (2013) reported haplotypes of cultivated bananas associated with wild bananas from Southeast Asia. However, most of these studies used only a limited number of samples from the centre of origin. Volkaert (2011) traced haplotypes found in cultivated bananas back to their presence in wild germplasm using the alcohol dehydrogenase (ADH3), auxin response factor 17 (ARF17), catalase (CAT2), granule-bound starch synthase (GBSS1) and isocitrate dehydrogenase (IDH1) genes in a multi-locus genotyping approach. A major finding of this study was that haplotypes present in several cultivars were not found among the wild accessions available in the International Transit Centre (ITC) of *Musa* Germplasm that is maintained at the Catholic University Leuven in Belgium. We, therefore, considered that an in-depth study of wild bananas from the centre of origin, particularly the western part of Indonesia where there is limited knowledge about wild *Musa acuminata*, could add to the understanding of their diversity and phylogeny and their contribution to the cultivated bananas. Nasution (1991) reported five wild *M. acuminata* varieties on Sumatra, which

is one of the main islands in western Indonesia. To extend the exploration and collection of genetic diversity of wild bananas, we sampled *M. acuminata* at 164 collection sites across Sumatra and studied the diversity using morphological observations and multi-locus genotyping. As a result of the analysis, we propose to merge two *M. acuminata* varieties into an existing subspecies and elevate another variety to the level of subspecies.

Materials and methods

Plant and data collection

Plant materials were collected during five week-long expeditions crossing Sumatra Island from North to South between March and June 2017. In wild populations with comparable morphology, samples were collected every 5–10 km, but plants with different morphologies were also sampled whenever these were spotted. One accession was collected in each sampling site, though when clear variation was observed at a sampling site, additional samples were taken. The leaf base, bud, fruits with seed and other tissues were collected and processed for preservation in the herbarium. Cigar leaves were collected and stored on ice for DNA isolation in the laboratory. For some accessions the corm was dug out for planting in the live collection at the Cibinong Botanical Gardens. GPS locations were recorded for each sample and plant morphology was documented according to the “Descriptors for Banana” (IPGRI 1996) and according to the morphological characteristics tentative groups were defined. Data for 20 wild banana accessions originating from other regions of Indonesia that were maintained in the collection of the National Research and Innovation Agency (BRIN) (Online Resource 2) were added to capture some of the overall genetic and morphological variation of Indonesian bananas. In this study, we also including two diploid AA cultivated banana accessions, i.e. ANS-26 “Pisang Dingin Lidi” and LIPI-232 “Pisang Rejang” that commonly found in Sumatra that may provide valuable insights into the genetic information of these bananas.

Species determination

Species determination and description of *Musa acuminata* were conducted by observing and cross-checking samples with type specimens, their protologues, and the general collections of *M. acuminata* housed in the Herbarium Bogoriense (BO) and digital herbaria provided by JSTOR. Morphological characterization was performed using the ‘Descriptors for Banana (*Musa* spp.)’ (IPGRI 1996). Traits were documented using a digital camera under standardized

conditions. Differentiating characteristics obtained from field observations were cross-referenced with published references and herbarium specimens to confirm species determination.

Data analysis

In order to determine and visualise the clustering of the collected accessions based on the multivariate morphological characters, a principal component analysis (PCA) was performed using ClustVis (Metsalu and Vilo 2015). Accessions and characters with more than 25% missing data were excluded. Eventually, 88 accessions (54% of total) and 65 morphological characters were retained (Online Resource 3). To determine potential spatial structure among the accessions a principal coordinates of neighbour matrices (PCNM) (Borcard and Legendre 2002) was conducted using the vegan library (Oksanen et al. 2020) in R (R-Core-Team 2020) on the same data matrix along with the GPS coordinates.

DNA was isolated following the CTAB method (Syamkumar et al. 2003) with modification, i.e., an addition of 0.1 g Polyvidone 25 (PVP) (Merck, Cat no. 1.07443) in 800 μ L extraction buffer containing 4% Cetyltrimethylammonium bromide (CTAB) (Merck, Cat no. 219374). DNA was also extracted from freeze-dried leaves in the BRIN collection using the Promega Wizard® Magnetic DNA Purification System (Promega, Madison, USA) or Geneaid™ Genomic DNA Mini Kit Plant (Geneaid, Taiwan), following the manufacturer's instructions.

For the population genetic study, 75 representative samples of wild *M. acuminata* from Sumatra were selected based on the geographical and morphological diversity together with 20 wild *M. acuminata* accessions and two diploid AA cultivars from the BRIN collection (Online Resource 2). DNA fragments of four nuclear protein coding regions (ADH3, CAT2, GBSS1 and IDH1) each located on a different chromosome were amplified by PCR (Table 1). These reactions were performed in a 15 μ L reaction volume using 0.075 unit/ μ L GoTaq® G2 DNA Polymerase (Promega, Madison, USA) with 32 μ M dNTPs, 0.53 μ M each primer,

1 $\times$ GoTaq® Flexy 2 buffer (Promega, Madison, USA) and 1.5–2.0 mM $MgCl_2$ following the GoTaq® G2 DNA Polymerase manual. The pre-denaturation for 4 min at 94 °C was followed by 40 cycles of 45 s at 94 °C, 45 s at 53 °C, 90 s at 72 °C, and 5 min for a final elongation at 72 °C. The resulting amplicons were checked by agarose gel electrophoresis and sent for sequencing (Macrogen Europe B. V., Amsterdam, The Netherlands). In order to avoid PCR recombination (Meyerhans et al. 1990; Cronn et al. 2002) and other cloning artefacts, we sequenced directly from the PCR products starting from specific internally located primers (Table 1).

Sequencing chromatograms were inspected in ChromasLite and polymorphic base calls were manually recoded when necessary. Haplotypes were reconstructed from diploid sequences in DnaSP version 6.12.01 (Librado and Rozas 2009) using the PHASE algorithm (Stephens et al. 2001). Diploid sequences with insertion-deletion (indel) polymorphisms were separated using Indelligent v.1.2 (Dmitriev and Rakitov 2008). Subsequently the haplotype networks were constructed using the statistical parsimony approach in TCS v2.1 (Clement et al. 2000). Indels longer than one nucleotide were reduced to single positions. To place the Sumatran bananas in a broader perspective, haplotypes from previous research comprising wild and cultivated bananas were included when constructing the haplotype networks (Volkaert 2011). The genotypes for each individual were then called based on the observed haplotypes. The expected heterozygosity, fixation index, and tests for deviations from the Hardy–Weinberg equilibrium (HWE) were calculated using GenAlEx 6.5 (Peakall and Smouse 2006, 2012) for all Indonesian accessions, all wild accessions from Sumatra and the three morphological subspecies. The analysis was repeated on a dataset in which all singleton haplotypes were joined with their closest non-singleton haplotype (Online Resource 4). Measures of population differentiation (F_{is} , F_{it} , F_{st} , G_{st} (Nei 1973), G'_{st} (Hedrick 2005), D (Jost 2008)) were calculated in the mmod package version 1.3.3 (Winter 2012) in R (R-Core-Team 2020).

Table 1 Nucleotide sequence (5' to 3') of the primers used in PCR and DNA sequencing

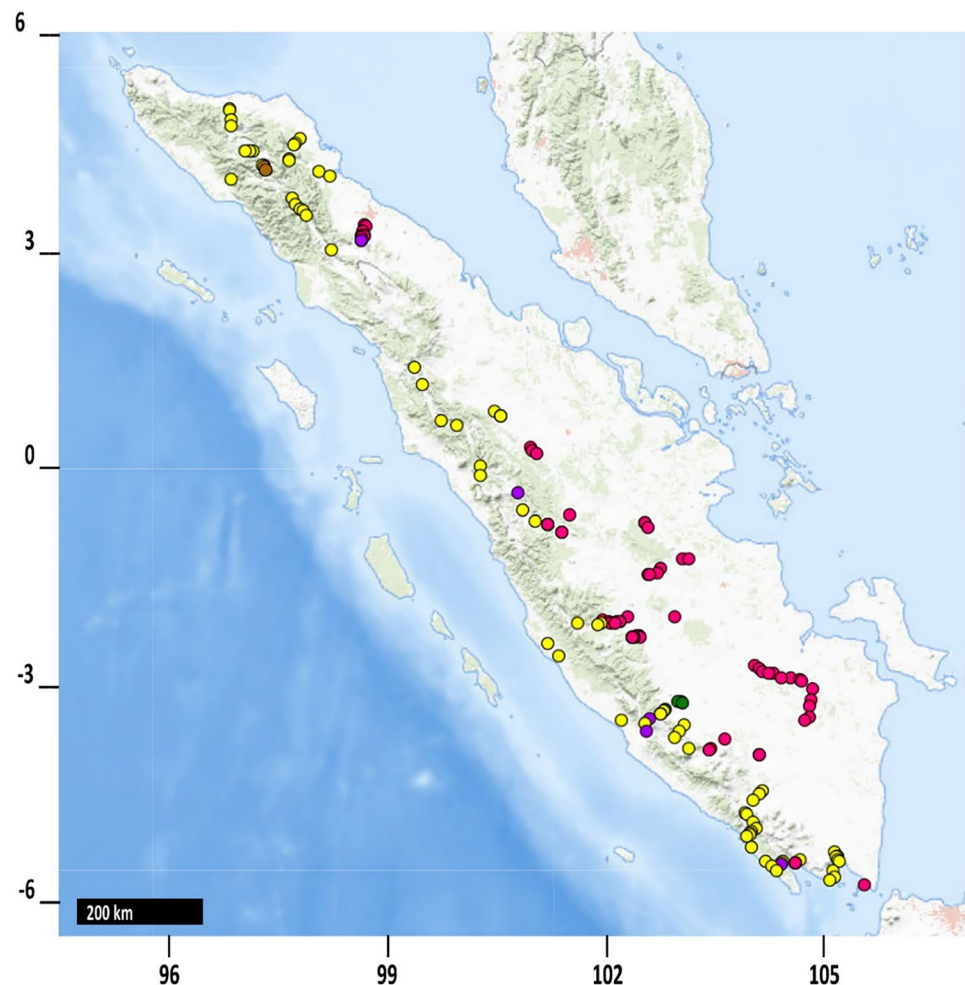
Locus	PCR primers	Sequencing primers	Function
ADH3	F: TTTTGGGAAGCCAAGGTAGGTG R: AACGCAGATATCATGGCGTCGAT	F: CCATGAGGCAGCAGGGTA R: TTACCTTCTTCAAATCTC	Alcohol dehydrogenase
CAT2	F: ACCAGGAGTACAAACCCCTA R: CGGAATAAGAGAAAATTCTGGT	F: ACTGTTATTCATGAGCG R: CCACTCCGGGTAGTTTCC	Catalase
GBSS1	F: AGCGTGCAGGTTGAGGTATTGC R: GTCGTACCTGCATGGAACACATC	F: GTTCGCTTCTTCCACTGCTA R: AAGAAGACCAGTGTGCCA	Granule-bound starch synthase
IDH1	F: GATTTCTTAGCTCAAGGTGAG R: GAACTCTTCGGTGTTTCAGATAC	F: AATTCTATGACTTATGGCA R: ATCTCGAGTAACCTGAGT	Isocitrate dehydrogenase

Population structure within the Indonesian and Sumatran accessions was analysed using the Bayesian clustering for multi-locus genotyping data implemented in the programme Structure v2.3.4 (Pritchard et al. 2000). Only individuals with molecular data for all four loci were included. The number of assumed clusters was set from 1 to 8 with 300,000 burn-in and 1,000,000 sampling steps, allowing for admixture and assuming correlated allele frequencies. The Structure analysis was also repeated using a dataset where singleton haplotypes had been joined with their closest non-singleton haplotype. To get insight into the potential number of subpopulations the results were further processed following the Evanno et al. (2005) method implemented in StructureHarvester on <https://taylor0.biology.ucla.edu/structureHarvester/> (Earl and vonHoldt 2012) and the Puechmaillle (2016) method in StructureSelector on <https://lmme.ac.cn/StructureSelector/index.html> (Li and Liu 2018).

Results

Data on a total of 164 wild *M. acuminata* accessions were collected (Online Resource 5) across seven provinces of Sumatra (Fig. 1). Already while collecting in the field, it was clear that most of the wild bananas could be tentatively grouped into three types based on morphological characteristics. One type was found below 750 m in most areas of the island and had a red inflorescence bud and angular seed. Another common type was found in lowland areas and mountain valleys below 1000 m, mostly in the western half of Sumatra, and had a purplish inflorescence bud and small globular seed with a smooth surface. In mountainous areas at an altitude of 750 m up to 1400 m, the third type occurred. It had darker purple inflorescence bracts and generally longer fruits containing flattened seeds with rough edges. We also observed various intermediate phenotypes deviating from the morphology of the widely distributed types such as accession SSB-33 (Online Resource 6). At a few locations in Southern Sumatra, we identified plants with a markedly longer petiole (usually up to half the length of the leaf blade)

Fig. 1 Collecting sites of all 164 wild *Musa acuminata* on Sumatra Island. The dots represent the collection sites and their colour represents the infraspecific taxa found at the site. Yellow dots subsp. *halabanensis*, red dots subsp. *malaccensis*, purple dots subsp. *sumatrana*, green dots var. *longepetiolata*, and brown dots banana of Mount Mugajah population. Grey shading represents areas above 1000 m



and a red inflorescence bud. Finally, we identified a small population of wild bananas with several distinct characteristics of the male bud and fruit at very high elevations (1200–1500 m) in the centre of the Aceh province.

General morphology of *Musa acuminata* in Sumatra

The *M. acuminata* accessions align with the following general description. Plants suckering freely; mature pseudostem height varies between 2 and 5 m (rarely > 5 m), slender to normal; the pseudostem colour is green yellow to green or green–brown to black purple or dark brown with or without pigmentation, shiny or dull; the sap is watery or milky. Petioles have small to large blotches at base; the petiole canal can be closed, overlapping, open with margin spreading, wide with erect margins, or straight with erect margins; the lamina is truncated at the top. The inflorescence bud colour is red to purple with very little to abundant wax, shaped like a top to rounded; the male flower colour is white to cream, with abundant pollen. The fruit bunch is horizontal, at an angle, or vertical. Fruits contain numerous black seeds, flat or angular or globular in shape; smooth or wrinkled. *Musa acuminata* grows in open areas along rivers, forest, palm plantations, along roadsides, rubber plantations or close to villages.

Based on key characters of wild *M. acuminata* we described 3 varieties and one infraspecific *M. acuminata*

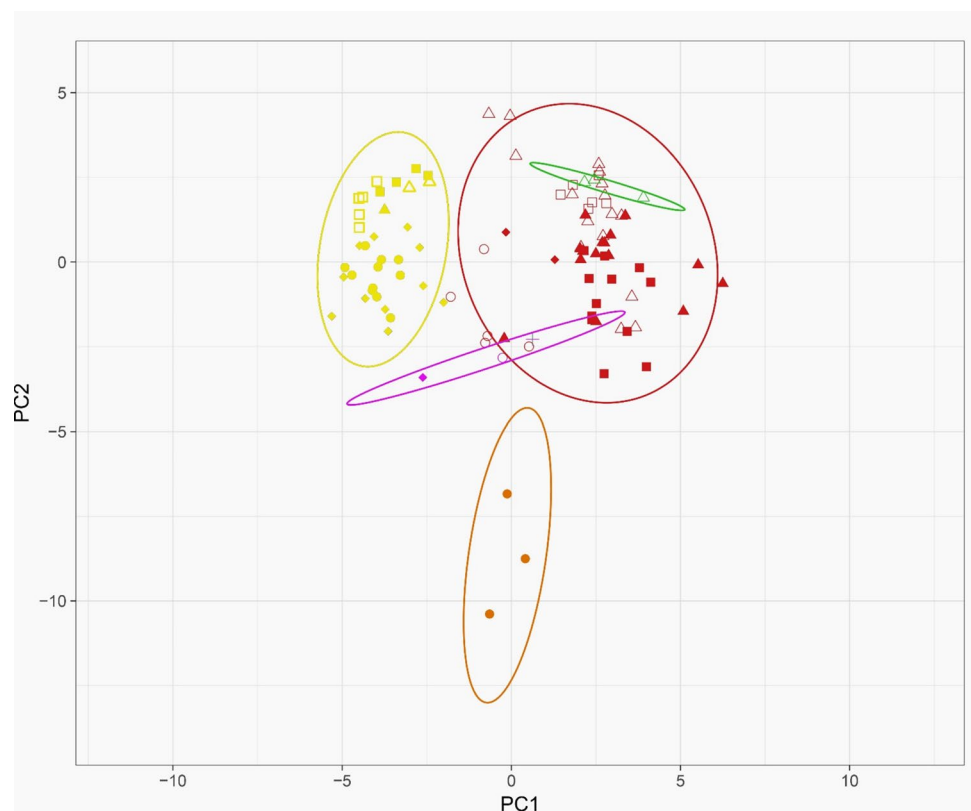
that not described in any publication yet. Accessions with flat angular seed described as var. *sumatrana* (Becc.) Nasution, accessions with red like a top male but described as var. *malaccensis* (Ridl.) Nasution, accessions with purple rounded male bud as var. *halabanensis* (Meijer) Nasution, accessions with similar characteristic of fruit and inflorescence to var. *malaccensis* but the petiole length of leaf is more than half of lamina as var. *longepetiolata* and accessions with brown male bud and longer petiole of fruit than other wild *M. acuminata* found in Sumatra as Mount Mugajah population.

In nearly all of the 38 observations of var. *halabanensis* in their natural environment we noted undeveloped fruits in one or more of the early hands while later hands were fully developed (Online Resource 7). This characteristic is not found in other varieties. Among var. *halabanensis* accessions, we observed two accessions, one in Aceh (ANS-32) and the other in Bengkulu province, (SSB-29), where the fruit peel split open at maturity (Online Resource 8).

Phenotypic analysis

The PCA of morphological data clearly separated the subsp. *malaccensis* from subsp. *halabanensis* with the Mount Mugajah population forming a separate cluster. However, the subsp. *malaccensis* cluster completely overlapped the var. *longepetiolata* cluster and partially overlapped the

Fig. 2 Principal component analysis of the morphological characters of 88 wild *Musa acuminata* accessions from Sumatra with less than 25% missing data. The colour of the symbol indicate the subspecies/variety, i.e. red subsp. *malaccensis*, yellow subsp. *halabanensis*, purple subsp. *sumatrana*, green subsp. *longepetiolata*, brown Mount Mugajah population. The shape of the dot indicates the province within Sumatra, i.e. filled circles Aceh, filled squares Bengkulu, filled triangles Jambi, filled diamonds Lampung, empty circles North Sumatra, empty squares Riau, empty triangles South Sumatra, plus signs West Sumatra



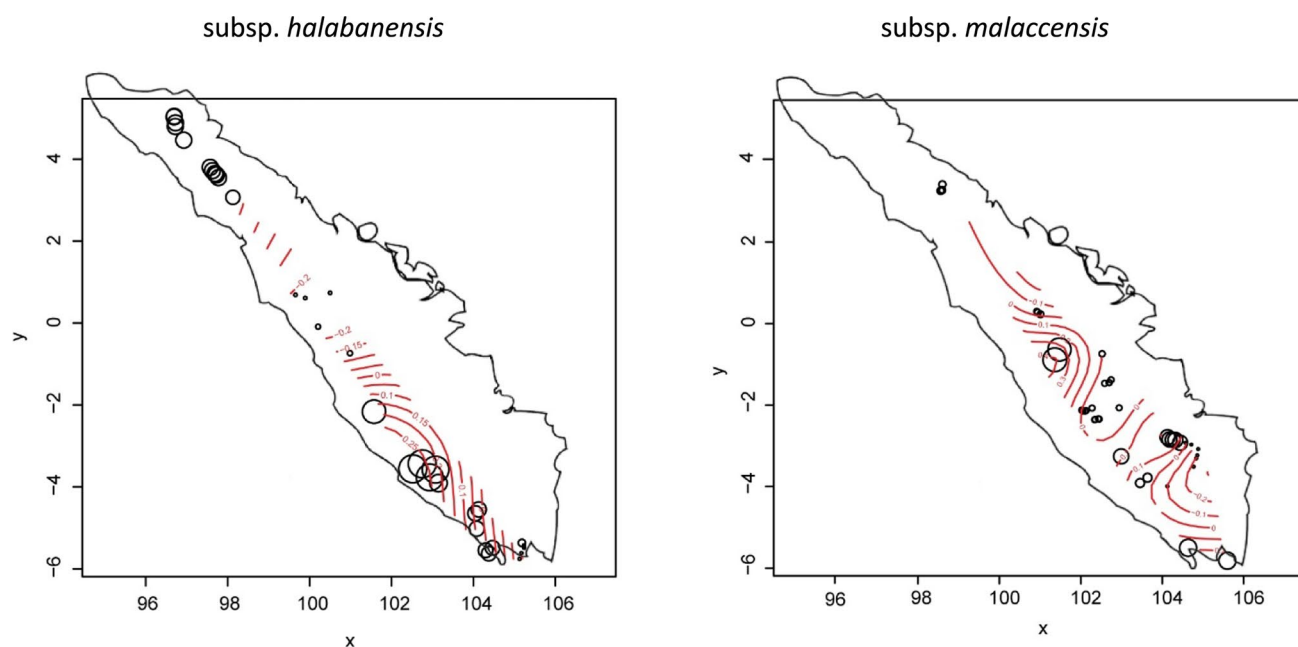


Fig. 3 A spatial multivariate analysis (eigen vector calculated using principal coordinate analysis of neighbour matrices, PCNM) of two wild *Musa acuminata* subspecies on Sumatra according to morphological characteristics. Dot sizes are proportionate to the Eigen val-

ues that represent the difference of morphological characters among accessions. The lines reflect spatial correlation that suggesting a relationship between locations. *x* latitude, *y* longitude

Table 2 Frequency of substitutions and insertions/deletions in exons and introns across the ADH3, CAT2, GBSS1, and IDH1 loci that were sequenced for a multi-locus genotyping analysis of 69 wild *Musa acuminata* accessions across Sumatra and 20 accession from BRIN collection

Locus	Substitutions		Insertions/deletions		Length (bp)		Total
	Exon	Intron	Exon	Intron	Exon	Intron	
ADH3	25	36	1	2	494	416	910
CAT2	23	76	2	3	666	490	1156
GBSS1	18	58	1	7	494	501	995
IDH1	17	82	0	8	301	804	1105
Total	83	252	4	20	1955	2211	4166

Table 3 Number of accessions and number of new haplotypes detected at the ADH3, CAT2, GBSS1, and IDH1 loci that were sequenced for a multi-locus genotyping analysis of 69 wild *Musa acuminata* accessions across Sumatra and 20 accession from BRIN collection

Locus	Accessions		Haplotypes			Total
	# wild Sumatra accessions	# BRIN accessions	New haplotypes	Haplotypes from Sumatra identical to references	Haplotypes from previous studies	
ADH3	57	12	16	8	38	54
CAT2	62	19	26	14	45	71
GBSS1	65	16	39	7	59	98
IDH1	67	14	23	9	39	62
Total			104	38	181	285

Table 4 Population genetic parameters at each locus for *M. acuminata* subsp. *malaccensis*, subsp. *sumatrana* and subsp. *halabanensis* from Sumatra including the number of individuals (*N*), number of alleles (*Na*), effective number of alleles (*Ne*), number of private haplotypes (*Pvt*), information index (*I*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), unbiased expected heterozygosity (*uHe*), fixation index (*F*) and *Fis*, *Fit*, *Fst*, *Gst* (Nei), *G'st* (Hedrick) and *D* (Jost) measures of population differentiation. Singleton haplotypes were merged with the nearest non-singleton

Population		ADH3	CAT2	GBSS1	IDH1
subsp. <i>halabanensis</i>	N	27	29	31	31
	Na	4	12	16	9
	Ne	1.361	4.123	9.376	5.339
	Pvt	1	8	11	7
	I	0.564	1.884	2.497	1.856
	Ho	0.148	0.448	0.581	0.355
	He	0.265	0.757	0.893	0.813
	uHe	0.270	0.771	0.908	0.826
	F	0.442	0.408	0.350	0.563
	F	0.442	0.408	0.350	0.563
subsp. <i>malaccensis</i>	N	24	25	26	26
	Na	10	13	15	8
	Ne	7.200	9.470	11.860	5.408
	Pvt	8	11	12	7
	I	2.098	2.396	2.574	1.839
	Ho	0.375	0.640	0.654	0.500
	He	0.861	0.894	0.916	0.815
	uHe	0.879	0.913	0.934	0.831
	F	0.565	0.284	0.286	0.387
	F	0.565	0.284	0.286	0.387
subsp. <i>sumatrana</i>	N	8	7	7	8
	Na	3	4	5	6
	Ne	1.293	2.970	2.649	5.333
	Pvt	0	1	1	3
	I	0.463	1.233	1.253	1.733
	Ho	0.250	0.143	0.143	0.000
	He	0.227	0.663	0.622	0.813
	uHe	0.242	0.714	0.670	0.867
	F	-0.103	0.785	0.770	1.000
	F	-0.103	0.785	0.770	1.000
All	Ho	0.254	0.492	0.562	0.369
	He	0.794	0.918	0.947	0.921
	Fis	0.429	0.468	0.434	0.650
	Fit	0.574	0.551	0.501	0.693
	Fst	0.430	0.156	0.119	0.125
	Nm	0.331	1.351	1.847	1.753
	Gst	0.414	0.133	0.096	0.104
	G'st	0.965	0.942	0.865	0.933
	D	0.928	0.929	0.844	0.921
	D	0.928	0.929	0.844	0.921

subsp. *sumatrana* cluster (Fig. 2). As the PCA cannot test the apparent association of genetic diversity with geographical origin, we conducted a principal coordinate analysis of neighbour matrices, PCNM, which provided variation of Eigenvalues that represent the difference of morphological characteristics among accessions. Based on this analysis, the

Table 5 The deviations from Hardy–Weinberg Equilibrium (*HWE*) for each locus for *M. acuminata* subsp. *malaccensis*, subsp. *sumatrana* and subsp. *halabanensis* and combined for all populations from Sumatra

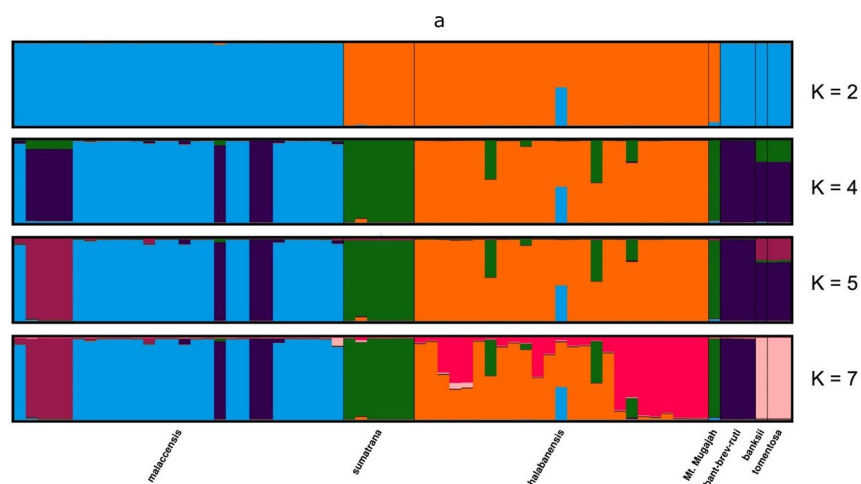
Population	Locus	DF	ChiSq	Probability	Significance
subsp. <i>halabanensis</i>	ADH3	6	27.204	<0.001	***
subsp. <i>halabanensis</i>	CAT2	66	160.789	<0.001	***
subsp. <i>halabanensis</i>	GBSS1	120	190.848	<0.001	***
subsp. <i>halabanensis</i>	IDH1	36	118.184	<0.001	***
subsp. <i>malaccensis</i>	ADH3	45	157.244	<0.001	***
subsp. <i>malaccensis</i>	CAT2	78	159.618	<0.001	***
subsp. <i>malaccensis</i>	GBSS1	105	142.800	0.008	**
subsp. <i>malaccensis</i>	IDH1	28	78.225	<0.001	***
subsp. <i>sumatrana</i>	ADH3	3	0.163	0.983	ns
subsp. <i>sumatrana</i>	CAT2	6	15.921	0.014	*
subsp. <i>sumatrana</i>	GBSS1	10	28	0.002	**
subsp. <i>sumatrana</i>	IDH1	15	40	<0.001	***
All Sumatra	ADH3	78	501.923	<0.001	***
All Sumatra	CAT2	276	711.023	<0.001	***
All Sumatra	GBSS1	435	805.252	<0.001	***
All Sumatra	IDH1	190	718.811	<0.001	***

Significance levels are indicated by asterisks with **P* < 0.05, ***P* < 0.01, and ****P* < 0.001, whereas *ns* not significant

existence of sub-populations within subsp. *halabanensis* and subsp. *malaccensis* in West, Central, and Eastern Sumatra is indicated by the differences in circle size that represent the Eigenvalues in Fig. 3.

Molecular analysis

To evaluate the genetic diversity of wild *M. acuminata* in Sumatra, we generated Sanger sequencing data from four nuclear gene encoding loci for 69 accessions and added 20 accessions from the BRIN collection. The ADH3 locus included a highly polymorphic microsatellite repeat which had to be excluded from further analysis because the correct number of repeat units was impossible to determine in most heterozygous individuals. Eventually, we obtained 910, 1156, 949 and 1105 nucleotide positions for ADH3, CAT2, GBSS1 and IDH1, respectively (Table 2). All haplotype sequences have been deposited in publicly accessible DNA repositories (Accession numbers OW737868–OW737920 (ADH3), OW371161–OW371251 (GBSS1),



b

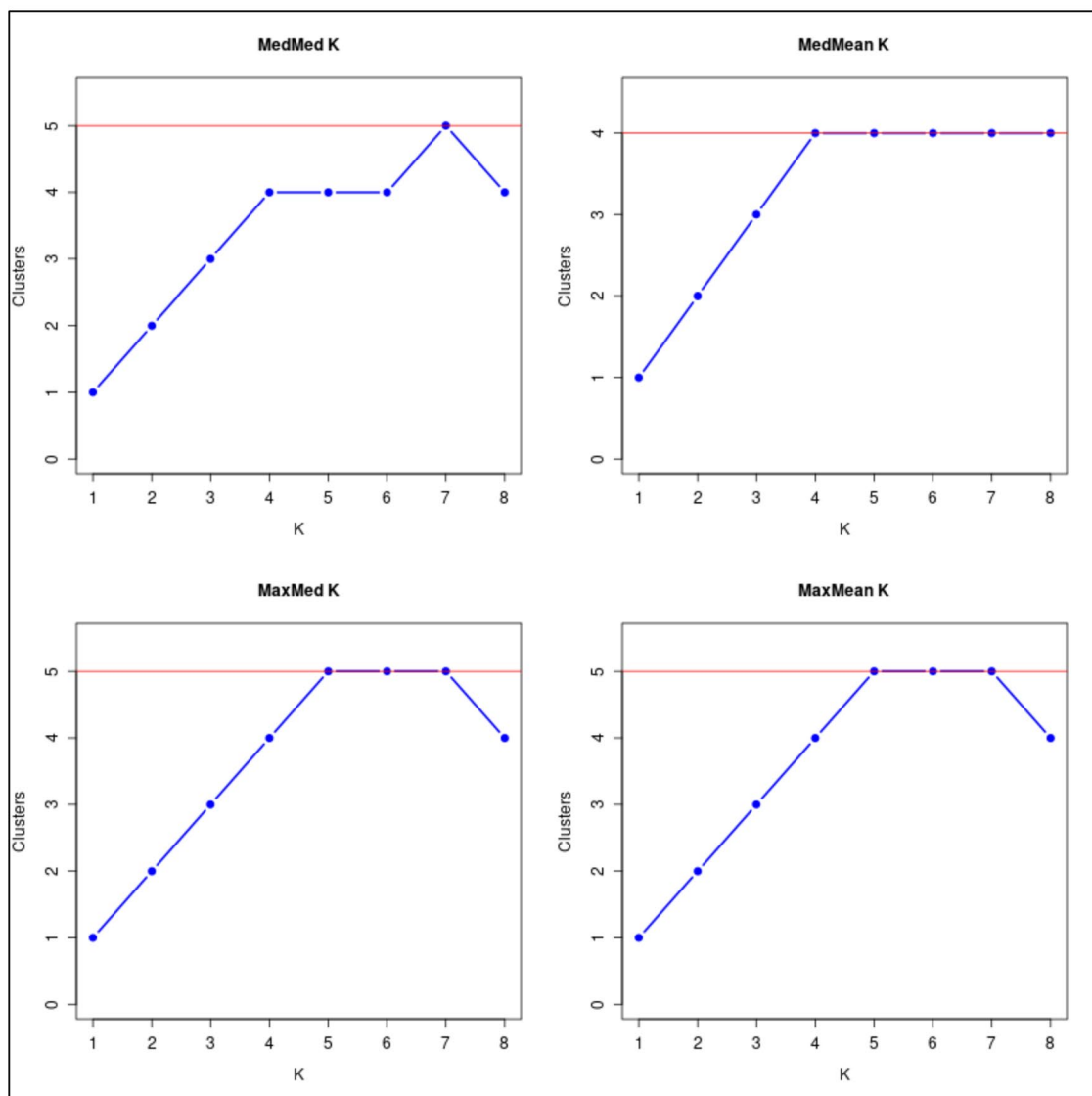


Fig. 4 Population structure analysis for 55 samples of wild *Musa acuminata* from Sumatra and 10 samples from other regions of Indonesia. **a** Partitioning of individuals among $K=2, 4, 5$ and 7 potential subpopulations. From $K=4$ the subsp. *malaccensis*, subsp. *sumatrana* and subsp. *halabanensis* are consistently separated. **b** Determination of the optimal K with the Puechmaille (2016) method. **c** Graphs plotting the rate of change of likelihood and “Delta K ” from the Evanno et al. (2005) method indicating that there might only be 2 subpopulations

OW112086-OW112111 (CAT2), OW111340-OW111363 (IDH1)). The sequenced accessions, combined with references from a previous study (Volkaert 2011), resulted in 54, 71, 98, and 62 haplotypes for ADH3, CAT2, GBSS1, and IDH1, respectively (Table 3). Across all four loci and all haplotypes from Sumatra and elsewhere we observed 83 substitutions (one per 24 bases) and four indels (1 per 489 position) in the coding regions and 252 substitutions (one per 9 bases) and 20 indels (one per 114 position) in the non-coding regions (Table 2). The indel sizes ranged from a single nucleotide to 276 bp long.

Among the *M. acuminata* accessions sampled from Sumatra and other parts of Indonesia we identified 142 haplotypes, of which 104 had not previously been detected (16, 26, 39 and 23 for ADH3, CAT2, GBSS1 and IDH1, respectively). The number of singleton haplotypes (found only once in the whole dataset) was large: 6, 9, 12, and 8 for ADH3, CAT2, GBSS1, and IDH1, respectively, or about one-third of all haplotypes unique to Sumatra at each locus.

Even though the number of individuals sampled per subspecies was relatively small, certainly when compared to the large number of haplotypes detected, population genetic analysis in GenAIEx revealed some interesting trends. The three main subspecies on Sumatra, viz. subsp. *halabanensis*,

subsp. *malaccensis* and subsp. *sumatrana*, contained almost distinct haplotype sets for the four loci, indicated by the large number of private alleles (50 in subsp. *malaccensis*, 43 in subsp. *halabanensis* and 9 in subsp. *sumatrana*), implying a high level of population differentiation. This differentiation of the populations is confirmed by the high or very high values for G_{st} , G'_{st} and D (Table 4). Surprisingly though, in spite of the large number of alleles there is a high level of homozygosity at individual loci, hence the observed heterozygosity is lower than expected for each locus, resulting in highly significant deviations from HWE for the total Sumatra population and for each of the subspecific populations, even after singleton haplotypes were merged with their closest relative (Table 5). Among the accessions from Sumatra, the observed homozygosity at ADH3, CAT2, GBSS1, and IDH1 was 72%, 45%, 38%, and 64%, respectively. Nevertheless, 55 of 69 wild *M. acuminata* accessions from Sumatra were heterozygous for at least one locus. Three out of 25 subsp. *halabanensis* plants and 5 out of 6 subsp. *sumatrana*, as well as the only accession of var. *longepetiolata* were homozygous across all 4 loci.

The multi-locus Bayesian clustering of accessions from Sumatra and other parts of Indonesia with the programme Structure v2.3.4 indicated clear groupings. The optimal K value was either 2 with Evanno et al. (2005) method or more with Puechmaille (2016) method (Fig. 4). At $K=2$ the subsp. *malaccensis* individuals were separated from all others. At $K=3$ the subsp. *halabanensis* was separated. However, it is only at $K=4$ that the three subspecies are consistently separated though then a few accessions within subsp. *malaccensis* clustered together with accessions from outside Sumatra (Fig. 4). Increasing the number of assumed clusters to seven separated an additional small cluster within the subsp. *malaccensis* accessions, joining three individuals

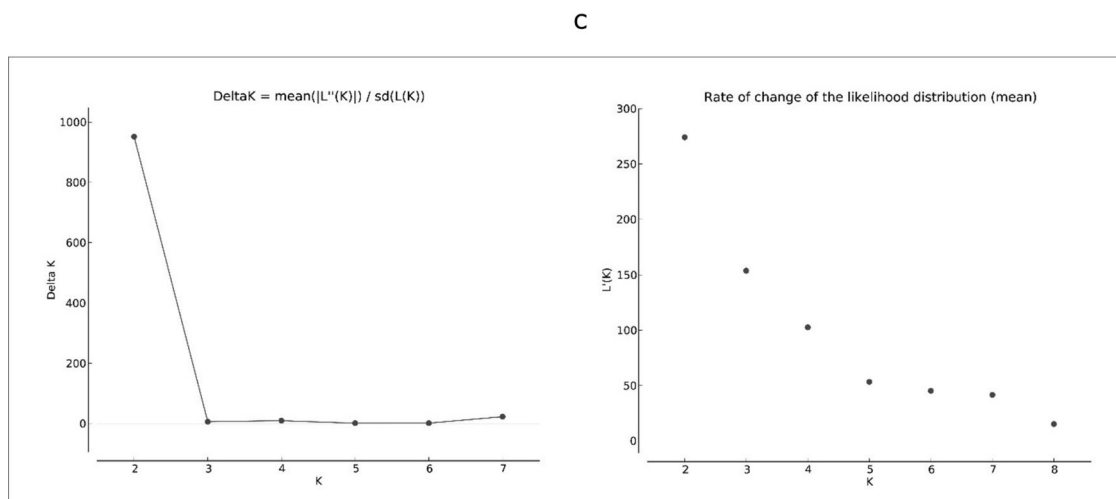


Fig. 4 (continued)

with the var. *bantamensis* and var. *rutlifles* accessions from the BRIN collection. Also, the accessions of subsp. *bank-sii* and subsp. *Tomentosa*—originating East of the Wallace line—separated from all others (Fig. 4). At the same time a consistent split within the subsp. *halabanensis* accessions became apparent separating those collected in Lampung province from the others. At all analyses of $K = 3$ and higher the single accession of the Mt. Mugajah population merged with subsp. *sumatrana*. The level of admixture was generally low except within the subsp. *halabanensis* where individuals showed evidence of admixture with the other two subspecies.

Haplotype networks reconstructed for the four loci clustered the accessions more or less according to the presumed subspecies (Fig. 5). Haplotypes of subsp. *halabanensis* grouped fairly well for all genes. In contrast, the haplotypes of subsp. *malaccensis* and subsp. *zebrina* populated a large

central area of the networks for each locus. Aside from some subspecies-specific haplotypes, the subsp. *sumatrana* accessions also contained haplotypes scattered among those of subsp. *halabanensis* and subsp. *malaccensis*. The accessions ANS-15 and ANS-17, which belonged to the morphologically distinct Mt. Mugajah population, have for each of the tested loci at least one singleton haplotype that is somewhat related to subsp. *sumatrana* haplotypes. Several accessions had haplotypes that are distantly related (Table 6, Fig. 5) which might indicate an admixture among wild *M. acuminata* through cross pollination between subspecies. Remarkably five accessions of subsp. *halabanensis*, SSB-34, J-21, J-16, SSB-19, and SSB-33, share haplotypes that are identical or very closely related to those found in a cultivated banana “Pisang Jari Buaya” on four loci (Table 7).

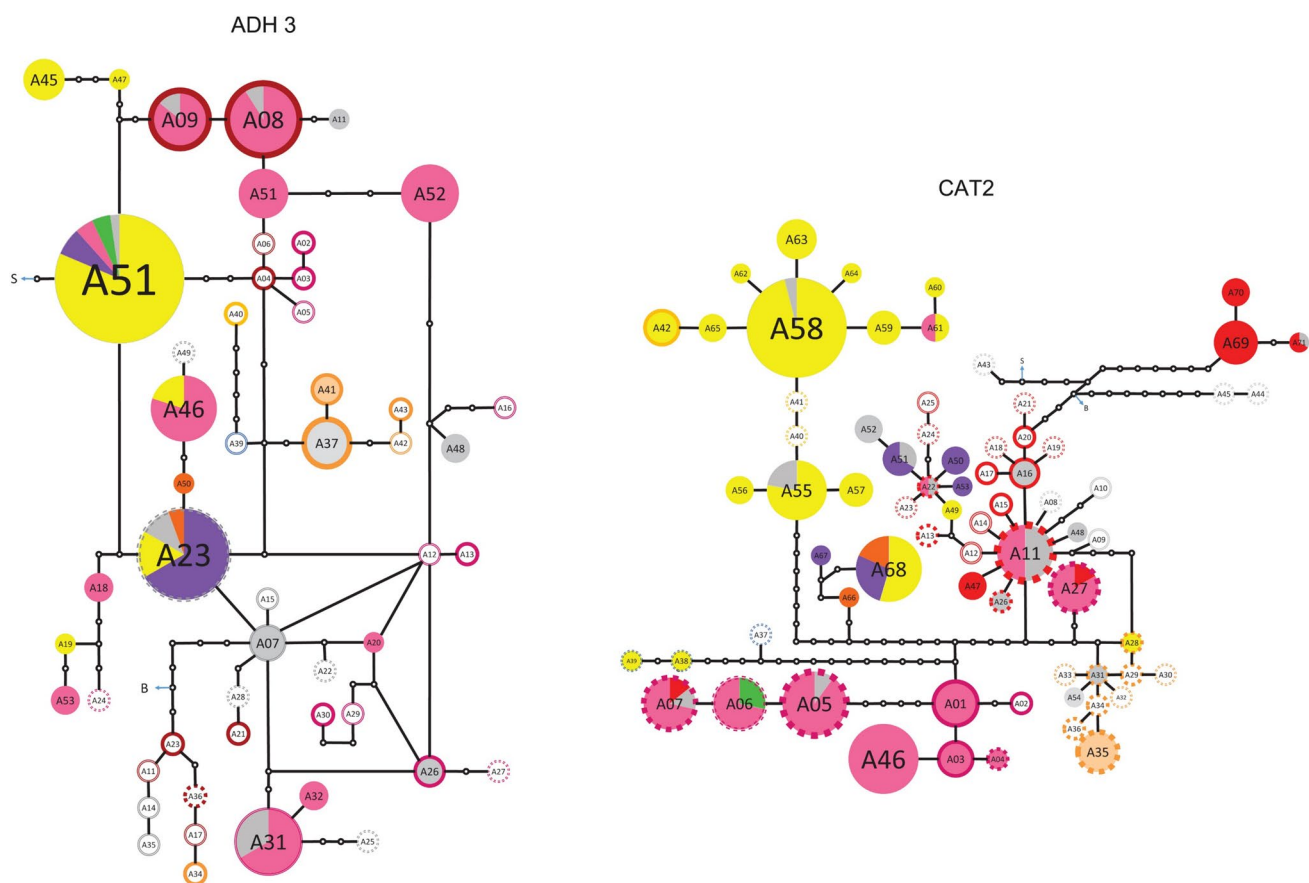


Fig. 5 Network and distribution of studied haplotypes of the ADH3, CAT2, GBSS1 and IDH1 loci among 69 accessions from Sumatra and 20 accessions from the BRIN collection of *Musa acuminata*. The colours representing subspecies/varieties/cultivars. The size of the

circles corresponds to haplotype frequency. Each line segment in the graph represents a mutational change. Small white circles indicate presumed intermediate haplotypes not detected among the studied accessions

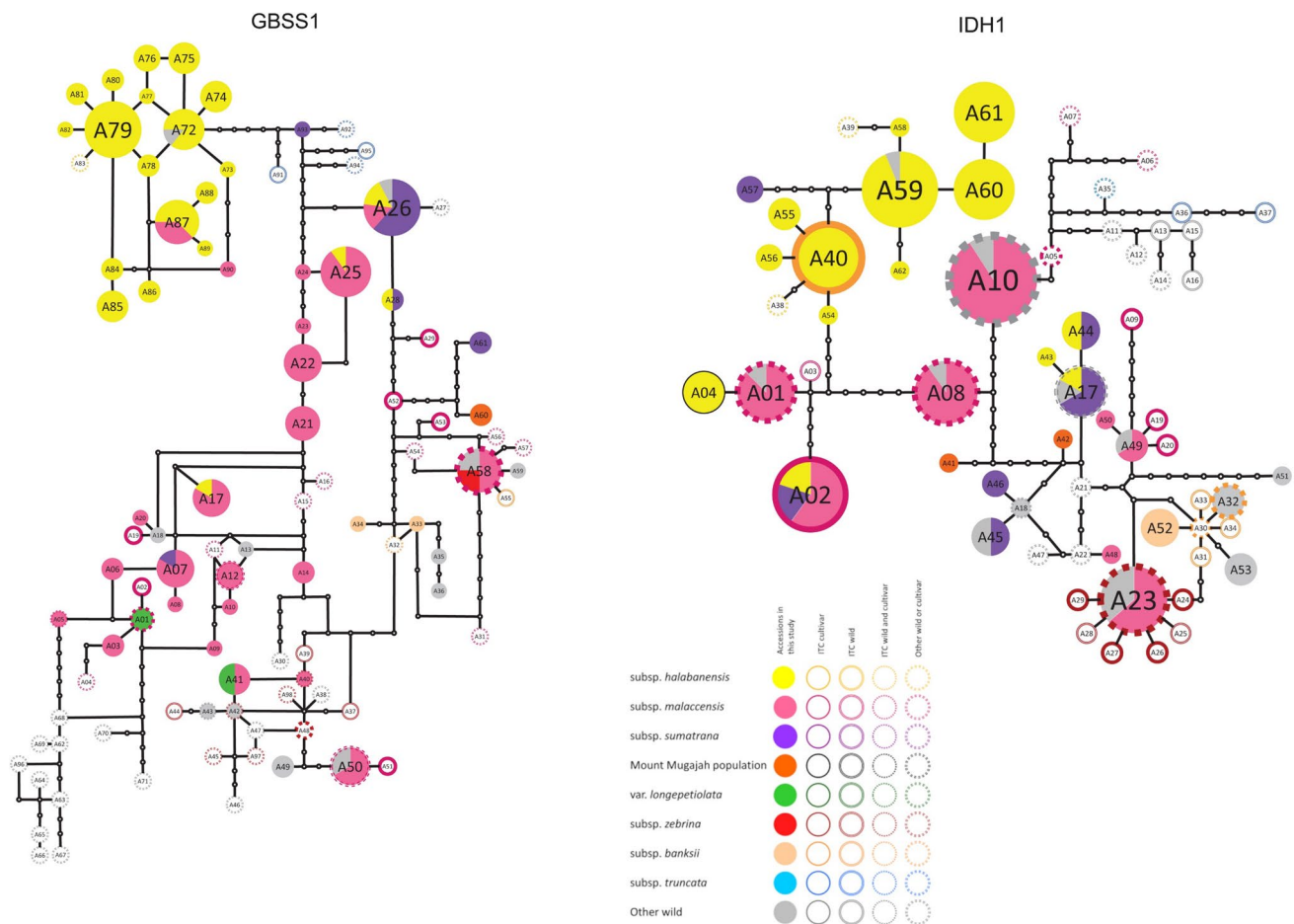


Fig. 5 (continued)

Discussion

In conservation biology, the identification of genetic diversity is crucial (Chung et al. 2020). Also, for crop improvement knowledge about genetic diversity is essential to ensure adequate representation in a breeding programme (Louwaars 2018). Export banana production represents perhaps the largest monoculture in contemporary agriculture with Cavendish clones globally dominating production (~ 50%) (FAO 2023) and the export trade (≥ 95%) (Voora et al. 2020). Emerging diseases could disrupt the banana production and trading chain as has happened in the past with the so-called “Panama disease” (Drenth and Kema 2021; Ploetz 2005). Therefore, we studied the genetic diversity of one of the progenitors of modern edible bananas, *M. acuminata* from Sumatra, Indonesia. The data represent an unparalleled exploration of genetic diversity in this species, providing a first picture of the spatial diversity across the largest island of the Indonesian archipelago. A previous study elucidated the

haplotype composition of wild and edible bananas from ITC collection and hypothesized that western Indonesia is important for the origin of cultivars (Volkaert 2011). Our study substantiates that hypothesis by investigating wild germplasm from Sumatra, which is underrepresented in (ex situ) genebanks as exemplified by just a single accession of *M. acuminata* subsp. *sumatrana* that was only recently introduced in the ITC (Ruas et al. 2017). Hence, we undertook an effort to explore the wild *M. acuminata* on Sumatra and describe its distribution and morphological and genetic diversity as a first step toward a conservation and deployment strategy.

Classification of the wild *Musa acuminata* of Sumatra

Nasution (1991) described two varieties from Sumatra with a top-shaped red coloured male bud, viz. var. *malaccensis* and var. *longepetiolata*, the main difference between the two being the very long petiole in the latter variety. We observed

Table 6 List of 91 accessions included in the multi-locus genotyping analysis and observed haplotypes for the four *ADH3*, *CAT2*, *GBSSI*, and *IDH1* loci

Accession	subsp./var./cv	ADH3			CAT2		GBSSI		IDH1
ANS-1	<i>malaccensis</i>	A07	A15	A01*	A27*	A20+	A25*	A10*	A10*
ANS-4	<i>sumatrana</i>	A17*	A20+	na	na	A07	A93+	A02*	A02*
ANS-5	<i>sumatrana</i>	A17*	A17*	A68	A68	A28	A28	A46	A46
ANS-8	<i>halabanensis</i>	A51	A51	A58	A59	A72	A73+	A59	A59
ANS-9	<i>halabanensis</i>	A51	A51	A60+	A61	A75	A75	A60	A60
ANS-15	Mt. Mugajah pop	A17*	A18+	A22*	A66+	A60	A60	A41+	A42+
ANS-17	Mt. Mugajah pop	na	na	A68	A68	na	na	na	na
ANS-20	<i>halabanensis</i>	A51	A51	na	na	A76	A88	A54+	A60
ANS-22	<i>halabanensis</i>	A51	A51	A55	A58	A74	A88	A60	A60
ANS-23	<i>sumatrana</i>	A17*	A51	A53+	A67+	A61	A61	A57	A57
ANS-25	<i>halabanensis</i>	A52	A52	A58	A58	A74	A74	A55	A55
ANS-29	<i>halabanensis</i>	A52	A52	A58	A58	A74	A87	A55	A59
ANS-31	<i>halabanensis</i>	A19	A19	na	na	na	na	A02*	A02*
ANS-32	<i>halabanensis</i>	A51	A53+	A55	A59	A75	A77+	A40	A40
ANS-34	<i>halabanensis</i>	A17*	A51	A59	A68	A28	A75	A18*	A59
ANS-35	<i>halabanensis</i>	A51	A51	A58	A58	A72	A76	A40	A40
ANS-37	<i>malaccensis</i>	A19	A19	A69	A70	A05*	A06	A02*	A02*
ANS-39	<i>malaccensis</i>	A19	A19	A69	A70	A07	A07	A02*	A02*
ANS-41	<i>malaccensis</i>	A19	A19	A69	A69	A07	A08+	A02*	A02*
J-1	<i>malaccensis</i>	A19	A19	A69	A71+	A07	A07	A02*	A02*
J-4	<i>malaccensis</i>	A07	A15	A27*	A46	A03	A10+	A10*	A10*
J-6	<i>malaccensis</i>	na	na	A6*	A46	A17	A22	A10*	A23*
J-8	<i>malaccensis</i>	na	na	na	na	na	na	A10*	A10*
J-12	<i>malaccensis</i>	A09*	A09*	A46	A46	A06	A87	A01*	A10*
J-14	<i>malaccensis</i>	na	na	A46	A46	A23+	A87	A08*	A10*
J-16	AA wild	A17*	A17*	A22*	A22*	A26	A26	A45	A45
J-19	<i>halabanensis</i>	A51	A51	A57	A57	A86	A86	A40	A40
J-21	<i>halabanensis</i>	A51	A51	A42*	A57	A26	A87	A40	A40
J-23	AA wild	A14	A14	na	na	A18+	A25*	A10*	A10*
J-26	<i>malaccensis</i>	A08*	A15	A46	A46	A25*	A87	A10*	A10*
J-27	<i>malaccensis</i>	A08*	A15	A05*	A27*	A22	A22	A08*	A10*
J-29	<i>malaccensis</i>	A09*	A28+	A01*	A05*	A17	A17	A10*	A10*
J-33	<i>malaccensis</i>	A08*	A15	A03*	A03*	A03	A17	A10*	A48+
RSB-01	<i>malaccensis</i>	A09*	A09*	A01*	A07*	A12*	A14	A02*	A10*
RSB-04	<i>halabanensis</i>	A51	A51	A58	A58	A87	A87	A59	A60
RSB-06	<i>halabanensis</i>	na	na	na	na	A76	A89+	na	na
RSB-09	<i>halabanensis</i>	A51	A51	A58	A62+	A72	A84	A59	A59
RSB-11	<i>halabanensis</i>	A17*	A51	na	na	A26	A87	A17*	A59
RSB-13	<i>sumatrana</i>	A17*	A17*	A68	A68	A26	A26	A44	A44
RSB-15	<i>halabanensis</i>	na	na	A65	A65	na	na	na	na
RSB-18	<i>malaccensis</i>	A09*	A54+	A03*	A58	A24+	A90+	A59	A59
RSB-19	<i>malaccensis</i>	A09	A09	A46	A46	A17	A25*	A08*	A08*
RSB-20	<i>malaccensis</i>	A08*	A15	A01*	A07*	A12*	A17	A08*	A10*
SSB-01	<i>malaccensis</i>	A07	A15	A04*	A27*	A22	A50	A01*	A01*
SSB-03	<i>malaccensis</i>	A21	A21	A05*	A06*	A21	A25*	A01*	A08*
SSB-06	<i>malaccensis</i>	A08*	A08*	A05*	A06*	A25*	A25*	A01*	A08*
SSB-10	<i>malaccensis</i>	A32	A32	A22*	A22*	A41*	A50	A23*	A50+
SSB-13	<i>malaccensis</i>	A31*	A31*	na	na	A41*	A50	A23*	A23*
SSB-19	<i>halabanensis</i>	na	na	A55	A55	A79	A79	A40	A40
SSB-22	<i>longepetiolata</i>	A08*	A08*	A06*	A06*	A01*	A01*	A04*	A04*
SSB-25	<i>halabanensis</i>	A51	A51	A56	A56	A85	A85	A60	A62+

Table 6 (continued)

Accession	subsp./var./cv	ADH3			CAT2		GBSS1		IDH1
SSB-29	<i>halabanensis</i>	A51	A51	A55	A58	A85	A85	A59	A59
SSB-30	<i>sumatrana</i>	A17*	A17*	A51	A51	A26	A26	A17*	A17*
SSB-33	<i>halabanensis-sumatrana</i>	A17*	A51	A51	A55	A26	A72	A17*	A59
SSB-34	<i>halabanensis</i>	A51	A51	A42*	A42*	A79	A84	A40	A40
SSL-01	<i>malaccensis</i>	A08*	A08*	A05*	A07*	A21	A21	A01*	A01*
SSL-04	<i>halabanensis</i>	A51	A51	A58	A58	A72	A79	A56	A56
SSL-10	<i>halabanensis</i>	A51	A51	A49+	A63	A79	A81	A43+	A61
SSL-11	<i>halabanensis</i>	A51	A51	A58	A58	A80	A80	A59	A61
SSL-14	<i>halabanensis</i>	na	na	A55	A58	A78	A81	A59	A59
SSL-15	<i>halabanensis</i>	A51	A51	A58	A58	A78	A79	A58+	A61
SSL-18	<i>halabanensis</i>	na	na	A55	A63	A72	A79	A61	A61
SSL-19	<i>sumatrana</i>	A17*	A17*	A50	A50	na	na	A17*	A17*
SSL-22	<i>halabanensis</i>	A51	A51	A58	A63	A72	A79	A60	A60
SSL-23	<i>malaccensis</i>	A31*	A31*	A11*	A47	A41*	A41*	A23*	A49
SSL-27	<i>halabanensis</i>	A51	A51	A58	A64+	A79	A82+	A60	A61
SSL-30	<i>malaccensis</i>	A31*	A31*	A11*	A47	A40*	A50	A23*	A49
SSL-33	<i>halabanensis</i>	A51	A51	A58	A63	A79	A79	A61	A61
SSL-34	<i>halabanensis</i>	A51	A51	A58	A58	A79	A79	A61	A61
CBN-1	<i>malaccensis</i>	A07	A07	A05*	A07*	A25*	A25*	A08*	A23*
CBN-3	<i>malaccensis</i>	A07	A09*	A05*	A07*	A12*	A21	A08*	A10*
CBN-4	<i>malaccensis</i>	A23	A23	A05*	A06*	A21	A22	A04*	A10*
CBN-9	<i>malaccensis</i>	A15	A22+	A06*	A46*	A22	A25*	A04*	A10*
PAN-01	<i>banksii</i>	A46*	A46*	A35*	A35*	A58*	A58*	A52	A52
PA-19	<i>banksii</i>	na	na	A35*	A35*	A58*	A58*	A52	A52
PA-79	<i>acuminata</i>	na	na	A11*	A16*	A58*	A58*	A32*	A32*
APH-192	<i>bantamensis</i>	A31*	A31*	A11*	A48+	A50	A50	A23*	A49
LIPI-218	<i>breviformis</i>	A31*	A35*	A11*	A16*	A49	A49	A23*	A23*
PNK-24	<i>flava</i>	na	na	A55	A58	na	na	na	na
PNK-26	<i>flava</i>	na	na	A71	A71	A33+	A34+	na	na
PAA-97	<i>halabanensis</i>	na	na	A68	A68	A26	A26	A44	A44
PAA-114	<i>halabanensis</i>	A51	A51	na	na	na	na	na	na
APH-258	<i>zebrina</i>	na	na	A11*	A11*	na	na	na	na
PAA-110	<i>sumatrana</i>	A17*	A17*	A68	A68	A26	A26	A45	A45
PH-01	<i>rutilifus</i>	A35*	A35*	A11*	A26*	A42*	A43*	A23*	A51+
PAN-26	<i>tomentosa</i>	A45*	A45*	A52	A52	A58*	A59+	A47	A47
PAR-100	<i>tomentosa</i>	A45*	A45*	A31*	A54+	A35+	A36+	A53	A53
APH-385	<i>zebrina</i>	na	na	A07*	A27*	na	na	na	na
LIPI-010	<i>malaccensis</i>	na	na	A03*	A05*	A09	A14	A01*	A10*
ANS-26	“Pisang DinginLidi”								
LIPI-232	“Pisang Rejang”	A09*	A26	A06*	A13*	A01*	A01*	A01*	A0*

Failed sequencing reactions for some accessions and loci are indicated with “na”. Haplotypes that are identical to those previously observed in accessions from the International Transit Centre and other accessions are indicated with *. Singleton haplotypes are marked with +

var. *longepetiolata* only at a few locations in a rather limited area of Southern Sumatra. Though plants belonging to the var. *longepetiolata* can be easily distinguished by their remarkably long petioles, neither the morphological (PCA) nor the haplotype data indicate that this variety would constitute a distinct genepool. Nevertheless, concluding that var. *longepetiolata* should be revised and merged altogether into

subsp. *malaccensis* is premature, since only very few accessions of var. *longepetiolata* were assessed in this study.

One group of wild bananas could be easily distinguished from all others on Sumatra by its purple coloured male bud and strikingly globular seed. Nasution (1991) described two varieties with globular seed but differing in male bud shape. One, var. *halabanensis*, had a rounded male bud, while the

Table 7 A comparison of the genotype of banana cultivar “Pisang Jari Buaya” with selected accessions of wild *Musa acuminata* from Sumatra having matching or very closely related genotypes

Accession	Origin	ADH3		CAT2		GBSS1		IDH1		Remarks
Pisang Jari Buaya	Indonesian cv	17	33	36	42	29	83	20	40	Volkaert pers obs
SSB34	South Sumatra	51	51	42	42	79	84	40	40	subsp. halabanensis
J21	Jambi	51	51	58	58	26	87	40	40	subsp. halabanensis
SSB-19	South Sumatra	–	–	55	55	79	79	40	40	subsp. halabanensis
SSB-33	Bengkulu	17	51	51	55	26	72	17	59	subsp. halabanensis
J16	Jambi	17	17	22	22	26	26	45	45	AA wild

The haplotypes identical to the “Pisang Jari Buaya” or those differing by a single substitution (GBSS1) are indicated in bold

other, var. *alasensis*, had a lanceolate male bud. The var. *alasensis* was described by Nasution (1991) from samples collected in the Alas River valley in Aceh. We did not find any wild *M. acuminata* with globular seed and lanceolate male bud during our resampling in that area. However, we observed that at an early stage the male bud of the wild bananas in the area is lanceolate which turns into a more rounded shape later on (Online Resource 9). According to the guidelines in the “Descriptors for banana (*Musa*)” (IPGRI 1996) the description of the male bud should be done at a later stage. Hence, we speculate that Nasution (1991) described var. *alasensis* from an early stage of the male bud. Our genotyping data also supports the conclusion that Nasution’s (1991) var. *alasensis* (accession maintained in the BRIN collection) should be placed in subsp. *halabanensis* as they share a number of haplotypes that are well differentiated from those of other subspecies.

In nearly all of the 38 observations of subsp. *halabanensis* in their natural environment we noted undeveloped fruits in one or more of the early hands while later hands were fully developed (Online Resource 7). This was similarly reported by Itino et al. (1991), but not substantiated with photographs. Their drawing seems to indicate that only the fruits of the hands at the underside of the (sub)horizontal bunch remained undeveloped. In our exploration, we found that a variable number of the early hands in bunches did not develop but this was independent of the orientation of the hand on the rachis. The incomplete development of fruits could result from a lack of pollination or from (partial) sterility without parthenocarpic fruit development. A lack of pollination is not very likely since the populations of wild bananas were generally quite large with several individuals flowering simultaneously with either male or female flowers. Sterility is common in cultivated or hybrid bananas, because of the presence of chromosomal translocations and/or inversion in heterozygous bananas leading to unbalanced chromosome segregations at meiotic division (Ahmad et al. 2020), resulting in aneuploidy and gamete lethality (Dodds 1943; Shepherd 1999; Wilson 1946a, b, c). Cultivated bananas in general are sterile, but the presence of parthenocarpy drives the fruit development (Simmonds 1953). Hence, cytogenetic

observation of the meiotic stages of pollen mother cells and pollination experiments could shed light on the cause of underdeveloped fruits. Since sterility due to cytogenetic irregularities would occur in all developing flowers in a bunch, as in the cultivated bananas, the undeveloped fruits in early hands observed here is most likely due to unknown developmental or physiological factors that require further study.

The cracked peel in subsp. *halabanensis* (Online Resource 8) probably results from a similar mechanism as the fruit peel splitting in *M. schizocarpa*. As the observation of the fruit splitting depends on the right stage of fruit development and may also be influenced by the environment, it is presently unclear whether this phenomenon would be characteristic for the subspecies as a whole or is limited to only some local populations as a result of genetic differentiation or environmental factors.

We sampled eight plants that matched Nasution’s (1991) description of var. *sumatrana*. They were easily recognisable by the pendulous bunch with long fruits, dark colour of the male floral bud and flat seed with rough edges. Nasution (1991) classified this type as a variety, but as we found this banana across five out of seven provinces of Sumatra at high elevation and with a constant morphology, we consider it as a subspecies. This is supported by both the PCA based on morphological data and by the haplotype analyses. Although the initial flowers are bisexual and self pollination would be possible, the molecular analysis indicates that this is not exclusively so since the subsp. *sumatrana* population maintains a level of diversity only slightly lower than that of the other subspecies that are non-hermaphroditic. Indeed, the subsp. *sumatrana* is genetically not completely isolated from the surrounding plants belonging to other subspecies. Aside from several unique haplotypes, the subsp. *sumatrana* individuals contained haplotypes that belong to the larger clusters of subsp. *malaccensis* and subsp. *halabanensis* haplotypes.

At three locations close to each other but all at high elevations on the southern side of Mount Mugajah, Central Aceh, we found a population of unusual bananas, i.e., accessions ANS-15, ANS-16 and ANS-17. These plants were distinct

due to a rounded male bud with clearly imbricated bract tips, the bracts lifting rather than rolling back, and the long pedicel of the fruit. Among the limited molecular data, we found five unique haplotypes. However, it is premature to describe it as a new subspecies or variety due to the limited sample size. Structure analysis with singletons merged into closely related haplotypes consistently grouped the single accession with subsp. *sumatrana*. It will require more in-depth sampling and analyses to determine whether this is an expression of environmental plasticity or a true genetic variation.

Among the 65 morphological traits that were recorded, seed morphology seemed to be a very robust characteristic for the identification of subsp. *sumatrana* which has flattened seed with rough edges and subsp. *halabanensis*, which has smaller globular seeds, differentiating them from all other wild *M. acuminata* whose seeds are irregularly angular and slightly depressed with a smooth or minutely tuberculate surface. According to Nasution (1991) an angular seed with wrinkled surface is common among wild *M. acuminata*, which he described in 13 *M. acuminata* varieties underscoring the importance of regenerative organs in traditional plant taxonomy (Ornduff 1978).

Haplotype assignment

More than two-thirds of the observed haplotypes across the four loci (104/142) had not been detected before (Table 3). Several of these new haplotypes occurred in just one individual, occasionally homozygous but mostly heterozygous. The fact that so many unique haplotypes have been detected in this study may raise the question whether spurious sequencing errors have been interpreted as mutations. In this investigation, we employed direct sequencing from PCR products initiated by specific internally positioned primers to prevent PCR recombination and other cloning artifacts. This approach ensured comprehensive coverage of most loci through sequencing in both directions, with thorough scrutiny for technical errors in both sequencing reactions. Actually, there might be even more polymorphic positions than we report as a number of ambiguous positions that would create additional singleton haplotypes were disregarded. Every polymorphic position that has been retained was either unambiguous on its own or confirmed by its presence in more than one genotype. Such large numbers of alleles in a species are not exceptional for natural populations of tropical plants (Wattanakupakin et al. 2015). Even if a few polymorphic positions have been erroneously included or wrongly assigned to a haplotype by the phasing algorithm, they will not significantly impact the network structure underlying the grouping into subspecies clusters.

The haplotype networks contain some level of homoplasy, especially for GBSS1 and ADH3. Aside from the haplotype

ascertainment difficulties because of the presence of polymorphisms that were observed only once, homoplasy may be due to past recombination between divergent alleles or parallel mutations and reverse mutations.

Molecular analysis supports the subspecies grouping

Nasution (1991) classified Indonesian wild *M. acuminata* according to morphological characteristics based on his explorations and specimens in the Herbarium Bogoriense. Bananas are known to have highly variable phenotypes depending on the environmental and seasonal conditions. Thus, characterization based on morphology alone may lead to spurious naming of forms as exemplified by the banana cultivars (Häkkinen 2013). For instance, subsp. *sumatrana* accessions grown in the BRIN gardens at low elevation (160 m asl) in Cibinong, West Java are morphologically very similar to subsp. *malaccensis* or subsp. *zebrina* (Online Resource 10). Molecular studies can confirm or correct the classification based on morphology (Rana et al. 2014; Soltis and Soltis 1998, 2000). To support the morphological observations, we used multi-locus genotyping to determine the haplotypes and genotypes of wild banana accessions across Sumatra. In the parsimony networks the haplotypes clustered largely according to the morphological subspecies assignments independent of their geographical origin.

The clustering based on the four loci sequences clearly separated the subsp. *halabanensis* and subsp. *sumatrana* accessions from subsp. *malaccensis* even though the Evanno et al. (2005) method to determine the correct number of subpopulations indicated the presence of just two subpopulations. Puechmaile (2016) demonstrated that the popular Evanno et al. (2005) method generally underestimates the appropriate number of clusters when sampling is unequal among subpopulations. Consequently, the Puechmaile (2016) parameters indicated the presence of potentially five subpopulations. Assuming the existence of a larger number of clusters revealed two additional subpopulations within plants classified as subsp. *malaccensis*. The separation of these accessions seems not based on any individual morphological characteristic and is discordant with the single var. *longepetiolata* individual in the analysis. However, it is not entirely unexpected as the haplotype networks indicated a very large diversity among the subsp. *malaccensis* and subsp. *zebrina*-derived haplotypes. A separation of the Southern accessions of the subsp. *halabanensis* from the others might indicate additional structure in that subspecies. However, any conclusions about separating clusters beyond the three subspecies should be interpreted as preliminary given the small number of loci and the limited number of accessions analysed.

In this study, we propose to merge two varieties, var. *halabanensis* and var. *alasensis* (Nasution 1991), into the validly described subsp. *halabanensis* (Meijer) Hotta. We justify this by the resampling in locations where var. *alasensis* was described by Nasution (1991), the Alas River valley, but additionally on the haplotype analysis. The haplotypes detected among all wild *M. acuminata* with globular seed, i. e. subsp. *halabanensis*, mostly grouped in one single cluster for each of the four loci indicating that this population is genetically quite distinct. Hence, across Sumatra these plants are more closely related to each other than to other subspecies that may be growing in nearby locations. However, haplotypes that typically occur in other subspecies can occasionally also be found in plants that otherwise concur to the subsp. *halabanensis* morphology. Therefore, the subsp. *halabanensis* gene pool is not separated from the other *M. acuminata* subspecies and thus a classification at the species level as initially proposed by Meijer (1961) is rejected. Describing these plants as *M. acuminata* subsp. *halabanensis* as suggested by Hotta (1989) seems to be the most reasonable classification given that these populations have a unique seed morphology and contain a cluster of characteristic haplotypes at each of the studied loci.

In addition to some unique haplotypes, the subsp. *sumatrana* accessions shared haplotypes with subsp. *halabanensis*, subsp. *malaccensis* and subsp. *zebrina*. *Musa acuminata* subsp. *sumatrana* generally grows at higher elevations whereas within the same area the other subspecies grow mostly at lower elevations. The occurrence of these subspecies in close proximity can explain the sharing of haplotypes through admixture one or several generations earlier. Though the barriers to gene flow are not absolute, the distinct morphology and ecology together with the presence of several private haplotypes justifies our decision to consider these bananas a distinct subspecies.

For the distinct Mount Mugajah population the limited DNA sequence data that has been obtained indicated the presence of unique haplotypes at all four loci, i.e., ADH3-A45, CAT2-A66, GBSS1-A60, IDH1-A41 and IDH1-A42 suggesting that it would be genetically distinct to other *M. acuminata* subspecies or varieties. However, all these haplotypes are rather close to the subsp. *sumatrana* haplotypes in the parsimony networks. Also, the structure analysis cannot differentiate the single individual included from the subsp. *sumatrana* individuals. More accessions of this type will have to be analysed to understand the genetic or environmental basis for their unique morphology before we can describe this type as a new infraspecific entity, be it as subspecies or variety.

The haplotypes of the subsp. *malaccensis* and var. *longepetiolata* accessions are distributed in the centre of the networks among those of subsp. *malaccensis* and subsp.

zebrina from other regions. This result accords with Volkaert (2011), who observed that subsp. *malaccensis* has a high genetic diversity and does not form a specific cluster in the network, which might indicate a polyphyletic origin for this subspecies as currently described based on morphology.

The observed admixtures at the subspecies level should not be surprising given the sometimes very patchy but intermingled occurrence of the different morphological types of wild bananas. It might even be visible in the morphology as in accession SSB-33, which had a fruit that was intermediate between subsp. *sumatrana* and subsp. *halabanensis* and has haplotypes of subsp. *halabanensis* admixed with others that cluster with subsp. *sumatrana* or subsp. *malaccensis* across the four loci studied. The occurrence of such intermediate morphologies was already reported by Nasution (1991).

Taken together, the haplotype composition of accessions can support strategies for the conservation of diversity of wild bananas. When in a highly diverse species like *M. acuminata* distinct clusters of haplotypes can be identified, they can guide the selection of type specimens that contain haplotypes exclusive to a single cluster and base the morphological description on those individuals.

Population genetics of wild *Musa acuminata* of Sumatra

Highly significant deviations from HWE were observed for all four loci whether the analysis was done for all wild accessions, only accessions from Sumatra, or for the individual subspecies within Sumatra. The high number of identified haplotypes and the comparatively modest number of analysed accessions resulted in a very large number of diploid genotypes occurring at very low frequencies and many potential genotypes completely missing, making the significance testing for HWE less reliable. The grouping of singleton haplotypes with their closest relatives reduced the number of haplotypes and genotypes but only marginally decreased the deviations from HWE. The clear deficiencies of heterozygotes may indicate that there are genetically isolated subpopulations between but also within each subspecies. The different islands of the Indonesian archipelago have been isolated and joined during glacial cycles (Husson et al. 2020) and thus it might be expected that many islands have distinct subpopulations. Our finding that there are highly significant deviations from HWE even within each subspecies on a single island indicates the presence of more fine-scaled genetically distinct subpopulations.

Since we found a significant deviation from HWE for each locus, we investigated whether PCNM, also known as Moran's Eigenvector maps, which analyses diversity data in combination with (geographical) distance data, could give some insights on the scale of population structure. However, the sampling of accessions was not designed with this

purpose in mind and thus the dataset is suboptimal for such an analysis. Because of the large number of haplotypes and genotypes at very low frequency and only four loci investigated, the genotype data could not be used. Analysing the individual SNP positions is not a solution either as the SNPs are tightly linked and thus do not represent independent data. Therefore, only the phenotypic data could be analysed. We chose to use Euclidian geographical distances as they can be easily calculated from the GPS data but that may not be the most appropriate distance measure. The mountainous topography of Western Sumatra and the occurrence of the subsp. *malaccensis* and subsp. *halabanensis* being restricted to lower elevations below 650 asl with subsp. *sumatrana* higher up above 600 asl (Online Resource 5), would indicate that the distances along river valleys might be more appropriate. Despite these caveats, some preliminary analyses such as the PCNM graphs indicate the occurrence of subpopulations that would accord with the complex topography of Sumatra. Hence, each watershed, each valley, may have its unique subpopulation of bananas for each subspecies, locally interbreeding but somewhat isolated from the populations of the other parts of the island. Consequently, further studies of the wild banana population structure on Sumatra or other islands should adopt a sampling strategy across the various topographies and include many more accessions.

Origins of edible bananas

Tracing the origin of various crops to their wild ancestors has revealed much about their domestication history but also has practical implications for breeding. The genetic origin of cultivated bananas has however proven quite elusive thus far.

A high level of heterozygosity among cultivated bananas has been revealed by various co-dominant markers such as SSRs (Creste et al. 2004; Retnoningsih et al. 2010) or RFLPs (Carreel et al. 2002). By exact matching of SSR fragment sizes potential diploid progenitor genotypes of the triploid Cavendish and Gros Michel banana cultivars could be identified (Raboin et al. 2005). However, the advantage of DNA sequence-based analyses over fragment size matching or even genome-wide SNPs is that not only exactly matching patterns are informative, but because the evolutionary relationships among haplotypes can be determined, even imperfect matches give clues to their origins. Previously, several studies have indicated that most cultivated bananas have originated from hybridisations between different subspecies mostly involving subsp. *malaccensis*, subsp. *zebrina* and subsp. *banksii* (Li et al. 2013; Martin et al. 2020; Volkaert 2011). However, it has proven to be very difficult to assign more specific genetic origins for the majority (Hippolyte et al. 2012). Recent efforts using whole genome analyses (Sardos et al. 2022; Martin et al. 2020) revealed multiple origins for cultivars and also indicated the existence

of "cryptic genepools". These are observed variants found in cultivars that thus far have eluded identification among wild relatives. One of the cultivar groups that contains a clear signal of an unknown genepool contribution is the "Pisang Jari Buaya" group (Volkaert 2011; Martin et al. 2020). A study of a few accessions of wild bananas from the BRIN collection in 2014 (unpublished) indicated that two accessions from Sumatra contained haplotypes almost identical to the haplotypes found in this cultivar group. It was hoped that a further collection and analysis of Sumatran wild bananas might reveal more details about their origin. The determination of haplotypes in wild and cultivated bananas can be of great value in determining the original location of the origin of banana cultivars. We demonstrate this by the analysis of the haplotypes present in the "Pisang Jari Buaya" cultivar. The present analysis provides data that support a likely contribution from a wild banana population from South Sumatra to this group of cultivars. The unique "Pisang Jari Buaya" haplotypes for the CAT2 (A42) and IDH1 (A40) loci were identified in the subsp. *halabanensis* accessions SSB-34 and J-21. For the GBSS1 locus the haplotype exactly matching "Pisang Jari Buaya" (A83) was not detected in the wild *M. acuminata* accessions but other highly similar haplotypes (A79, A84, and A87) were observed, all fitting in the subsp. *halabanensis* cluster. In addition, accession J-21 has a GBSS1 haplotype belonging to the subsp. *zebrina* cluster (A26) which is just one single mutation different from the GBSS1 haplotype found in "Pisang Jari Buaya" (A27). This is a strong indication that subsp. *halabanensis* has contributed as a parent of "Pisang Jari Buaya" and that the origin of this cultivar may be in Southern Sumatra, Jambi, Bengkulu, or Lampung provinces at one of the various locations where subsp. *malaccensis* and subsp. *halabanensis* populations are growing in close proximity. Note that according to the structure analysis the subsp. *halabanensis* accessions from Southern Sumatra might be slightly distinct from those elsewhere. However, the haplotypes of ADH3 are still distinct between "Pisang Jari Buaya" and any of the subsp. *halabanensis* accessions suggesting that our sampling depth for molecular analysis was too sparse to detect the contemporary relatives of the population that may have given rise to this cultivar. Conclusion we can not only determine the subspecific composition of PJB, but we can be quite certain that it originated from a population in southern Sumatra.

The discovery of haplotypes in cultivars that cannot be found in wild accessions (Volkaert 2011) and hints of "cryptic gene pools" from clustering analysis based on genome-wide variants (Martin et al. 2020) indicate that there are gaps in the germplasm collections of bananas. Only a single accession of subsp. *sumatrana* and subsp. *halabanensis* have recently been introduced to the germplasm collection at ITC. Consequently, the detection of a large fraction of the haplotypes (104 out of 142) among wild bananas from

Sumatra that thus far have never been found among accessions maintained by ITC is not surprising. It confirms that the representation of the wild genotypes in the ITC is far from complete. Even though there are accessions of wild subsp. *malaccensis* available through the ITC collection, new haplotypes were detected among Sumatran samples of this subspecies. The large majority of wild germplasm in the ITC collection originated from explorations in and around Papua New Guinea, which turns out to be relatively poor in genetic diversity. The representation from other regions and subspecies is limited and mostly not well documented. Another indication for the missing diversity is the detection of several unique haplotypes among just two accessions of var. *tomentosa* collected from Sulawesi and maintained in the BRIN germplasm collection. As long as substantial germplasm collections from other regions of the wild *M. acuminata* are not included, molecular studies attempting to establish relationships between cultivated bananas and their wild relatives will be either inconclusive or in the worst case even misleading.

Because of the high levels of heterozygosity found in cultivated bananas, previous studies have indicated that most might have originated from hybridisations between different subspecies mostly involving subsp. *malaccensis*, subsp. *zebrina* and subsp. *banksii* (Li et al. 2013; Martin et al. 2020; Volkaert 2011). In our exploration of the diversity among the Sumatran wild bananas we discovered many wild plants that have genotypes indicating admixture between various subspecies, i.e. subsp. *malaccensis* with subsp. *halabanensis* or subsp. *sumatrana* with subsp. *halabanensis*. This should not be a surprise given the intermixing mosaic of morphologically distinct subspecies observed at several locations. This also implies that the mountainous topography of Sumatra lends itself very well to natural admixture. Thus, the admixed genotypes observed in cultivated bananas could well have happened naturally at various places where the individual types occur in sufficiently close proximity. Contrary to some of the main hypotheses postulated for the origin of edible bananas (Perrier et al. 2009), there would not be any need for humans facilitating the admixture by transporting semi-domesticated cultivars from one location to another.

None of the other main islands in Indonesia, such as Java, Kalimantan, or Sulawesi, have been explored in much detail, leaving the possibility open for other major taxonomic discoveries that could further our understanding of the banana diversity and its deployment in breeding. Therefore, exploring wild *M. acuminata* and other ancestors of contemporary edible bananas across the Indonesian archipelago as well as the neighbouring regions is important from both a conservation and breeding perspective.

Conclusions

The present study contributes to the knowledge of the diversity of wild *M. acuminata* of Sumatra by combining morphological variations at the infraspecific level with multi-locus genotyping data, resulting in a fine-tuned diversity with consequences for taxonomy. This study revealed that the diversity of wild *M. acuminata* of Sumatra comprises five types i.e., subsp. *malaccensis*, subsp. *halabanensis*, subsp. *sumatrana*, var. *longepetiolata* and one type that has never been described and will need further study to classify it into the *M. acuminata* group. According to the wide distribution on the Sumatra, subsp. *sumatrana* considerate as subspecies. Moreover, the molecular data show that many haplotypes are not present in the international collections and hence are inaccessible for diversifying breeding programs. Yet, the haplotypes of several wild accessions are similar to the haplotypes identified in contemporary banana cultivars, suggesting that indeed, some edible bananas may have originated from Sumatra. This finding indicates that the genetic knowledge of Indonesian bananas is incomplete. Further exploration of wild ancestors of banana cultivars across diverse topographies of the Indonesian archipelago is important from a conservation standpoint, particularly with an eye on competing claims for agricultural or urban land use.

Taxonomic treatment

Key to the infraspecific taxa of *Musa acuminata* on Sumatra

- 1a Length of petiole less than half of the lamina length 2
- 1b Length of petiole half or more of the lamina length.....
..... **var. *longepetiolata***
- 2a Male bud ovoid, like a top or lanceolate, tip not imbricate
..... 3
- 2b Male bud rounded, tip imbricate.....
..... **Mount Mugajah population**
- 3a Seed angular 4
- 3b Seed globular, smooth **subsp. *halabanensis***
- 4a Seed wrinkled **subsp. *malaccensis***
- 4b Seed flat with rough edge **subsp. *sumatrana***

Musa acuminata* subsp. *malaccensis (Ridl.) N.W.Simmonds. Kew Bull. 11: 466. 1957, "1956". ≡ *Musa malaccensis* Ridl., Trans. Linn. Soc. London, Bot., Botany ser. II. 3: 385. 1893. ≡ *Musa acuminata* Colla var.

malaccensis (Ridl.) Nasution, Mem. Tokyo Univ. Agric. 32: 75. 1991.—TYPE: Malaya, Pahang, Tanjong Gajah Mati, 1891, *H.N. Ridley s.n.* (lectotype: SING 062891!, designated by Nasution [1991: 45]).

Diagnosis: Male bud like a top, bract shape medium to large shoulder, apex slightly pointed or pointed, slightly waxy, outer bract overlapping younger ones at apex of bud, external bract red to red purple sometimes with yellow streaks, internal bract whitish to red, apex acute, sometimes tinted with yellow, lifting one or rarely two or more bracts at a time, revolute before falling; Seeds angular and wrinkled (Online Resource 11).

Description: Plants are suckering freely, vertical. The mature pseudostems are 2–4.5 m tall, 15 up to 30 cm in diameter at the base, covered with dead remains of older sheaths; sheaths are light green-yellow with reddish, brown, purplish, or blackish blotches; underlying colour cream; very little to moderately waxy and the sap milky or watery. Petioles are 40–60 cm long, the petiole canal is straight with erect margins or margins curved inward, and the petiole margin colour is green, pink-purple to red or purple to blue. The lamina are lanceolate to oblong, 150–250 × 35–70 cm, with a asymmetric or symmetric base, both sides rounded or one side rounded, one side pointed or both sides pointed, apex truncate, abaxial light green with waxy or nonwaxy, adaxial green. The colour of the abaxial midrib is mostly yellow or green, and the colour of adaxial midrib light green. The inflorescence peduncle is 20–60 cm long, 3–12 cm in diameter, and the colour is green to dark green with brown blotches, pubescent. The rachis is bare, and mostly falling vertically or sub-horizontal. The male bud is like a top, 5–9 × 10–18 cm. The bract shape is medium to large shoulder, apex slightly pointed or pointed, slightly waxy, 5–9 × 10–18 cm. The outer bract overlapping the younger ones at apex of bud, the external bract is red to red purple sometimes with yellow streaks. The internal bract is whitish to red, with an acute apex, sometimes tinted with yellow, lifting one or more bracts at a time, revolute before falling. The initial flowers carpellate is with angular ovary, yellowish green and the ovules is in four rows per locule. The stigma is flat, 6 × 6 mm, greyish. The style is 3–3.2 cm long, creamy. The compound tepals are white or cream, with 2 prominently thickened light yellow keels, and with deep yellow and curved outward apex. The free tepal is ovate, translucent white, apex acute, little or no visible sign of the tip development. The anthers are in the same level of the lobes on the compound tepal. The male flower has five creamy filaments of staminode. The later flowers pistillate with white, cream, rusty brown or pinkish anther lobes. The ovary is creamy, without additional pigmentation. The stigma is creamy. The fruit bunch is horizontal, compact,

and has 5–9 hands. The fruits is in two rows. Individual fruits are straight or straight in the distal part, up to 10 cm long (including pedicel), 1–1.5 cm in diameter, prominently ridged, and the apex is obtuse. The pedicel is up to 1 cm long, and straight. The immature fruit peel is green, and not waxy. The fruit pulp is white, containing up to 180 seeds per fruit, which are angular, wrinkled, 3.1–6.6 mm long, 3.7–5.4 mm wide, 3.6 mm depth.

Distribution: This subspecies was found in North Sumatra, Jambi, Riau, Bengkulu, South Sumatra, and Lampung province in habitats at forest borders at open areas, open canyons, river beds, villages, palm oil and rubber plantation borders and along roadsides.

Specimens examined: Mount Kerinci, Jambi province, *Bunne Meyer 10346* (3 sheets) (BO); Palembang city, South Sumatra, 2 Nov 1929, *C.G.J. van Steenis 3950* (4 sheets) (BO); Asahan regency, North Sumatra province, *H.S. Yates 1690* (2 sheets) (BO); Mount Kerinci, 4 Apr 1954, *A.H.G. Alston 14236* (1 sheet) (BO); Jalan H. Makam, Bongko district, Jambi province, 12 Apr 2017, *F. Ahmad FA-J 12* (4 sheets) (BO); Desa Muara Kilis, Tebo Hilir, Muara Tebo district, Jambi province, 14 Apr 2017, *F. Ahmad FA-J 29* (3 sheets) (BO); Jejawi district, Okan Komerang Ilir regency, South Sumatra province, 25 Apr 2017, *F. Ahmad FA-SSB 10* (4 sheets) (BO); Batu Raja district, Prabumulih city, South Sumatra province, 18 May 2017, *F. Ahmad* and *Y.S. Poerba AP-SSL 3* (4 sheets) (BO); Desa Siabu, Salo, Bangkang district, Kampar regency, Riau province, 8 May 2017, *F. Ahmad* and *Y.S. Poerba AP-RSB 01* (4 sheets) (BO); Kuantan Mudik district, Riau province, 12 May 2017, *F. Ahmad* and *Y.S. Poerba AP-RSB 21* (4 sheets) (BO); Desa Seberida, Batang Dangkal district, Indra Giri Hulu regency, Riau province, 13 May 2017, *F. Ahmad* and *Y.S. Poerba AP-RSB 22* (5 sheets) (BO); Jalan Raya Pesisir, Desa Banding, Raja Basa district, Lampung Selatan regency, Lampung province, 22 May 2017, *F. Ahmad* and *Y.S. Poerba AP-SSL 30* (4 sheets) (BO).

Musa acuminata* subsp. *halabanensis (Meijer) Hotta, Occas. Pap. Kagoshima Univ. Res. Center S. Pacific 16: 68. 1989. ≡ *Musa halabanensis* Meijer, Acta Bot. Neerl. 10: 250. 1961. ≡ *Musa acuminata* Colla var. *halabanensis* (Meijer) Nasution, Mem. Tokyo Univ. Agric. 32: 51–53. 1991.—TYPE: Indonesia, North Sumatra, Pematang Siantar, 9.III.1958, *W. Meijer 7264* (holotype: L [web!]).

= *Musa acuminata* Colla var. *alasensis* Nasution, Mem. Tokyo Univ. Agric. 32: 48–51, Fig. 19. 1991.—TYPE: Indonesia, Aceh, Kotacane, Ketambe, Rusdy 1638 (holotype, BO!).

Diagnostis: Male bud ovoid; bract shape medium shoulder, apex mostly intermediate or obtuse, moderately waxy, outermost bract overlaps at apex of bud or sometimes slightly imbricate, external bract purple–brown, internal bract whitish, apex obtuse, not tinted with yellow colour, lifting one or two or more bracts at a time, revolute before falling; Initial hand or hands of the bunch generally partly or not developed; Seed globular, mostly smooth (Online Resource 12).

Description: Plants suckering are freely. The mature pseudostems are 2–4 m tall, 15 up to 30 cm in diameter at the base, covered with dead remains of older sheaths. Sheaths are mostly brown-black, black-purple, green-yellow with brown–purplish or black–purplish blotches, underlying colour cream, very little or no visible sign of wax, and the sap is milky or watery. The petioles are 40–60 cm long, the petiole canal is straight with erect margins, and the petiole margin is colour pink-purple to red or purple to blue. The lamina of the leaves are lanceolate to oblong, 150–250 × 35–70 cm, with a asymmetric or symmetric base, both sides rounded or one side rounded, one side pointed or both sides pointed, truncated apex, the colour of abaxial is light green and shiny, and adaxial green and shiny, colour of abaxial midrib mostly yellow or light green with pink-purple or purple to blue, and with light green blotches of adaxial midrib. Inflorescences initially horizontal rather abruptly turning down at male bud stage. The peduncle is 20–40 cm long, 3–7 cm in diameter, and the colour is green to dark green or with brown blotches, slightly pubescent to very pubescent. The rachis is bare, and mostly with a curve or falling vertically. The male bud is ovoid, 7–9 × 9–15 cm. The bract shape is medium shoulder, apex mostly intermediate or obtuse, moderately waxy, ca. 15 × 7 cm, mostly not imbricate. The old bract is overlap at apex of bud or slightly imbricate, the colour of external bract is purple–brown, the colour of internal bract is whitish, apex obtuse, not tinted with yellow colour, lifting one or more bracts at a time, revolute before falling. The basal flower is carpellate, and prominently ridged, with creamy or whitish colour. The Ovules are in four rows per locule. The stigma is flat, 6 × 6 mm, greyish. The style is 3–3.2 cm long, creamy. The compound tepals of male flowers are white or cream, with 3 prominently yellow lobes, and the apex is deep yellow and curved outward. The free tepal is ovate, with translucent white, and the apex is acute with a developed sign. The anthers are in the same level of the lobes on the compound tepal. The male flower has five creamy filaments of staminode, and the anther lobes white or creamy. The ovary is creamy, without additional pigmentation. The stigma is yellow. The fruit bunch orientation is horizontal, with 7–10 hands, which fully developed with seed, but sometimes initial hands or even most hands not developing into fruit. The bunch is mostly lax if the fruit is

not developed to very compact if the fruit is developed and bearing seeds. The fruits are in two rows, 14–18 individual fruits per hand. Individual fruits are straight, up to 10 cm long (including pedicel), 1–1.5 cm in diameter, with prominently ridged, and the apex is bottle-necked, slightly sessile. Immature fruit peels are green, not waxy, and the fruit pulp is white, containing up to 300 seeds to per fruit, which are globular, mostly smooth, 3.1–6.6 mm long, 3.7–5.4 mm width, 3.6–5 mm depth.

Distribution: This subspecies was found in most provinces of Sumatra in habitats at forest borders, open areas, open canyons, river beds, villages, palm oil and rubber plantation borders and along roadsides.

Specimens examined: Anai waterfall, Padang Panjang city, West Sumatra province, 11 Jan 1981, *M. Hotta* 26153 (4 sheets) (BO); Batu Gajabuih, Desa Ulu Gadut, about 15 km east Padang city, West Sumatra province, 11 Feb 1981, *M. Hotta* and *H. Okada* 60 (2 sheets) (BO); Desa Ulu Gadut, about 10 km east Padang City, West Sumatra province, 24 Jan 1981, *M. Hotta* 26128 (1 sheet) (BO); Eastern foot of Mount Sago, south of Payakumbuh regency, West Sumatra province, 28 Aug 1983, *M. Hotta* 26599 (1 sheet) (BO); Mount Talang, Desa Laras Talang, West Sumatra province, 1919, *Bunne Meyer* 5559 (1 sheet) (BO); Batu Gajabuih, Desa Ulu Gadut, about 15 km east Padang city, West Sumatra province, 6 Jan 1981, *M. Hotta* 25853 (2 sheets) (BO); Desa Lawe Mangkudu, Ketambe district, Southeast Aceh regency, Nanggroe Aceh Darussalam province, 26 Feb 2013, *Y.S. Poerba* and *K.U. Nugraheni* PN 17 (1 sheet) (BO); Desa Sukamaju, Pesisir Barat district, Pesisir Barat regency, Lampung province, 20 May 2017, *F. Ahmad* and *Y.S. Poerba* AP-SSL 17 (3 sheets) (BO).

***Musa acuminata* Colla subsp. *sumatrana* (Becc.) Ahmad, Volkaert, Sulist. & Poerba, **com. nov.** ≡ *Musa sumatrana* Becc., André, Ill. Hort. 27: 37.1880. ≡ *Musa acuminata* Colla var. *sumatrana* (Becc.) Nasution, Mem. Tokyo Univ. Agric. 32: 80. (991.—TYPE: Indonesia, Sumatra, Padang, Aug 1878, *O. Beccari* 489 (lectotypes: K000292216 [web!], K000292218 [web!])).**

Diagnostis: Male bud like a top or lanceolate; bract shape small or medium to large shoulder, apex slightly pointed or intermediate, moderately waxy, convolute, old bract overlap at apex of bud, external bract purple, internal bract red with yellow pigmentation, apex slightly pointed or intermediate, not tinted with yellow colour, lifting one or two or more bracts at a time, revolute before falling; Initial flowers bisexual with pollen producing anthers; Seeds angular, flattened with rough edges (Online Resource 13).

Description: Plants are suckering freely. The mature pseudostems are 3–6 m tall, 15 up to 30 cm in diameter at the base, covered with dead remains of older sheaths. Sheaths are purplish or blackish, underlying colour cream with pink-purple pigmentation, very little to moderately waxy and the sap is milky. Petioles are 40–50 cm long, and 2–3 cm in diameter. The petiole canal is straight with erect margins or margins curved inward to margins overlapping, and the petiole margin colour is purple to blue. The laminae are lanceolate, 150–250 × 40–60 cm, with a slightly asymmetric to symmetric base, both sides rounded or one side rounded and one side pointed, apex truncate, abaxial medium green with waxy, adaxial green not waxy. The colour of the abaxial midrib is yellow green with pink pigmentation, and the colour of the adaxial midrib is light green. Inflorescences hang vertically, and the bunches are mostly very compact. The peduncle is 30–40 cm long, 6–8 (–10) cm in diameter, and the colour is green to dark green, pubescent. The rachis is mostly curved, and initial flowers are hermaphrodite, followed by male flowers. The inflorescence bud is like a top or lanceolate, 4–5 × 12–15 cm. The bract shape is small or medium to large shoulder, apex slightly pointed or intermediate, moderate waxy, 6–7 × 10–11 cm, convolute. The outer bract overlaps at apex of bud, the external bract colour is dark purple, internal red with yellow pigmentation, with a slightly pointed or intermediate apex, not tinted with a yellow colour, lifting one or more bracts at a time, revolute before falling. The ovary is cream coloured and ovules are in four rows per locule. The male flower has a whitish compound tepals, with three prominently thickened light yellow keels. The free tepals are translucent white, ovate, apex acute with a developed sign. The fruit bunch has 20–25 hands, with fruits in two rows, 20–25 fruits in the third hand. Individual fruits are straight, up to 15 cm long (including pedicel), 1–1.5 cm in diameter and the apex is bottle necked. The pedicel is ca. 1 cm long, straight or straight in the distal part. Immature fruit peels are green, not waxy, and the fruit pulp is white, containing up to 200 seeds per fruit, which are angular flat with rough edge, 7–8.3 mm × 5–6.5 mm in area, and 2–3 mm depth.

Distribution: This subspecies was found in Aceh, North Sumatra, West Sumatra, Bengkulu, and Lampung provinces at elevations of 750–1500 m asl at forest borders, open areas or at some shaded river beds, villages, and along roadsides.

Specimens examined: Desa Sungai Kering, Kerinci regency, Jambi province, 3 Mar 1954, A.H.G. Alston 14 120 (2 sheets) (BO); Desa Lae Pondom, Sumbul Pegagan district, Dairi regency, North Sumatra province, 28 Feb 2013, Y.S. Poerba and K.U. Nugraheni PN 44c (6 sheets) (BO); Desa Tiayang Layar, Pancur Batu district, Deli Serdang regency, North Sumatra, 1 May 2013, Y.S. Poerba and K.U. Nugraheni PN 73 (3 sheets) (BO); Desa Ketangkuhen, Sibolangit district,

Deli Serdang regency, North Sumatra province, 1 May 2013, Y.S. Poerba and K.U. Nugraheni PN 67 (3 sheets) (BO), PN 68 (4 sheets) (BO); Desa Sikeben, Sibolangit district, Deli Serdang regency, North Sumatra province, 1 May 2013, Y.S. Poerba and K.U. Nugraheni PN 65 (4 sheets) (BO).

Musa acuminata* Colla var. *longepetiolata Nasution, Mem. Tokyo Univ. Agric. 32: 63, Fig. 24. 1991.—TYPE: Indonesia, Palembang, Musi Rawas, Jun 1982, Rusdy 1630 (holotype: BO [n.v.]).

Diagnosis: In general, the habitus, bunch, male bud, and flower were similar to subsp. *malaccensis*, but the petiole length is at least half of its lamina (Online Resource 14).

Description: Plants are suckering freely. The mature pseudostems are 2–3 m tall, 15 up to 25 cm in diameter at the base, covered with dead remains of older sheaths. Sheaths are light green with reddish, brown or purplish or blackish blotches, underlying colour cream, very little to moderately waxy and the sap is milky or watery. The petioles are 60–100 cm long, the petiole canal is straight with erect margins, and the petiole margin colour is green, pink-purple to red or purple to blue. The laminae are lanceolate, 120–250 × 35–50 cm, with a base asymmetric or symmetric, one side pointed or both sides pointed, apex truncate. The colour of the abaxial is light green with waxy, and the colour of adaxial is green. The colour of abaxial midrib is mostly green, and the colour of adaxial midrib is light green. Inflorescences are horizontal, and the bunches are mostly moderately compact. The peduncle is 20–60 cm long, 4–6 cm in diameter, and the colour is green to dark green with brown blotches, pubescent. The rachis is bare or neutral flowers, and mostly falling vertically or horizontal. The male bud is like a top, 5–9 × 10–15 cm. The bract shape is medium to large shoulder, with slightly pointed or pointed apex, slightly waxy, 5–9 × 10–18 cm, and the old bract is overlap at apex of bud or young bract slightly overlap, the external bract is red to red purple, the internal bract is whitish to red, the apex is acute, which not tinted with yellow colour or tinted with yellow, lifting one or two or more bracts at a time, revolute before falling, bract mostly very little or few wax. The basal flowers are carpellate, which prominently ridged, and yellowish green. The ovules are in four rows per locule. The stigma is flat, 6 × 6 mm, and greyish. The styles are 3 cm long, and creamy. The compound tepals of male flowers are white or cream, with two prominently thickened light yellow keels, and the apex is deep yellow and curved outward. The free tepals are translucent white ovate, and the apex is acute, with little or no visible sign a developed sign. The anthers are in the same level of the lobes on the compound tepal. The male flower has five creamy filaments of staminode, and the anther lobes is white, cream, rusty brown or pinkish. The colour of ovaries is creamy, without additional

pigmentation, and the stigma is creamy. The fruit bunches are horizontal, compact, and has 5–9 hands. The fruits are in two rows. The individual fruits is straight or straight in the distal part, up to 10 cm long (including pedicel), 1–1.5 cm in diameter, prominently ridged, and the apex is obtuse. The pedicel is up to 1 cm long, straight. The immature fruit peel is green, not waxy, and the fruit pulp is white containing seeds up to 150 seeds per fruit, which are angular, wrinkled, 3–6 mm long, 3–5 mm width, 3.5 mm depth.

Distribution: This subspecies was only collected in Bengkulu and South Sumatra in habitats at forest borders, open areas, villages, and along roadsides.

Mount Mugajah population

Diagnosis: The distinctive characteristics of this population are the rounded male inflorescence bud with an imbricated tip and one to three bracts lifting, not revolute before falling; The pedicel of the fruit is rather long (Online Resource 15).

Description: Plants suckering freely. The mature pseudostems are 2.5–3 m tall, 15 up to 25 cm in diameter at the base, covered with dead remains of older sheaths. Sheaths are black-purple, underlying colour cream, very little to moderately waxy and the sap is milky. Petioles are 40–50 cm long, and 2–3 cm in diameter. The petiole canal is erect margins, and the petiole margin colour is purple to blue. The laminae are lanceolate, 150–200 × 35–60 cm, with a symmetric base, both sides rounded, apex truncate, abaxial light green with waxy, adaxial green not waxy, colour of abaxial midrib pink purple, colour of adaxial midrib pink purple. Inflorescences at an angle, and the bunched are lack. The peduncle is 30–40 cm long, 3–5 cm in diameter, and the colour is green to dark green or with brown blotches, pubescent. The rachis is bare. The male bud is rounded, 7–8 × 10–13 cm. The bract shape is small shoulder, apex obtuse, waxy, 7–8 × 10–13 cm, imbricate. The young bract is slightly overlap, external bract purple-brown, internal bract red with a yellow pigmentation, not tinted with a yellow colour, lifting one or two or more bracts at a time, not revolute before falling, and with some moderate wax. The ovary is cream coloured. The ovules are in four rows per locule. The compound tepals of male flowers are whitish, with 3 prominently thickened light yellow keels. The free tepals are translucent white, ovate, apex acute with a developed sign. The male flower has five whitish filaments or staminodes. The fruit bunch has 8–9 hands, with fruits in 2 rows, 20–25 fruits in the third hand. Individual fruits are straight, up to 15 cm long (including pedicel), 1–1.5 cm in diameter, apex is blunt-tipped; pedicel is ca. 2 cm long, straight or straight in the distal part. Immature fruit peels are green, not waxy, fruit pulp white, containing up to 200 seeds

per fruit, which are flat, smooth, 6.4–7.2 mm long, 4–5.6 mm wide, 2.9–3.6 mm depth.

Distribution: We found this variety at only three locations a few kilometres apart at high elevation (1500 m) in central Aceh growing at forest borders and open areas along the roadside.

Specimen examined: Pantau Cuaca district, Gayo Lues regency, Nanggroe Aceh Darussalam province, 22 Mar 2017, F. Ahmad, Y.S. Poerba and H.A. Volkaert ANS-17 (4 sheets) (BO).

Information on Electronic Supplementary Material

Online Resource 1. Intraspecific classification of *Musa acuminata* over time.

Online Resource 2. Banana accessions of the Research Center for Genetic Engineering, Research Organization of Life and Environment, National Research and Innovation Agency (BRIN) collection that were included for genetic diversity in this study.

Online Resource 3. Accessions used to determine and visualise the clustering based on the multivariate morphological characters, a principal component analysis (PCA)

Online Resource 4. Genotypes of accessions with singleton haplotypes joined with their closest non-singleton haplotype relative.

Online Resource 5. Observation points of wild *Musa acuminata* from Sumatra exploration with location data.

Online Resource 6. Accession (SSB-33) with intermediate characteristics of the fruit between *M. acuminata* subsp. *sumatrana* and *M. acuminata* susp. *halabanensis* indicated by a horizontal bunch such as *M. acuminata* susp. *halabanensis*, but with typical long fruit such as in *M. acuminata* subsp. *sumatrana* fruits and containing angular seeds, rather like subsp. *malaccensis*.

Online Resource 7. Undeveloped fruits of *Musa acuminata* subsp. *halabanensis* with few seed or seedless.

Online Resource 8. Bunches of *Musa acuminata* subsp. *halabanensis* with cracked peel, similar to the phenomenon found in *M. schizocarpa*.

Online Resource 9. Characteristics of the male bud during its development.

Online Resource 10. A *Musa acuminata* subsp. *sumatrana* accession grown in the BRIN gardens at low elevation (160 m asl) in Cibirong, West Java which are morphologically very similar to *M. acuminata* subsp. *malaccensis* or *M. acuminata* var. *zebrina*.

Online Resource 11. Photographs of wild *Musa acuminata* subsp. *malaccensis* in Sumatra.

Online Resource 12. Photographs of wild *Musa acuminata* subsp. *halabanensis* in Sumatra.

Online Resource 13. Photographs of wild *Musa acuminata* subsp. *sumatrana* in Sumatra.

Online Resource 14. Photographs of wild *Musa acuminata* var. *longepetiolata* in Sumatra.

Online Resource 15. Photographs of wild *Musa acuminata* 'Mount Mugajah population' in Sumatra.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00606-025-01958-x>.

Acknowledgements We thank the Head of Research Center for Biology for providing the exploration permit. We gratefully acknowledge the Center for Agricultural Biotechnology at Kasetsart University Kamphaengsaen Campus for providing laboratory space and computer facilities. Our thankfulness goes to the local governments of the provinces of Nangroe Aceh Darussalam, North Sumatra, West Sumatra, Riau, Jambi, Bengkulu, South Sumatra and Lampung for facilitating the exploration. We thank Bayu Arief Pratama for discussing the spatial multivariate analysis, Dian Mulyana for field activities and Herlina for technical assistance during the course of this project. Banana research at Wageningen University and Research was supported by the Dutch Dioraphte Foundation Endowed Chair in Tropical Phytopathology of GHJK at the WUR-Laboratory of Phytopathology.

Author contributions HAV: Conceptualization, Supervision, Methodology, Software, Formal analysis, Formal analysis Writing—Original Draft. FA: Methodology, Software, Investigation, Resources, Formal analysis, Visualization, Writing—Original Draft. GHJK: Conceptualization, Supervision, Writing- Reviewing and Editing, Project administration. HJ: Conceptualization, Supervision, Writing- Reviewing and Editing. YSP: Methodology, Resources, Investigation. LDS: Investigation. SS: Investigation.

Funding This research partially funded by the Royal Netherlands Academy of Arts and Sciences (KNAW) Indonesia-Netherlands Collaborative Program (SPIN) and the Dutch Dioraphte Foundation.

Data availability No datasets were generated or analysed during the current study.

Declaration

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmad F, Bourke P, Schouten H, Volckaert HA, Kema GHJ (2020) Cytogenetics of structural rearrangements in *Musa* hybrids and cultivars. In: Kema GHJ, Drenth A (eds) Achieving sustainable cultivation of bananas. Vol. 2. Germplasm and genetic improvement. Burleigh Dodds Science Publishing, Cambridge, pp 1–28. <https://doi.org/10.19103/as.2019.0054.13>
- Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Model* 153:51–68
- Carreel F, Gonzales de Leon D, Lagoda P, Lanaud C, Jenny C, Horry J-P, Tezenas du Montcel H (2002) Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. *Genome* 45:679–692. <https://doi.org/10.1139/g02-033>
- Chung MY, Son S, Herrando-moraira S, Tang CQ, Maki M, Kim Y, López-pujol J, Hamrick JL, Chung MG (2020) Incorporating differences between genetic diversity of trees and herbaceous plants in conservation strategies. *Conserv Biol* 34:1142–1151. <https://doi.org/10.1111/cobi.13467>
- Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molec Ecol* 9:1657–1659. <https://doi.org/10.1046/j.1365-294x.2000.01020.x>
- Creste S, Tulmann Neto A, Vencovsky R, de Oliveira SA, Figueira A (2004) Genetic diversity of *Musa* diploid and triploid accessions from the Brazilian banana breeding program estimated by microsatellite markers. *Genet Resources Crop Evol* 51:723–733
- Crichton RR, Vezina A, van den Bergh I (2016) An online checklist of banana cultivars. *Acta Hort* 2:13–18
- Cronn R, Cedroni M, Haselkorn T, Grover C, Wendell JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor Appl Genet* 104:482–489
- Drenth A, Kema GHJ (2021) The vulnerability of bananas to globally emerging disease threats. *Phytopathology* 111:2146–2161. <https://doi.org/10.1094/PHYTO-07-20-0311-RVW>
- Dmitriev DA, Rakitov RA (2008) Decoding of superimposed traces produced by directsequencing of heterozygous indels. *PLoS Comput Biol* 4:e1000113. <https://doi.org/10.1371/journal.pcbi.1000113>
- Dodds KS (1943) Genetical and cytological studies of *Musa*. V. Certain edible diploids. *Genetics* 45:113–139
- Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resources* 4:359–361
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molec Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294x.2005.02553.x>
- FAO (2020) Banana market review 2019. FAO, Rome
- FAO (2023) Banana facts and figures. FAO, United Nations. Available at: <https://www.fao.org/economic/est-commodities/oilcrops/bananas/bananafacts/en/>. Accessed 29 Dec 2023
- Häkkinen M (2013) Reappraisal of sectional taxonomy in *Musa* (Musaceae). *Taxon* 62:809–813. <https://doi.org/10.12705/624.3>
- Häkkinen M, Väre H (2008) Typification and check-list of *Musa* L. names (Musaceae) with nomenclatural notes. *Adansonia* 30:63–112
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* 59:1633–1638
- Hippolyte I, Jenny C, Gardes L, Bakry F, Rivallan R, Pomies V, Cubry P, Tomekpe K, Risterucci AM, Roux N, Rouard M, Arnaud E, Kolesnikova-Allen M, Perrier X (2012) Foundation characteristics of edible *Musa* triploids revealed from allelic distribution of SSR markers. *Ann Bot (Oxford)* 109:937–951. <https://doi.org/10.1093/aob/mcs010>
- Hotta M (1989) Identification list of *Ensete* and *Musa* (Musaceae) in SE Asia and West Malesia. South Pacific Ocean Research Report, Nankai Lab Occasional papers, Kagoshima
- Husson L, Boucher F, Sar A-C, Sepulchre P, Cahyarini S (2020) Evidence of Sundaland's subsidence requires revisiting its biogeography. *J Biogeogr* 47:843–853. <https://doi.org/10.1111/jbi.13762>
- IPGRI (1996) Descriptors for banana (*Musa* spp.). IPGRI, Rome
- Itino T, Kato M, Hotta M (1991) Pollination ecology of the two wild bananas, *Musa acuminata* subsp. *halabanensis* and *M. salaccensis*: chiropterophily and ornithophily. *Biotropica* 23:151–158

- Kress JW, Specht CD (2006) The evolutionary and biogeographic origin and diversification of the tropical monocot order Zingiberales. *Aliso* 22:49
- Jost L (2008) GST and its relatives do not measure differentiation. *Molec Ecol* 17:4015–4026
- Li YL, Liu JX (2018) StructureSelector: a web based software to select and visualize the optimal number of clusters using multiple methods. *Molec Ecol Resources* 18:176–177. <https://doi.org/10.1111/1755-0998.12719>
- Li LF, Wang HY, Zhang C, Wang XF, Shi FX, Chen WN, Ge XJ (2013) Origins and domestication of cultivated banana inferred from chloroplast and nuclear genes. *PLoS ONE* 8:e80502. <https://doi.org/10.1371/journal.pone.0080502>
- Librado L, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1450–1452
- Louwaars NP (2018) Plant breeding and diversity: a troubled relationship? *Euphytica* 214:1–9. <https://doi.org/10.1007/s10681-018-2192-5>
- Martin G, Cardi C, Sarah G, Ricci S, Jenny C, Fondi E, Perrier X, Glaszmann JC, D'Hont A, Yahiaoui N (2020) Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *PL J* 102:1008–1025. <https://doi.org/10.1111/tj.14683>
- Meijer W (1961) Notes on wild species of *Musa* from Sumatra. *Acta Bot Neerl* 10:248–256
- Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucl Acids Res* 43:W566–W570. <https://doi.org/10.1093/nar/gkv468>
- Meyerhans A, Vartanian JP, Wain-Hobson S (1990) DNA recombination during PCR. *Nucl Acids Res* 18:1687–1691
- Nasution RE (1991) A taxonomy study of the species *Musa acuminata* Colla with its intraspecific taxa in Indonesia. *Mem Tokyo Univ Agric* 32:1–122
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321–3323. <https://doi.org/10.1073/pnas.70.12.3321>
- Oksanen AJ, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, Hara RBO, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2020) Package ‘vegan’: community ecology package version 2.5-7. Available at: <http://CRAN.Rproject.org/package=vegan>
- Ornduff R (1978) Reproductive characters and taxonomy. *Syst Bot* 3:420–427
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molec Ecol Notes* 6:288–295. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
- Peakall R, Smouse PE (2012) GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Perrier X, Bakry F, Carreel F, Jenny C, Horry JP, Lebot V, Hippolyte I (2009) Combining biological approaches to shed light on the evolution of edible bananas. *Ethnobot Res Appl* 7:199–216
- Ploetz RC (2005) Panama disease: an old nemesis rears its ugly head. Part 1. The beginnings of the banana export trades. *PL Health Prog* 6:1. <https://doi.org/10.1094/PHP-2005-1221-01-RV>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Puechmaille SJ (2016) The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molec Ecol Res* 16:606–627. <https://doi.org/10.1111/1755-0998.12512>
- R-Core-Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Raboin LM, Carreel F, Noyer JL, Baurens FC, Horry JP, Bakry F, Du Montcel HT, Ganry J, Lanaud C, Lagoda PJL (2005) Diploid ancestors of triploid export banana cultivars: molecular identification of 2n restitution gamete Donors and n gamete donors. *Molec Breed* 16:333–341. <https://doi.org/10.1007/s11032-005-2452-7>
- Rana TS, Narzary D, Mahar KS, Meena B (2014) Methods and approaches in plant molecular systematics. In: Rana T, Nair K, Upreti D (eds) *Plant taxonomy and biosystematics: classical and modern methods*. New India Publishing Agency, New Delhi, pp 383–416
- Retnoningsih A, Megia R, Hartana A (2010) Molecular verification and diversity analysis of Indonesian BB, AAB and ABB banana cultivars. *Tree Forest Sci Biotechnol* 4:69–76
- Ruas M, Guignon V, Sempere G, Sardos J, Hueber Y, Duvergey H, Andrieu A, Chase R, Jenny C, Hazekamp T, Irish B, Jelali K, Adeka J, Ayala-Silva T, Chao CP, Daniells J, Dowiya B, Effa Effa B, Gueco L, Herradura L, Ibobondji L, Kempenaers E, Kilangi J, Muhangi S, Ngo Xuan P, Paofa J, Pavis C, Thiemele D, Tossou C, Sandoval J, Sutanto A, Vangu Paka G, Yi G, van den Houwe I, Roux N, Rouard M (2017) MGIS: managing banana (*Musa* spp.) genetic resources information and high-throughput genotyping data. Database 2017:bax046. <https://doi.org/10.1093/database/bax046>
- Sardos J, Breton C, Perrier X, van den Houwe I, Carpentier S, Paofa J, Rouard M, Roux N (2022) Hybridization, missing wild ancestors and the domestication of cultivated diploid bananas. *Frontiers Pl Sci* 13:1–17. <https://doi.org/10.3389/fpls.2022.969220>
- Shepherd K (1999) Cytogenetics of the genus *Musa*. International Network for the Improvement of Banana and Plantain, Montpellier
- Smith M, Pillay M (2020) An overview of genetic improvement in bananas over the last century. In: Kema GHJ, Drenth A (eds) *Achieving Sustainable Cultivation of Bananas. Vol. 2. Germplasm and genetic improvement*. Burleigh Dodds Science Publishing, Cambridge, pp 1–28
- Simmonds NW (1956) Botanical results of the banana collecting expedition, 1954–5. *Kew Bull* 11:463–489
- Simmonds NW (1953) The development of the banana fruit. *J Exp Bot* 4:87–105. <https://doi.org/10.2307/23686166>
- Simmonds NW (1962) The evolution of the bananas. Longmans, London
- Soltis DE, Soltis PS (1998) Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis PS, Soltis DE, Doyle JJ (eds) *Molecular systematics of plants II: DNA sequencing*. Kluwer, Boston, pp 1–42. <https://doi.org/10.1007/978-1-4615-5419-6>
- Soltis ED, Soltis PS (2000) Contributions of plant molecular systematics to studies of molecular evolution. *PL Molec Biol* 42:45–75. <https://doi.org/10.1023/A:1006371803911>
- Stephens M, Nicholas JS, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Amer J Human Genet* 68:978–989
- Stover R, Simmonds N (1987) Bananas: tropical agriculture series, 3rd ed. Longman, New York. 468p
- Swangpol S, Volkaert H, Sotto RC, Seelanan T (2007) Utility of selected non-coding chloroplast DNA sequences for lineage assessment of *Musa* interspecific hybrids. *BMB Rep* 40:577–587. <https://doi.org/10.5483/BMBRep.2007.40.4.577>
- Syamkumar S, Lowrance B, Sasikumar B (2003) Isolation and amplification of DNA from rhizomes of turmeric and ginger. *PL Molec Biol Report* 21:171a–171e
- Valmayor RV (1998) Classification et caractérisation de *Musa exotica*, *M. alisanaya* et *M. acuminata* ssp. *errans*. *InfoMusa* 10:35–39
- Volkaert H (2011) Molecular analysis reveals multiple domestications of edible bananas. *Acta Hort* 897:143–152. <https://doi.org/10.17660/ActaHortic.2011.897.15>

- Voora V, Larrea C, Bermudez S (2020) Global market report: Banana. In: Baliño S (ed) Sustainable commodities marketplace series. International Institute for Sustainable Development, Manitoba
- Wattanakulpakin T, Iamtham S, Grubbs KC, Volkaert HA (2015) Population genetic analysis of *Xylia xylocarpa* (Fabaceae—Mimosoideae) in Thailand. *Tree Genet Genomes* 11:7. <https://doi.org/10.1007/s11295-014-0825-y>
- WFO (2024) World Flora Online Version (2024). Available at: <http://www.worldfloraonline.org>. Accessed 12 Feb 2024
- Wilson GB (1946a) Cytological studies in the Musae. I. Meiosis in some triploid clones. *Genetics* 31:241–258
- Wilson GB (1946b) Cytological studies in the Musae. II. Meiosis in some diploid clones. *Genetics* 31:475–482
- Wilson GB (1946c) Cytological studies in the Musae. III. Meiosis in some seedling clones. *Genetics* 31:483–493
- Winter D (2012) mmod: an R library for the calculation of population differentiation statistics. *Molec Ecol Res* 12:1158–1160. <https://doi.org/10.1111/j.1755-0998.2012.03174.x>

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