Received: 21 November 2023 | Accepted: 7 July 2024

DOI: 10.1002/tpg2.20497

The Plant Genome

ORIGINAL ARTICLE

Special Section: Modern Improvement of Tropical Crops

Genetic diversity and population structure in banana (*Musa* spp.) breeding germplasm

Violet Akech[1,2] | **Therése Bengtsson**[2] | **Rodomiro Ortiz**[2] | **Rony Swennen**[1,3] | **Brigitte Uwimana**[1] | **Claudia F. Ferreira**[4] | **Delphine Amah**[5] | **Edson P. Amorim**[4] | **Elizabeth Blisset**[6] | **Ines Van den Houwe**[7] | **Ivan K. Arinaitwe**[8] | **Liana Nice**[6] | **Priver Bwesigye**[8] | **Steve Tanksley**[6] | **Subbaraya Uma**[9] | **Backiyarani Suthanthiram**[9] | **Marimuthu S. Saraswathi**[9] | **Hassan Mduma**[10] | **Allan Brown**[10]

[1]International Institute of Tropical Agriculture, Uganda, Uganda

[2]Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden

[3]Department of Biosystems, KU Leuven, Leuven, Heverlee, Belgium

[4]Brazilian Agricultural Research Corporation (Embrapa), Brasília, Brasil

[5]International Institute of Tropical Agriculture, Ibadan, Nigeria

[6]Nature Source Improved Plants, Ithaca, New York, USA

[7]The Alliance of Bioversity and CIAT—Musa Germplasm Transit Centre, Heverlee, Belgium

[8]National Agricultural Research Organization, Kampala, Uganda

[9]National Research Centre for Banana (NRCB) of India, Tiruchirappalli, India

[10]International Institute of Tropical Agriculture, Arusha, Tanzania

Correspondence
Violet Akech, International Institute of Tropical Agriculture, Uganda.
Email: violet.akech@slu.se, V.Akech@cgiar.org

Assigned to Associate Editor Damaris Odeny.

Funding information
Bill and Melinda Gates Foundation, Grant/Award Number: INV-002979 (formally OPP1213871)

Abstract

Bananas (*Musa* spp.) are one of the most highly consumed fruits globally, grown in the tropical and sub-tropical regions. We evaluated 856 *Musa* accessions from the breeding programs of the International Institute of Tropical Agriculture of Nigeria, Tanzania, and Uganda; the National Agricultural Research Organization of Uganda; the Brazilian Agricultural Research Corporation (Embrapa); and the National Research Centre for Banana of India. Accessions from the in vitro gene bank at the International Transit Centre in Belgium were included to provide a baseline of available global diversity. A total of 16,903 informative single nucleotide polymorphism markers were used to estimate and characterize the genetic diversity and population structure and identify overlaps and unique material among the breeding programs. Analysis of molecular variance displayed low genetic variation among accessions and diploids and a higher variation among tetraploids ($p < 0.001$).

Abbreviations: AMOVA, analysis of molecular variance; DArTseq, diversity array technology sequencing; DAPC, discriminant analysis of principal component; EAHB, East African Highland Bananas; IITA, International Institute of Tropical Agriculture; ITC, International Transit Centre; MAF, minor allele frequency; NARO, National Agriculture Research Organization; NSIP, Nature Source Improved Plant; PC, principal component; PCoA, principal coordinate analysis; SNP, single nucleotide polymorphism.

Structure analysis revealed two major clusters corresponding to genomic composition. The results indicate that there is potential for the banana breeding programs to increase the diversity in their breeding materials and should exploit this potential for parental improvement and to enhance genetic gains in future breeding efforts.

Plain Language Summary

Banana is an important staple food and popular fruit, grown in the tropical and sub-tropical regions. Pests and pathogens reduce banana production, and the use of resistant cultivars is the ultimate solution. This study used DNA from 856 genotypes from banana breeding programs in Brazil, India, Nigeria, Tanzania, Uganda, and the in vitro genebank at the International Transit Centre in Belgium to determine the level of diversity available and identify unique materials with potential valuable genes. The results indicate low genetic diversity. There is a need to increase the diversity in parental materials to broaden their genetic base, and this study provides important clues on how to achieve this, such as the material from Brazil, with host plant resistance to pathogens and pests could benefit other programs. Banana breeders should exploit this potential to enhance genetic gains in future breeding efforts by increased cooperation and germplasm exchange among breeding programs.

1 | INTRODUCTION

With an estimated global production of 140 million metric tons, bananas and plantains (*Musa* spp.) are the world's leading fruit crop (Evans et al., 2020). The crop is non-seasonal and produces fruits throughout the year, contributing to food security as a major staple food and as a source of supplemental income for hundreds of millions of people in the global tropical and sub-tropical South (Lescot, 2020; Nayar, 2010). Currently, the largest producer of bananas is Asia (55.9%), followed by Africa (24.6%), America (16.1%), the Pacific islands (1.7%), Oceania (1.3%) and Europe (0.4%) (FAO, 2022).

Bananas are divided into edible cultivars and non-edible wild species. The wild species are diploids ($2n = 2x$) and include *Musa acuminata* Colla (AA) and *M. balbisiana* Colla (BB), the sources of the A and B genome, respectively. Both species belong to the section *Eumusa* of the genus *Musa*. The edible parthenocarpic cultivars have previously been reported to have evolved from these two wild species, and cross-compatibility between the two *Musa* species allowed natural hybridization leading to the development of *M. acuminata* × *M. balbisiana* hybrids (Deepthi, 2016; N. W. Simmonds & Shepherd, 1955). However, recent studies have reported hybridizations with other unknown ancestral contributors from uncharacterized gene pools (Jeensae et al., 2021; Martin et al., 2020; Sardos et al., 2022). The intra- and interspecific hybridization with reduced and unreduced gametes of the wild species resulted in cultivars with a mixture of genomic combinations and ploidy levels that include diploid ($2n = 2x$) AA and AB; triploid ($2n = 3x$) AAA, AAB, and ABB; and tetraploid ($2n = 4x$) AAAB, AABB, and ABBB (Manzo-Sánchez et al., 2015; Nayar, 2010).

Bananas were domesticated in Southeast Asia and the Pacific regions (Castillo & Fuller, 2012) before the crop was introduced to other regions of the world (N. Simmonds, 1962). Their cultivation in Africa started between 2000 and 6000 years ago (De Langhe et al., 1994; Perrier et al., 2011). Based on a broad view of archaeological, cultural, and linguistic evidence combined with genetic results, the Indian Ocean islands such as Madagascar, Zanzibar, Comoros, and Pemba were entryways of bananas into Africa from Southeast Asia by various waves of migration of the Austronesian populations (Blench, 2009; De Langhe et al., 2009; Perrier et al., 2019). Edible bananas may have entered the East African region (including Burundi, the Democratic Republic of Congo, Kenya, Rwanda, Uganda, and Tanzania) through multiple introductions between the first and 16th centuries A.D. (Karamura, 1998; Nayar, 2010). Large variability caused by somatic mutations or possibly epigenetics gave rise to a distinct group of about 70 cultivars endemic to the East African region referred to as the East African Highland Bananas (EAHB). These include both the triploid East African highland cooking and beer banana (*Musa*, AAA), thereby making East Africa a secondary center of banana diversity (De Langhe et al., 2009).

Using botanical and linguistic evidence, it is hypothesized that triploid plantains (AAB) may have reached Africa more than 3000 years ago (De Langhe et al., 1994) from the

Philippines and the Eastern contact areas between the Philippines and New Guinea (Lejju et al., 2006). However, there are suggestions that they may have been brought to Africa from India or Sri Lanka by the Bornean people migrating along the South Arabia coastal regions (Fuller & Madella, 2009; Lejju et al., 2006). A large number of the African plantain cultivars that have never been recorded elsewhere suggest that they underwent a sustained diversification in Africa over a long time (Blench, 2009; De Langhe, 2007), majorly due to natural mutations. These mutations gave rise to secondary plantain cultivars that have been maintained in regions of intense cultivation by vegetative propagation (De Langhe, 1964; De Langhe & de Maret, 2004). Consequently, like East Africa for EAHB, West and Central Africa became a center of secondary diversity for the plantain (Swennen, 1990). In Africa, East African cooking bananas (which include the "matooke"—AAA and "mchare"—AA), the brewing types ("mbidde"—AAA), and plantains (AAB genome) make up ~64% of all bananas grown on the continent (Lescot, 2020). In 2018, an estimated 12.4 million metric tons of plantains (93% from West and Central Africa) and 9.8 million metric tons of highland, beer, and other cooking bananas (88% from East Africa) were produced in Africa, where 3.2 million farming households depend on plantains and 2.5 million depend on highland bananas, ABB (which includes Sukali Ndizi for dessert; Kayinja and Kisubi for beer making), and other banana cultivars (Lescot, 2020).

Banana crossbreeding is impeded by the polyploid nature of the crop which is characterized by near sterility, poor seed set, and low germination (Batte et al., 2019). Edible cultivated bananas such as the EAHB and plantains have a low genetic variation (Kitavi et al., 2016; Nyine et al., 2017; Ortiz, 1997). Despite these challenges, a handful of banana breeding programs are developing high-yielding and resistant banana cultivars. Among these breeding programs, the International Institute of Tropical Agriculture (IITA) leads in Africa where they breed for cooking bananas, namely Mchare in Tanzania; Matooke in Uganda together with the Ugandan National Agriculture Research Organization (NARO), representing the Great Lakes Region of East Africa and plantain in Nigeria representing plantain for the lowlands of the degraded forest of West Africa. The Brazilian Agricultural Research Corporation (Embrapa) representative of Latin America breeds Prata (Pome)/Silk/Plantain, and the National Research Centre for Banana (NRCB) of India representative of Asia breeds Silk/Pome/Dessert/Plantain/ABB cooking banana. All these programs focus their breeding objectives on developing hybrids with consumer-preferred qualities in terms of color, texture, and taste, with varying threshold values depending on the target product (Amorim, dos Santos-Serejo, et al., 2011; Silva et al., 2001; Madalla et al., 2023; Marimo et al., 2019; Nowakunda et al., 2023; Sathiamoorthy et al., 2001), short stature, reduced crop cycle, and good agronomic attributes that lead to high yield and resistance to a complex of diseases and pests that affect bananas and plantains (Amorim, Amorim, et al., 2011; Brown et al., 2017; Ortiz et al., 1995; Swennen & Vuylsteke, 1993; Vuylsteke et al., 2010).

Core Ideas

- There is low genetic diversity among Musa breeding materials.
- The lowest genetic distance was between the breeding materials from Tanzania for Mchare and Brazil for Prata.
- There is a need for broadening genetic diversity for parental improvement especially for the diploids.
- There is a possibility of using the selected DArTseq-SNP markers for further genomic research in Musa.

The crossbreeding process, however, depends mostly on existing genetic diversity within *Musa* breeding populations, cultivars, and gene banks, which largely determine the potential of plant improvement that can be expected (Brown et al., 2017). It is, therefore, important to assess the genetic diversity of existing germplasm and breeding populations to guide informed crosses or breeding schemes for the development of new and improved cultivars and avoid inbreeding and narrowing the genetic base in advanced generations (Yao et al., 2008). While many studies such as Christelová et al. (2017) and Bawin et al. (2019) have conducted studies of materials available in collections, this is the first study that evaluates accessions used by several breeding programs.

In the present study, we determined the genetic diversity and population structure of *Musa* germplasm available in the breeding programs representative of the four largest consumers of bananas worldwide; IITA in Nigeria for plantain, Tanzania for Mchare, IITA—NARO breeding programs for Matooke in Uganda, Embrapa in Brazil, and NRCB in India. Additionally, accessions from the in vitro banana gene bank at the International Transit Centre (ITC) of Alliance of Bioversity International and CIAT in Belgium were included to identify potential diversity available that can be used to broaden the narrow genetic base for banana and plantain breeding.

This study aimed to (1) determine whether DArTseq-single nucleotide polymorphism (SNP) markers can be used to study diversity and population genetics of *Musa*, (2) understand the level of genetic diversity available and the pattern of population structure among the *Musa* breeding materials from five breeding programs, and (3) identify compatible new sources of germplasm harboring valuable variation for improving resistance to biotic stresses.

TABLE 1 Plant materials used in the study.

Source	Breeding materials	Abbreviation	Ploidy			Total
			2x	3x	4x	
IITA—Nigeria	Plantain	Plantain	12	5	18	35
IITA—Tanzania	Mchare	Mchare	157	49	1	207
IITA and NARO—Uganda	Matooke	Matooke	58	12	44	114
Embrapa—Brazil	Prata (Pome)/Silk/Plantain	Prata	70	31	41	142
NRCB—India	Silk, Pome, Dessert, Plantain, cooking	SwS (sweet and starchy)	33	112	5	150
ITC—Belgium	Conservation by ITC	ITC	199	6	3	208
Total			529	215	112	856

Abbreviations: IITA, International Institute of Tropical Agriculture; ITC, International Transit Centre; NARO, National Agriculture Research Organization; NRCB, National Research Centre for Banana.

2 | MATERIALS AND METHODS

2.1 | Plant material and sampling

The accessions were provided by five breeding programs: IITA in Nigeria (4%) representing the triploid Plantain breeding program; Tanzania (24%) representing the Mchare and diploid breeding program; Uganda (13%) (including those collected from NARO) representing the triploid Matooke breeding program (henceforth referred to as "Plantain," "Mchare," and "Matooke" breeding material, respectively); Embrapa (17%) representing the triploid Prata (Pome)/Silk and plantain breeding program; and NRCB (18%) representing the Silk/Pome, Dessert, plantain, and ABB cooking breeding program (henceforth referred to as Prata and SwS breeding material, respectively). Additionally, breeding material from accessions in the ITC in Belgium (24%; henceforth referred to as ITC; Table 1) were included. The accessions from the breeding programs were grown in situ in different field trials and germplasm collections, while the accessions from ITC were held ex-situ as in vitro tissue culture plants before sampling.

Samples sent from the breeding programs were collected from the cigar (youngest and emerging) leaf samples, and 200 mg of each leaf sample was cut into small pieces of 9 cm^2, without the midrib, and gently placed in a 1.2-mL round-bottomed tube (containing one 4-mm stainless steel ball) in a 96-well tube rack. The samples were dried using silica gel pads (Dry & Dry brand from L2K Commerce) at 4˚C for 14 days, and dryness was confirmed with a humidity indicator, with an allowed maximum humidity of 10%. For samples sent from ITC, ~3 g of freeze-dried leaf tissues taken from the in vitro plants were pre-packed in air-tight sealed aluminum foil bags. The dried samples were consolidated, and then pulverized with the GenoGrinder (Spex Sample prep MiniG 1600 machine) run twice at 1400 rpm for 1 min at the Nature Source Improved Plants facilities (NSIP) in Ithaca, USA, and then shipped to DArT, Australia, for DNA extraction and genotyping.

Initially, 913 individual samples were sent to DArT for genotyping. For this study, 57 accessions from sections other than *Eumusa* were removed, and the remaining 856 *Musa* accessions (168 wild germplasm, 482 cultivars, 119 breeding clones, 58 hybrids, 14 chromosome-doubled, and 1 bluggoe) used as breeding materials, comprising different ploidy levels with diverse genomic constitutions, were used (Table 1). The number of genotypes (population size) is unevenly distributed among the compared groups in this study. The samples were allocated a seven-digit number as their sample ID (Table S1, Column A). A sample id as "2522162$2521592" was created by merging two or more individual samples as technical replicates confirmed to be of the same genotype after genotyping (hence combining sample 2522162 and sample 2521592 as sample 2522162$2521592).

Passport data and classification of the banana accessions as members of subgroups (such as "Pome," "Silk," "Cavendish," "Mchare," "Matooke," and "Plantain") were provided by the breeders and ITC curator, while other information was obtained from the *Musa* Germplasm Information System (MGIS; http://www.crop-diversity.org/banana/) (Ruas et al., 2017) and is presented in Table S1.

2.2 | Genotyping by DArTseq technology

DNA was extracted using the Diversity Array Technology plant DNA extraction protocol (DArT 2000). Genotyping was performed using the DArTSeq genotyping platform using the option of 1.2 million reads. Libraries were constructed following the DArTSeq complexity reduction method (Kilian et al., 2016), and genomic DNA was digested using a combination of *Pst*I and *Mse*I enzymes. Next-generation sequencing was carried out using HiSeq2000 (Illumina, USA). The sequences containing the SNP markers were aligned by NSIP to version

4 of the "DH Pahang" banana reference genome (Belser et al., 2021).

All accessions were called as diploids during genotype calling to reduce complexity in data analysis and use tools developed for diploids, as done in other studies (Christelová et al., 2017; Osterman et al., 2021).

The final calls were coded as "0/0" for homozygous reference allele, "1/1" for homozygous alternate allele, and "0/1" for heterozygotes for the downstream analysis for this study.

Before filtering, technical replicates, duplicates, and accessions (highlighted in italic in Table S1) from sections other than *Eumusa* were removed. The data were further filtered using the "dartR" package in R (Gruber et al., 2018) by removing loci with a call rate below 95%, individuals with a call rate below 80%, loci with a minor allele frequency (MAF) below 0.01, and monomorphic markers. Finally, loci with more than 25% heterozygosity were removed, resulting in a final dataset that was used for further analysis. The generated dataset was further divided into different subsets based on the breeding program and ploidy, which were filtered using the same thresholds as used for the full dataset (Table 2).

2.3 | Analysis of genetic diversity

Analysis of molecular variance (AMOVA) was conducted using the "poppr" R package (Kamvar et al., 2014) to estimate the genetic differentiation within and among accessions as well as to assess the population differentiation among the genetic groups using the complete panel and the different (sub)sets as per the abovementioned sources of the material (Table 1). To test for the significance of the AMOVA results, the "randtest" function from the R package "ade4" (Dray & Dufour, 2007) was used with 999 permutations. The pairwise population differentiation (F_{ST}) to determine the between-group differentiation was computed using the "dartR" package in R (Gruber et al., 2018).

2.4 | Analysis of population structure

The genetic structure of the *Musa* genetic resources from the five breeding programs (641 retained after filtering) was analyzed using principal coordinate analysis (PCoA) and discriminant analysis of principal components (DAPC). All analyses were performed for the subsets diploids and tetraploids using the retained SNP for each subset as shown in Table 2.

The PCoA was carried out using the "dartR" package in R (Mijangos et al., 2022) that acts as a wrapper for the "glPca" function in "adegenet" (Jombart & Ahmed, 2011). The "ggplot2" package (Wickham, 2009) was used to visualize the pattern of variation in two-dimensional plots.

TABLE 2 Number of single nucleotide polymorphism (SNP) markers and accessions before and after each filtering step for the entire dataset and the three subsets.

	Filtering individuals			Filtering the markers					
Subsets	Initial count	% of individual lost at 80% call rate	Final number of individuals	Initial count (loci)	% of loci lost at call rate >95%	% of monomorphic loci lost	% of loci lost at MAF > 1%	% of loci lost at heterozygous call ≤25%	Final number of markers retained
Entire dataset	856	1.3	845	44,902	27.4	5.6	26.3	25.5	16,903
Breeding programs[a]	648	1.2	641	44,902	29	15.9	22.7	39.6	12,455
Breeding programs; diploids	330	0.9	327	44,902	23.3	20.3	14.5	16.1	19,678
Breeding programs; tetraploids	109	0.9	108	44,902	46	30.5	17.4	60	6,143

Abbreviation: MAF, minor allele frequency.

[a]Excluding materials from ITC—Belgium.

Population structure was assessed to determine genetic subpopulation using the Bayesian model-based clustering implemented in the DAPC method in R using the "adegenet" package (Jombart & Ahmed, 2011; Jombart et al., 2010). The optimal number of clusters in the DAPC analysis was inferred using "Silhouette" clustering method of K-means analysis by varying the possible number of clusters from 1–20, and the function "fviz_nbclust function" from the "Factoextra" R package (Kassambara & Mundt, 2020) was used to confirm the number of clusters. In addition, a bar plot of eigenvalues for the discriminant analysis was used to select discriminant functions to be retained. The number of retained principal components (PC) for DAPC analyses was calculated using a cross-validation method implemented in the "xvalDapc" function from the "adegenet" R package. DAPC scatter plots were developed on the clusters identified through K-means using 100 retained principal components

3 | RESULTS

3.1 | SNP polymorphism and diversity

The alignment to version 4 of double haploid Pahang reference genome resulted in 44,902 unfiltered SNP markers. After filtering, 37.6% of the markers (16,903) and 845 individuals were retained (Table 2).

Out of these retained markers, 34.8% were highly polymorphic with a polymorphic information content (PIC) ≥ 0.25 (Figure S1A). Observed heterozygosity (Ho) ranged from 0 to 0.31 with an average value of 0.11 (Figure S1C). The markers were well distributed across the 11 banana chromosomes, with chromosome 4 having the highest number of markers (12.3%) and chromosome 2 having the lowest number of markers (6.6%; Figure S1D).

3.2 | Population genetic differentiation and molecular variance analysis

The genetic divergence between the populations from the five breeding programs (641 retained after filtering) was determined by calculating the pairwise genetic differentiation (F_{ST}; Table 3).

The F_{ST} among breeding programs varied from 0.007 to 0.177 (Table 3). The highest pairwise genetic differentiation was found between the breeding material of SwS and Matooke ($F_{ST} = 0.177$), while the lowest was observed between Prata and Mchare ($F_{ST} = 0.007$). There were zero F_{ST} values between the diploids from Plantain and Prata populations and F_{ST} value of 0.011 between the diploids from Mchare and Prata programs. For tetraploids, the highest was between SwS and Matooke breeding materials ($F_{ST} = 0.436$), and the lowest was between Matooke and Plantain breeding materials ($F_{ST} = 0.118$; Table S2).

TABLE 3 Pairwise genetic differentiation (F_{ST}) between breeding programs.

	Prata	SwS	Plantain	Mchare
SwS	0.044			
Plantain	0.075	0.131		
Mchare	0.007	0.054	0.076	
Matooke	0.098	0.177	0.079	0.093

The AMOVA revealed a low but significant ($p < 0.001$) difference among breeding programs and among the diploid germplasm of the five breeding programs, which accounted for 14% and 16% of the total variation, respectively (Table 4). Significant ($p < 0.001$) and high differentiation was obtained among groups of tetraploid accessions from the five breeding programs (accounting for 31% of the total variation). The global genetic difference among groups was measured by pairwise F_{ST} and within individuals (F_{IS}) as given in Table 4.

3.3 | Population structure

3.3.1 | Principal coordinate analysis

Nei's unbiased genetic distance was used to calculate the two-dimensional PCoA among all genotypes. The PCoA based on breeding programs explained 41.2% and 10.4% of the total variation in the first and second principal coordinates, respectively (Figure 1). Five clusters were observed with considerable admixtures among the genotypes, with one cluster containing most of the accessions from the Matooke breeding program together with a few accessions from all the other breeding programs. Similarly, the materials from Plantain form their own distinct small cluster. The other three clusters each contained accessions from Mchare, Prata, and SwS programs. Prata accessions clustered closely with Mchare accessions, while SwS and Mchare breeding materials were scattered and seemed more diverse.

3.3.2 | Population structure analysis for all accessions

To understand the pattern of genetic structure in this panel of accessions DAPC based on a Bayesian information criterion were used to detect subpopulations or clusters among all the breeding materials (641 retained after filtering) from the five breeding programs, the diploids (327 retained) and tetraploids (108 retained). DAPC analysis detected a peak at $K = 2$, suggesting the presence of two clusters (Figure S4) for all the

TABLE 4 Analysis of molecular variance (AMOVA) showing the genetic differentiation within and among the different subsets as revealed by single nucleotide polymorphisms.

Source of variation		Degrees of freedom	Sum of squares	Variance components	Variation (%)	*F*- statistics	*p*-value
Breeding programs	Among	4	2.702	0.005	13.5	$F_{ST} = 0.140$	0.001
	Within	636	21.360	0.034	86.5	$F_{IS} = 0.860$	0.001
	Total	640	24.062	0.039	100		
Breeding programs; diploids	Among	4	2.044	0.008	15.9	$F_{ST} = 0.160$	0.001
	Within	322	14.279	0.044	84.1	$F_{IS} = 0.840$	0.001
	Total	326	16.324	0.053	100		
Breeding programs; tetraploids	Among	4	1.005	0.0124	30.6	$F_{ST} = 0.305$	0.001
	Within	103	2.901	0.028	69.4	$F_{IS} = 0.694$	0.001
	Total	107	3.906	0.04055219	100		

Abbreviations: F_{ST} = pairwise genetic differentiation between the groups; F_{IS} = inbreeding coefficient within the groups.

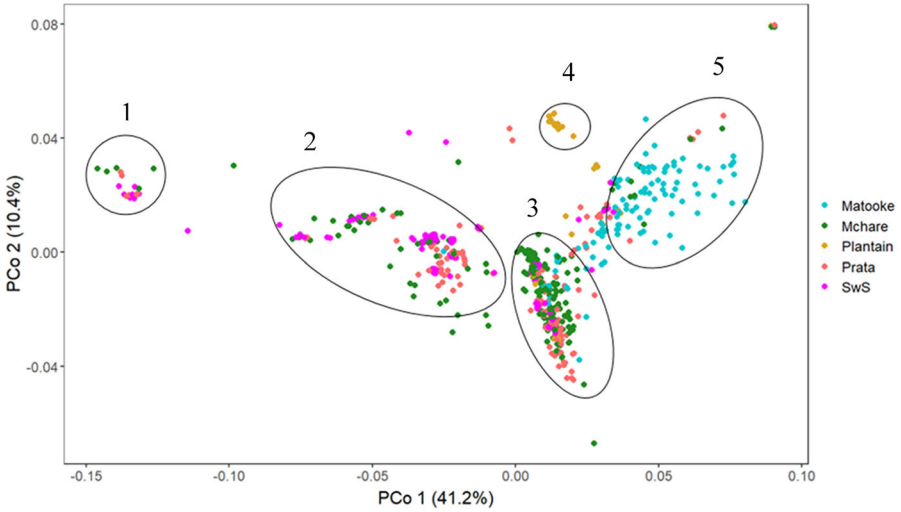

FIGURE 1 Genetic distance revealed by a principal coordinate analysis (PCoA) based on Nei's unbiased genetic distance of single nucleotide polymorphisms (SNP) among the accessions. A strong population structure was revealed by the PCoA based on diploids from each breeding program with the first and second principal coordinates explaining 87.1% and 2.3% of the total variation, respectively (Figure S2). Similarly, as observed for the entire panel of accessions, five major clusters were observed among diploids, with the major two containing diploids from all breeding programs, and the other three containing mainly accessions from Mchare, Prata, and SwS. Diploid accessions from Matooke formed a separate distinct cluster along the first axis. The PCoA based on tetraploids in each breeding program explained 35.2% and 15.8% of the total variation in the first and second principal coordinates, respectively, and revealed a clear population structure with five distinct clusters formed (Figure S3), each containing predominantly accessions from a distinct program. The tetraploid accessions from SwS are clearly separated from the rest of the populations along the PCo 1. The tetraploids from Prata formed two separate clusters. Two individuals (Sample_id 2522093 and 2522094) from Plantain clustered together with the tetraploids from Matooke population and one individual (Sample_id 2521907) from Matooke clustered with the tetraploids from the Plantain population.

subsets of the data. DAPC was also set to cluster according to the breeding program where the accession was sourced to further study the relationship between accessions from different programs. The membership probability of each accession to be assigned into different clusters was 100% for all accessions, and no admixture or accession with multiple affiliations was detected by DAPC analysis.

The DAPC scatter plot of the accessions showed a clear separation of accessions according to their breeding program, along the first discriminant function that comprised most of the genetic variation (Figure 2A). The breeding materials from Matooke, Plantain, and SwS each formed their own cluster, while Mchare and Prata grouped together as one cluster. This distinction was even more apparent when

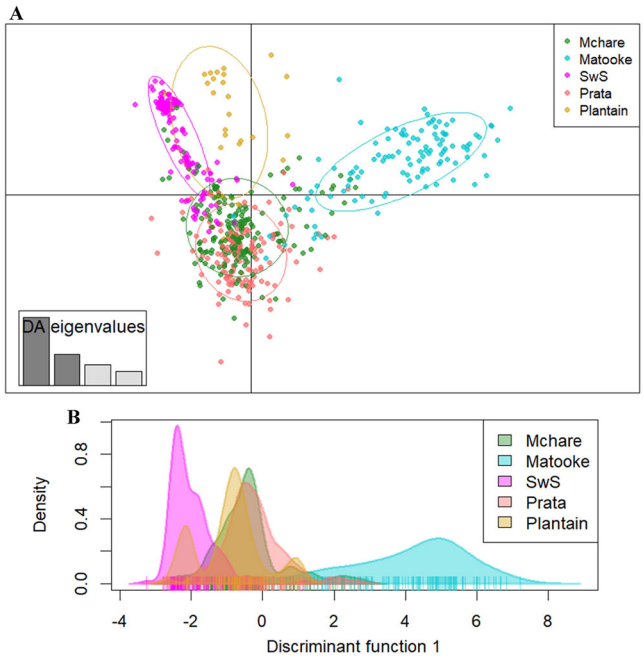

FIGURE 2 Discriminant analysis of principal components (DAPC). (A) Scatter plot of the 641 *Musa* accessions from the five breeding programs using two discriminant functions with an inset plot of discriminant analysis (DA) eigen values; (B) plot of the densities of individuals on the first discriminant function that displays cluster differences.

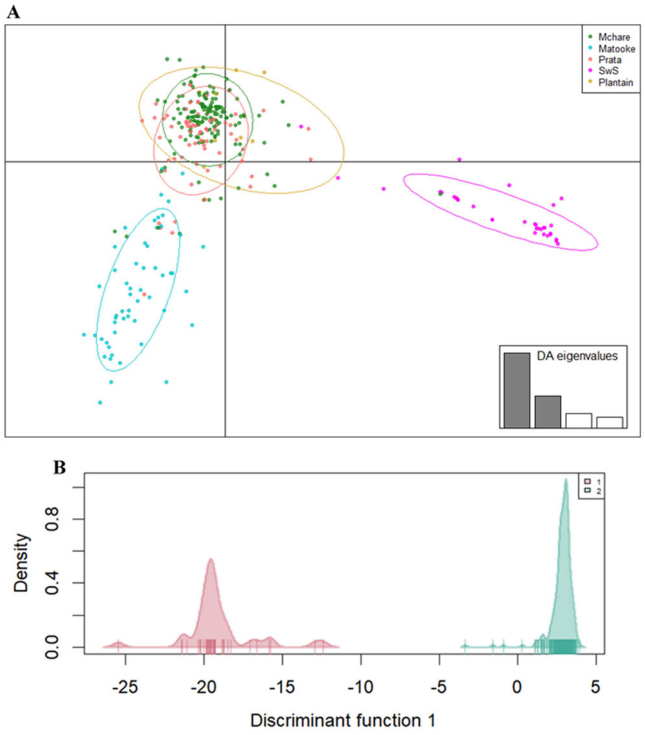

FIGURE 3 Discriminant analysis of principal components (DAPC). (A) Scatter plot of the 327 diploid *Musa* accessions from the five breeding programs using two discriminant functions with an inset plot of discrimant analysis (DA) eigen values and (B) plot of the densities of individuals on the first discriminant function that displays cluster differences.

densities were plotted only along the first discriminant function (Figure 2B). The plot showed a close relationship between accessions of Prata and those from the Mchare breeding program, which clustered closely together along both axes. This result is consistent with the PCoA.

3.3.3 | Population structure analysis for diploids and tetraploids

DAPC analysis for the diploid subsets identified two clusters as shown in the DAPC biplot and the plot of densities below with the first discriminant function explaining 92% of the variation (Figure 3A,B).

The DAPC analysis for the diploid accessions accounted for 84.9% of the cumulative variance using 50 principal components retained, and separated the 327 accessions into two clusters at $K = 2$ with 42 accessions (12.8%) in cluster 1, and the highest number of accessions (285 accessions, 87.2%) were assigned to cluster 2 (Figure 3A). The DAPC biplot on the first discriminant function showed a clear separation of the diploid accessions into two distinct clusters at $K = 2$ (Figure 3A). Cluster 1 comprised 10 accessions sourced from Prata, 20 accessions from SwS, 2 accessions from plantain, 11 accessions from Mchare, and no accessions from Matooke breeding programs. Cluster 2 consisted of 13 accessions sourced from SwS, 146 accessions from Mchare, 61 from Prata, 10 from plantain, and all the 55 accessions from the Matooke breeding program (Table S3). Both clusters contained almost equal numbers of cultivar and wild diploids, while diploids classified as "Hybrid" and "Improved 2x" type were in cluster 2. The diploids clustered according to their genome group with all accessions in cluster 1 from the B genome except for four accessions. Noticeably, two banksii accessions (ITC0467 and ITC0806) clustered together with balbisiana accessions in cluster 1. All accessions in cluster 2 are from the A genome except for two (Pisang Nangka and Eti Kehel annotated as "A/B" and "BB," respectively). The highest within-population variation or gene diversity (He) was observed in cluster 1 (He = 0.15) followed by cluster 2 (He = 0.11), while the pairwise F_{ST} distance between the clusters was 0.02 (Table S3).

The DAPC biplot (Figure 4A) together with the plot of densities of individuals on the first discriminant function (Figure 4B) showed a clear separation of the 108 tetraploid accessions into the two clusters with no admixed individuals, explaining 71% of the variation. Cluster 1 was the largest and contained 77 accessions (71.3%) and cluster 2 contained 31 accessions (28.7%; Figure 4A). The highest number of accessions observed in cluster 1 was from breeding programs of

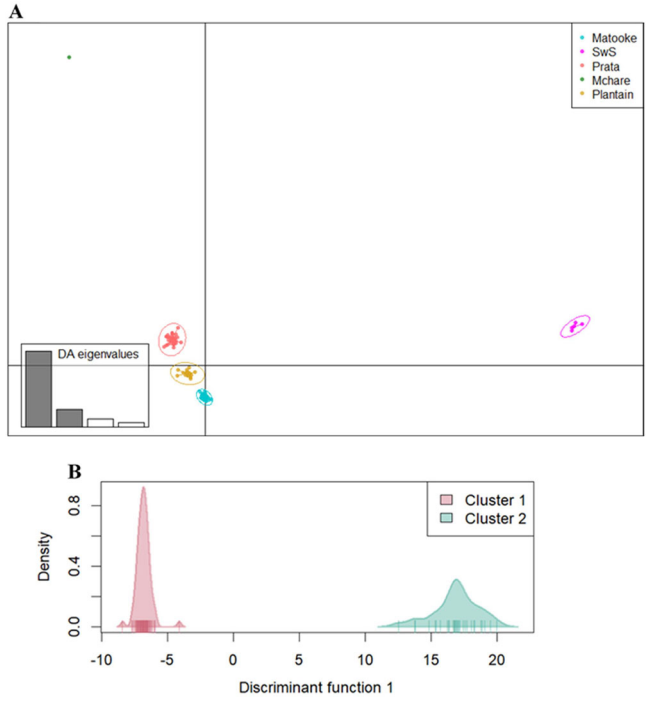

FIGURE 4 Discriminant analysis of principal components (DAPC). (A) Scatter plot of the 108 tetraploid *Musa* accessions from the five breeding programs using two discriminant functions with an inset plot of discriminant analysis (DA) eigen values and (B) plot of the densities of individuals on the first discriminant function that displays cluster differences.

Matooke (57.1%) followed by Plantain and Prata (23.4% and 19.5%, respectively). In cluster 2, the highest number of accessions observed was from Prata (80.6%), followed by SwS (16.1%), and none from plantain and Matooke. Gene diversity was 0.14 and 0.09 for clusters 1 and 2, respectively, and the pairwise F_{ST} distance between the clusters was 0.01 (Table S3). All the accessions in cluster 1 were from the A genome and except for the 13 PITA hybrids from Plantain that are from the AB genome. Accessions in cluster 2 were all from the AB genome. Noticeably, all the tetraploids from Matooke clustered together in cluster 1. The single tetraploid from Mchare clustered together with five tetraploids from SwS and 25 tetraploids from Prata. The clustering of tetraploids is observed to be based on genome group and breeding aim or product, with all the plantain-improved hybrids from Plantain (PITAs) except for one and the matooke-derived hybrids clustering together in cluster 2. The cluster allocations per diploid and tetraploid accession according to DAPC are summarized in Table S1.

4 | DISCUSSION

Information on the genetic diversity and structure of breeding germplasm is of great importance to breeders in order to make informed crosses among diverse parents for the development of new and improved cultivars. This study provides the first overview of the genetic variation in *Musa* breeding populations from five of the seven global banana breeding programs using DArTseq SNP markers.

4.1 | Marker informativeness

The 16,903 high-quality retained DArTseq SNPs were informative with a PIC value ≥0.25 (Botstein et al., 1980; Serrote et al., 2020) for 34.8% of the markers. The percentage of polymorphic loci of more than 59% and an MAF larger than or equal to 0.01 for all markers indicate that the retained SNPs are reliable to detect genetic variation in bananas (Luo et al., 2019). This is relevant for future genomic investigations in banana breeding efforts and demonstrates the possibility of using the selected DArTseq-SNP markers for genomic investigations such as linkage mapping, individual identification and quality control in *Musa*. This may serve as a foundation for future breeding and conservation efforts.

4.2 | Population structure and relationships

The different approaches to detecting existing population structure in the panel of germplasm appeared to provide complementary information. DAPC performed well in detecting clusters of diversity, and results were confirmed by PCoA. While PCoA depicted five clusters and DAPC revealed two major clusters with each cluster containing accessions from each breeding program further showing the low population differentiation. Among the diploids, cluster 1 from DAPC corresponds to clusters 4 and 5 of the PCoA, and cluster 2 corresponds to clusters 1, 2, and 3 of the PCoA in Figure S2. For the tetraploids, cluster 1 from DAPC corresponds to clusters 3, 4, and 5, while cluster 2 corresponds to clusters 1 and 2 of the PCoA. The low genetic differentiation between populations as indicated by the low F_{ST} values between the DAPC clusters could be due to large gene pool exchanges (Eltaher et al., 2018), through the use of common diploids from the oldest banana breeding program from Honduras, by all the breeding programs. Moreover, the recent funding of the banana breeding programs in Africa has improved germplasm exchange between the banana breeding programs, with material mostly going to Mchare and Matooke breeding programs. Accessions from the Mchare breeding program were to a higher extent spread across all clusters, indicating that this population was more diverse due to germplasm exchange.

The low F_{ST}, AMOVA, and DAPC suggested a separation of no more than two clusters. The five groups suggested by the PCoA show that a genetic structure already exists or accessions from different breeding programs might have a

common ancestor (Luo et al., 2019; Meirmans, 2015). This study demonstrated that the majority of genetic variance exist within breeding programs rather than between them. This was evident in the AMOVA and the DAPC clustering, for example, as the cultivars grouped together irrespective of their source breeding program. Similar results were observed in white Guinea yam where landraces from different geographical regions grouped together (Agre et al., 2021; Bhattacharjee et al., 2020). The tetraploid population was more structured as it formed distinct clusters in the PCoA, indicating that variation within the tetraploids is lower compared to between them.

The accessions in each group were observed to cluster according to genome group irrespective of their ploidy levels. This could be explained by the ancestry and evolution of the banana cultivars through the intra- and interspecific hybridization of the two wild species *M. acuminata* Colla (AA) and *M. balbisiana* Colla (BB; Manzo-Sánchez et al., 2015; Nayar, 2010), complex interspecific chromosome mosaic patterns in the ancestral groups that contributed to the cultivated banana genomes (Martin et al., 2020), and/or structural variations between the A and B genomes (Baurens et al., 2018). Our findings are supported by similar clustering trends based on genome groups in the Philippino *Musa* gene pool using SNPs (Gardoce et al., 2023). In that research, similar to our observations, two major clusters directly corresponding to the ploidy level of the B genome groups and a single major cluster containing A genome groups irrespective of ploidy levels were observed. Also, Onyango et al. (2010) reported distinct clusters in East African "Apple Banana" (AAB genome) and "Muraru" (AA genome) dessert bananas using microsatellite markers, corresponding to their genome composition. The presence of wild diploids in two different clusters in our study seems to be explained entirely by the presence or absence of the B genome. Similarly, clustering dictated primarily by the presence or absence of the B genome has been reported in other studies (De Jesus et al., 2013; Doloiras-Laraño et al., 2018). Tetraploids in this study clustered based on breeding aim or product. This indicates that regional preferences influence breeding targets and shape the diversity of breeding products. Hence, selection pressure toward those preferences creates distinct groups of parents used in the development of the end-user-preferred products. This unique variation within the breeding programs is because product profiles describe an ideal variety with the necessary characteristics intended to replace the older varieties that still dominate a particular market.

Twenty-one accessions from the Matooke breeding program observed to cluster with some accessions from the Prata breeding program were improved diploids originally from Brazil but under field evaluation in Uganda as indicated by their passport data. Likewise, the diversity of Mchare came mostly from 2x accessions imported from ITC. Female diploid parents used in the Mchare breeding program normally referred to as "Mchare" clustered together in cluster 2 by DAPC, indicating they are genetically similar. The same observation was made by Christelová et al. (2017), where the AA cv. African set that contained the "Mchare" formed a distinct cluster IX. Evidence from our results indicated a very close relationship between the Mchare (referring to the Mchare diploid cultivars used as female parents) from Tanzania and the Cavendish and Prata of Brazil, as well as with the Silk and Pome bananas of India. There is also a close relationship between the SwS and the Prata bananas. This offers more support to the hypothesis that Mchare is the donor of the 2n gamete contributing "zebrina" and "banksii" to the Cavendish genome (Hippolyte et al., 2012; Raboin et al., 2005) and more recently, also shown to have contributed a large "malaccensis" component to Cavendish (Martin et al., 2020). Martin et al. (2023) confirmed that both Gros Michel and Cavendish resulted from transmission of a 2x gamete from Mchare with full genome restitution without recombination.

The tetraploid "Sample_id 2521907" (accession name: 25974S-17), which was stated to be Matooke-derived (AA) yet clustered with accessions of the "AB" genome like the plantains from Nigeria, seems to be a case of mislabeling during field establishment. This tetraploid in the crossing blocks at IITA in Uganda research fields does not match the Matooke descriptors (Batte et al., 2018) and has exhibited a phenotype similar to the PITA plantain hybrid (Tenkouano et al., 2019). The two tetraploids from Plantain that clustered together with those from Matooke were identified as chromosome-doubled diploids (AA), which explains their grouping with the AAAA-Matooke accessions. The 13 tetraploid improved plantain hybrids called "PITA" were in the same cluster with improved tetraploids from the Matooke program. This could be because for these tetraploids, Calcutta 4 is a common parent that was used for their development as a source of resistance to diseases and pests (Brown et al., 2017; Ortiz, 2015; Swennen & Vuylsteke, 1993; Tenkouano et al., 2003). The single tetraploid accession that was sourced from the Mchare breeding program always clustered with accessions from the Prata program. This accession identified as PV 42–53, and from passport data mined from MGIS is originally from Brazil and is not used in the Mchare breeding program for any breeding purposes. The PCoA scores indicate a presence of two kinds of Calcutta 4 accessions. The samples from the Mchare and Plantain breeding programs ("2521592$2522162" and "2522085," respectively) are different from the sample from the Matooke program ("2521894"). This was confirmed by the genetic distance matrix of the diploids (data not attached). These look phenotypically the same. We are following up on these discrepancies. Thus, the discriminatory powers of this set of SNP markers to detect molecular differences and similarities could be utilized by breeders for quality control in the breeding germplasm.

4.3 | Genetic differentiation of populations

A low differentiation between populations was revealed by the low estimated F_{ST} for the breeding programs, low to high for diploids and a high to very high differentiation between the tetraploids. This indicates that the only difference among these populations could be due to the differences in quality traits targeted by the different breeding programs. According to Wright (1968) and Luo et al. (2019), F_{ST} values greater or equal to 0.15 are considered high and significant for discriminating individuals within a population indicating the presence of subpopulations. A significant and very high F_{ST} was observed between breeding materials from SwS and Matooke, diploids from SwS and all the other programs and similarly between tetraploids from SwS and all the other populations. This could be due to the use of common diploid parents by all other programs except for SwS (Amorim, dos Santos-Serejo, et al., 2011; Batte et al., 2019; Brown et al., 2017; Ortiz, 2015). The F_{ST} value was the lowest between populations from Prata and Plantain, and between Prata and Mchare, an observation that supports suggestions that Mchare bananas are common ancestors for dessert bananas, Cavendish, Gros Michel, Prata, and Silk (Christelová et al., 2017; Hippolyte et al., 2012; Perrier et al., 2011), which except for Gros Michel are also bred for in Brazil.

The AMOVA results similarly imply a high genetic variation among the accessions within each breeding program and also demonstrate the unique alleles within each program probably due to the unique product profiles targeted by each program. This is more distinct, especially for the quality traits such as taste and texture that are unique to consumers as reported for the East African Highland cooking bananas representative of the Matooke and Mchare banana breeding programs (Madalla et al., 2023; Marimo et al., 2019; Sanya et al., 2020). For example, among the tetraploids, 38% are Matooke-derived tetraploids specifically selected for the Matooke end-user product profile, and 16.5% are plantain-improved tetraploids targeted for the plantain product profile.

4.4 | Implications to the breeding programs

The definition of source of breeding material in this study is limited to and defined as the breeding program in which the accession was collected, and it is known that some breeding materials have been used in more than a single breeding program. The low genetic diversity is compounded by the repeated use of a few parental clones and their progeny as parents in breeding schemes even across different breeding programs. For example, improved diploids SH 3142, SH 3217, SH 3362, TMB2 × 9128-3, and TMB2 × 7197-2 are used across all IITA and NARO breeding programs with SH 3362 also used at Embrapa-Brazil (Amorim, dos Santos-Serejo, et al., 2011; Batte et al., 2019; Brown et al., 2009; Ortiz, 2015). Also, Calcutta 4 has been utilized for decades as a source of resistance to the Sigatoka complex, yellow Sigatoka, fusarium wilt, banana weevil, and burrowing nematodes (Krishnamoorthy & Kumar, 2005; Ortiz, 2015; Tenkouano et al., 2003) across banana breeding programs.

There is a need for strategies for parental improvement especially for the diploids to identify and increase useful diversity and genetic variation if genetic gain is to be realized in banana breeding (Sanchez et al., 2023). The clustering in this study provides important clues for how to increase useful diversity (variation for key traits) in breeding programs. For example, all improved diploids clustered together in cluster 1 irrespective of their source, but the high and significant F_{ST} value of 0.90 for accessions within this cluster (Table S1) is indicative of high differentiation and perhaps existence of high genetic variation among these diploids. Five of these improved diploids (CNPMF 0731, CNPMF 0513, CNPMF 0998, CNPMF 1323, and 013019-01) from the Prata breeding program have been reported to have resistance to Fusarium wilt race 1 (Gonçalves et al., 2019). These diploids, all from the A genome might be good candidates to explore as another source of resistance genes for Fusarium, for the benefit of the Mchare breeding program. The Prata breeding program, which also breeds for plantains, had seven improved diploids reported to have resistance to black Sigatoka (Gonçalves et al., 2021), and three of these (M53, CNPMF 0496, and CNPMF 0519) also carry resistance to weevils and nematodes (E. Amorim, Embrapa, unpublished data, 2023). The use of these seven diploids could benefit both the plantain and Matooke improvement program as new sources of resistance genes for either hybrid development for plantain or parental improvement for both plantain and matooke. The use of unrelated parents in crossing schemes would increase genetic variation for new cultivars, which is expressed as additive genetic variance in the breeder's equation. Since genetic variation is a great contributor to the genetic gain equation, its increase would increase overall genetic gain in bananas. This study will aid the breeders from all the banana breeding programs represented here to make more informed decisions regarding the choice of diverse and unique parents.

AUTHOR CONTRIBUTIONS

Violet Akech: Conceptualization; data curation; formal analysis; investigation; writing—original draft; writing—review and editing. **Therése Bengtsson**: Conceptualization; data curation; formal analysis; investigation; supervision; writing—review and editing. **Rodomiro Ortiz**: Conceptualization; investigation; methodology; writing—review and editing. **Rony Swennen**: Conceptualization; funding

acquisition; methodology; project administration; writing—review and editing. **Brigitte Uwimana**: Methodology; writing—review and editing. **Claudia F. Ferreira**: Writing—review and editing. **Delphine Amah**: Writing—review and editing. **Edson P. Amorim**: Writing—review and editing. **Elizabeth Blisset**: Data curation. **Ines Van den Houwe**: Writing—review and editing. **Ivan K. Arinaitwe**: Writing—review and editing. **Liana Nice**: Data curation; writing—review and editing. **Priver Bwesigye**: Writing—review and editing. **Steve Tanksley**: Writing—review and editing. **Subbaraya Uma**: Writing—review and editing. **Backiyarani Suthanthiram**: Writing—review and editing. **Marimuthu S. Saraswathi**: Writing—review and editing. **Hassan Mduma**: Writing—review and editing. **Allan Brown**: Conceptualization; funding acquisition; investigation; methodology; project administration; supervision; writing—review and editing.

ACKNOWLEDGMENTS

This study was part of the PhD research of the first author, with funding from the IITA coordinated project "Accelerated Breeding for Better Bananas (ABBB)." The authors would like to thank all donors who supported this work through contribution to the CGIAR Fund. The authors would furthermore like to thank Rosemary Simba, Awio Miriam, and Temitope Ademulegun in the IITA Banana Breeding programs and all others who helped to collect samples from the fields and prepare the plant materials for shipping.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The list of all plant entries and their metadata (Table S1), the unfiltered genotypic data as *vcf* file (named "banana_2024"), filtered genotype data as *vcf* file (named "Allfilt_2024") and the locus Metadata as *csv* file (named "Locusinfo_2024") is available on figshare.com and datadryad.org https://figshare.com/s/b21422f6f40b11a2f1f6 Reserved figshare DOI: 10.6084/m9.figshare.25442146 https://doi.org/10.5061/dryad.c2fqz61jk.

ORCID

Violet Akech https://orcid.org/0000-0003-2330-0682
Rodomiro Ortiz https://orcid.org/0000-0002-1739-7206
Rony Swennen https://orcid.org/0000-0002-5258-9043

REFERENCES

Agre, P. A., Dassou, A. G., Loko, L. E., Idossou, R., Dadonougbo, E., Gbaguidi, A., Mondo, J. M., Muyideen, Y., Adebola, P. O., & Asiedu, R. (2021). Diversity of white Guinea yam (*Dioscorea rotundata* Poir.) cultivars from Benin as revealed by agro-morphological traits and SNP markers. *Plant Genetic Resources*, *19*(5), 437–446. https://doi.org/10.1017/s1479262121000526

Amorim, E. P., Amorim, V. B. O., Silva, S. O., & Pillay, M. (2011). Quality improvement of cultivated Musa. In M. Pillay & A. Tenkouano (Eds.), *Banana breeding: progress and challenges* (pp. 252–280). CRC Press.

Amorim, E. P., dos Santos-Serejo, J. A., Amorim, V. B. O., Ferreira, C. F., & Silva, S. O. (2011, Oct. 10). *Banana breeding at Embrapa cassava and fruits* [Paper presentation]. VII International Symposium on Banana: ISHS-ProMusa Symposium on Bananas and plantains: Towards Sustainable Global Production, Salvador, Brazil.

Batte, M., Mukiibi, A., Swennen, R., Uwimana, B., Pocasangre, L., Hovmalm, H. P., Geleta, M., & Ortiz, R. (2018). Suitability of existing Musa morphological descriptors to characterize East African highland 'matooke' bananas. *Genetic Resources and Crop Evolution*, *65*(2), 645–657. https://doi.org/10.1007/s10722-017-0562-9

Batte, M., Swennen, R., Uwimana, B., Akech, V., Brown, A., Tumuhimbise, R., Hovmalm, H. P., Geleta, M., & Ortiz, R. (2019). Crossbreeding East African highland bananas: Lessons learnt relevant to the botany of the crop after 21 years of genetic enhancement. *Frontiers in Plant Science*, *10*, 81. https://doi.org/10.3389/fpls.2019.00081

Baurens, F. C., Martin, G., Hervouet, C., Salmon, F., Yohomé, D., Ricci, S., Rouard, M., Habas, R., Lemainque, A., Yahiaoui, N., & D'Hont, A. (2018). Recombination and large structural variations shape interspecific edible bananas genomes. *Molecular Biology and Evolution*, *36*(1), 97–111. https://doi.org/10.1093/molbev/msy199

Bawin, Y., Panis, B., Vanden Abeele, S., Li, Z., Sardos, J., Paofa, J., Ge, X.-J., Mertens, A., Honnay, O., & Janssens, S. B. (2019). Genetic diversity and core subset selection in ex situ seed collections of the banana crop wild relative *Musa balbisiana*. *Plant Genetic Resources*, *17*(6), 536–544. https://doi.org/10.1017/S1479262119000376

Belser, C., Baurens, F.-C., Noel, B., Martin, G., Cruaud, C., Istace, B., Yahiaoui, N., Labadie, K., Hřibová, E., Doležel, J., Lemainque, A., Wincker, P., D'Hont, A., & Aury, J. M. (2021). Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. *Communications Biology*, *4*(1), 1047. https://doi.org/10.1038/s42003-021-02559-3

Bhattacharjee, R., Agre, P., Bauchet, G., De Koeyer, D., Lopez-Montes, A., Kumar, P. L., Abberton, M., Adebola, P., Asfaw, A., & Asiedu, R. (2020). Genotyping-by-sequencing to unlock genetic diversity and population structure in white yam (*Dioscorea rotundata* Poir.). *Agronomy*, *10*(9), 1437. https://doi.org/10.3390/agronomy10091437

Blench, R. (2009). Bananas and plantains in Africa: Re-interpreting the linguistic evidence. *Ethnobotany Research and Applications*, *7*, 363–380. https://doi.org/10.17348/era.7.0.363-380

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, *32*(3), 314.

Brown, A., Tumuhimbise, R., Amah, D., Uwimana, B., Nyine, M., Mduma, H., Talengera, D., Karamura, D., Kuriba, J., & Swennen, R. (2017). Bananas and plantains (*Musa* spp.). *Genetic improvement of tropical crops* (pp. 219–240). Springer. https://doi.org/10.1007/978-3-319-59819-2_7

Castillo, C., & Fuller, D. Q. (2012). Bananas: The spread of a tropical forest fruit as an agricultural staple. In J. L. Thorp & M. A. Katzenberg (Eds.), *The Oxford handbook of the archaeology of diet* (pp. 518–538). Oxford University Press.

Christelová, P., De Langhe, E., Hřibová, E., Čížková, J., Sardos, J., Hušáková, M., Van den houwe, I., Sutanto, A., Kepler, A. K.,

Swennen, R., Roux, N., & Doležel, J. (2017). Molecular and cytological characterization of the global Musa germplasm collection provides insights into the treasure of banana diversity. *Biodiversity and Conservation*, *26*(4), 801–824. https://doi.org/10.1007/s10531-016-1273-9

Deepthi, V. P. (2016). Taxonomic scoring and genomic grouping in bananas. *Flora and Fauna*, *22*, 151–158.

De Jesus, O. N., Silva, S. D. O. E., Amorim, E. P., Ferreira, C. F., de Campos, J. M. S., Silva, G. D. G., & Figueira, A. (2013). Genetic diversity and population structure of *Musa* accessions in ex situ conservation. *BMC Plant Biology*, *13*(1), 41. https://doi.org/10.1186/1471-2229-13-41

De Langhe, E. (1964). The origin of variation in the Plantain Banana. *Overdruk uit de Mededelingen van de Landbouwhogeschool en de Opzoekingsstations van de Staat te Gent*, *29*(1), 45–80.

De Langhe, E. (2007). The establishment of traditional plantain cultivation in the African rain forest: A working hypothesis. In J. I. Timothy, P. Denham, & L. Vrydaghs (Eds.), *Rethinking agriculture* (pp. 361–370). Routledge.

De Langhe, E., & de Maret, P. (2004). Tracking the banana: Its significance in early agriculture. *The prehistory of food* (pp. 387–404). Routledge. https://doi.org/10.4324/9780203203385-33

De Langhe, E., Swennen, R., & Vuylsteke, D. (1994). Plantain in the early Bantu world. *Azania: Archaeological Research in Africa*, *29*(1), 147–160. https://doi.org/10.1080/00672709409511667

De Langhe, E., Vrydaghs, L., De Maret, P., Perrier, X., & Denham, T. (2009). Why bananas matter: An introduction to the history of banana domestication. *Ethnobotany Research and Applications*, *7*, 165–177. https://doi.org/10.17348/era.7.0.165-177

Doloiras-Laraño, A. D., Garcia, R. N., Sandoval, C. M. C., Lalusin, A. G., Gueco, L. S., Huelgas, V. C., & Tecson-Mendoza, E. M. (2018). DNA fingerprinting and genetic diversity analysis of Philippine Saba and other cultivars of *Musa balbisiana* Colla using simple sequence repeat markers. *Philippines Journal Crop Science*, *43*, 1–11.

Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*, 1–20. https://doi.org/10.18637/jss.v022.i04

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F., Poland, J., & Baenziger, P. S. (2018). Genetic diversity and population structure of F_3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing. *Frontiers in Genetics*, *9*, 76. https://doi.org/10.3389/fgene.2018.00076

Evans, E. A., Ballen, F. H., & Siddiq, M. (2020). Banana production, global trade, consumption trends, postharvest handling, and processing. In *Handbook of banana production, postharvest science, processing technology, and nutrition* (pp. 1–18). Wiley. https://doi.org/10.1002/9781119528265.ch1

FAO. (2022). *Crops production and trade statistics*. https://www.fao.org/faostat/en/#data/QCL

Fuller, D. Q., & Madella, M. (2009). Banana cultivation in South Asia and East Asia: A review of the evidence from archaeology and linguistics. *Ethnobotany Research and Applications*, *7*, 333–351. https://doi.org/10.17348/era.7.0.333-351

Gardoce, R. R., Manohar, A. N. C., Mendoza, J. V. S., Tejano, M. S., Nocum, J. D. L., Lachica, G. C., Gueco, L. S., Cueva, F. M. D., & Lantican, D. V. (2023). A novel SNP panel developed for targeted genotyping-by-sequencing (GBS) reveals genetic diversity and population structure of *Musa* spp. germplasm collection. *Molecular Genetics and Genomics*, *298*(4), 857–869. https://doi.org/10.1007/s00438-023-02018-0

Gonçalves, Z. S., de Jesus Rocha, A., Haddad, F., de Oliveira Amorim, V. B., Ferreira, C. F., & Amorim, E. P. (2021). Selection of diploid and tetraploid banana hybrids resistant to *Pseudocercospora fijiensis*. *Agronomy*, *11*(12), 2483. https://doi.org/10.3390/agronomy11122483

Gonçalves, Z. S., Haddad, F., de Oliveira Amorim, V. B., Ferreira, C. F., de Oliveira, S. A. S., & Amorim, E. P. (2019). Agronomic characterization and identification of banana genotypes resistant to Fusarium wilt race 1. *European Journal of Plant Pathology*, *155*(4), 1093–1103. https://doi.org/10.1007/s10658-019-01837-5

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*(3), 691–699. https://doi.org/10.1111/1755-0998.12745

Hippolyte, I., Jenny, C., Gardes, L., Bakry, F., Rivallan, R., Pomies, V., Cubry, P., Tomekpé, K., Risterucci, A.-M., & Roux, N. (2012). Foundation characteristics of edible Musa triploids revealed from allelic distribution of SSR markers. *Annals of Botany*, *109*(5), 937–951. https://doi.org/10.1093/aob/mcs010

Jeensae, R., Kongsiri, N., Fluch, S., Burg, K., & Boonruangrod, R. (2021). Cultivar specific gene pool may play an important role in *Musa acuminata* Colla evolution. *Genetic Resources and Crop Evolution*, *68*(4), 1589–1601. https://doi.org/10.1007/s10722-020-01088-y

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 1–15. https://doi.org/10.1186/1471-2156-11-94

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *Peer Journal*, *2*, e281. https://doi.org/10.7717/peerj.281

Karamura, D. A. (1998). *Numerical taxonomic studies of the East African Highland bananas (Musa AAA-East Africa) in Uganda*. University of Reading.

Kassambara, A., & Mundt, F. (2020). *Factoextra: extract and visualize the results of multivariate data analyses*. R package version 1.0.7.

Kilian, A., Sanewski, G., & Ko, L. (2016). The application of DArTseq technology to pineapple. *Acta Horticulturae*, *1111*, 181–187. https://doi.org/10.17660/actahortic.2016.1111.27

Kitavi, M., Downing, T., Lorenzen, J., Karamura, D., Onyango, M., Nyine, M., Ferguson, M., & Spillane, C. (2016). The triploid East African Highland Banana (EAHB) genepool is genetically uniform arising from a single ancestral clone that underwent population expansion by vegetative propagation. *Theoretical and Applied Genetics*, *129*, 547–561. https://doi.org/10.1007/s00122-015-2647-1

Krishnamoorthy, V., & Kumar, N. (2005). Preliminary evaluation of diploid banana hybrids for yield potential, male fertility and reaction to *Radopholus similis*. *Plant Genetic Resources Newsletter*, *141*, 39–43.

Lejju, B. J., Robertshaw, P., & Taylor, D. (2006). Africa's earliest bananas? *Journal of Archaeological Science*, *33*(1), 102–113. https://doi.org/10.1016/j.jas.2005.06.015

Lescot, T. (2020). Banana genetic diversity. *Close-Up Fruitrop*, *269*, 98–102.

Luo, Z., Brock, J., Dyer, J. M., Kutchan, T., Schachtman, D., Augustin, M., Ge, Y., Fahlgren, N., & Abdel-Haleem, H. (2019). Genetic diversity and population structure of a *Camelina sativa* spring panel. *Frontiers in Plant Science*, *10*, 184. https://doi.org/10.3389/FPLS.2019.00184

Madalla, N. A., Swennen, R., Brown, A., Carpentier, S., Van den Bergh, I., Crichton, R., Marimo, P., Weltzien, E., Massawe, C., & Shimwela, M. (2023). Farmers' preferences for East African highland cooking banana 'Matooke' hybrids and local cultivars. *Agriculture & Food Security*, *12*(1), 1–20. https://doi.org/10.1186/s40066-023-00407-7

Manzo-Sánchez, G., Buenrostro-Nava, M., Guzmán-González, S., Orozco-Santos, M., Youssef, M., & Medrano, R. M. E.-G. (2015). Genetic diversity in bananas and plantains (*Musa* spp.). *Molecular approaches to genetic diversity* (pp. 93–121). IntechOpen. https://doi.org/10.5772/59421

Marimo, P., Karamura, D., Tumuhimbise, R., Shimwela, M., Bergh, I. v. d., Batte, M., Massawe, C., Okurut, A., Mbongo, D., & Crichton, R. (2019). *Post-harvest use of banana in Uganda and Tanzania: Product characteristics and cultivar preferences of male and female farmers* (RTB Working Paper 2019-3). International Potato Center.

Martin, G., Baurens, F.-C., Labadie, K., Hervouet, C., Salmon, F., Marius, F., Paulo-de-la-Reberdiere, N., Van den Houwe, I., Aury, J. M., D'Hont, A., & Yahiaoui, N. (2023). Shared pedigree relationships and transmission of unreduced gametes in cultivated banana. *Annals of Botany*, *131*, 1149–1161. https://doi.org/10.1093/aob/mcad065

Martin, G., Cardi, C., Sarah, G., Ricci, S., Jenny, C., Fondi, E., Perrier, X., Glaszmann, J. C., D'Hont, A., & Yahiaoui, N. (2020). Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *The Plant Journal*, *102*(5), 1008–1025. https://doi.org/10.1111/tpj.14683

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, *24*(13), 3223–3231. https://doi.org/10.1111/mec.13243

Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C., & Georges, A. (2022). dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. *Methods in Ecology and Evolution*, *13*(10), 2150–2158. https://doi.org/10.1111/2041-210X.13918

Nayar, N. (2010). The bananas: Botany, origin, dispersal. *Horticultural Reviews*, *36*(3), 117–164. https://doi.org/10.1002/9780470527238.ch2

Nowakunda, K., Khakasa, E., Ceballos, H., Kenneth, A., Tumuhimbise, R., Bugaud, C., Asasira, M., Uwimana, B., Bouniol, A., Nuwamanya, E., Forsythe, L., Marimo, P., Dufour, D., & Tushemereirwe, W. (2023). East African highland cooking banana: Towards an efficient selection of hybrids with user-preferred food quality traits. *Journal of the Science of Food and Agriculture*, *104*(8), 4551–4560. https://doi.org/10.1002/jsfa.13070

Nyine, M., Uwimana, B., Swennen, R., Batte, M., Brown, A., Christelová, P., Hribová, E., Lorenzen, J., & Dolezel, J. (2017). Trait variation and genetic diversity in a banana genomic selection training population. *PLoS ONE*, *12*(6), 0178734. https://doi.org/10.1371/journal.pone.0178734

Onyango, M., Haymer, D., Keeley, S., & Manshardt, R. (2010). Analysis of genetic diversity and relationships in East African 'Apple Banana' (AAB genome) and 'Muraru' (AA genome) dessert bananas using microsatellite markers. *Acta Horticulturae*, *879*, 623–636. https://doi.org/10.17660/ActaHortic.2010.879.69

Ortiz, R. (1997). Secondary polyploids, heterosis, and evolutionary crop breeding for further improvement of the plantain and banana (*Musa* spp. L.) genome. *Theoretical and Applied Genetics*, *94*, 1113–1120. https://doi.org/10.1007/s001220050524

Ortiz, R. (2015). *Plant breeding in the omics era*. Springer International Publishing. https://doi.org/10.1007/978-3-319-20532-8

Ortiz, R., Ferris, R., & Vuylsteke, D. (1995). Banana and plantain breeding. In S. Gowen (Ed.), *Bananas and plantains* (pp. 110–146). Chapman & Hall.

Osterman, J., Hammenhag, C., Ortiz, R., & Geleta, M. (2021). Insights into the genetic diversity of Nordic red clover (*Trifolium pratense*) revealed by SeqSNP-based genic markers. *Frontiers in Plant Science*, *12*, 748750. https://doi.org/10.3389/fpls.2021.748750

Perrier, X., De Langhe, E., Donohue, M., Lentfer, C., Vrydaghs, L., Bakry, F., Carreel, F., Hippolyte, I., Horry, J.-P., & Jenny, C. (2011). Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proceedings of the National Academy of Sciences*, *108*(28), 11311–11318. https://doi.org/10.1073/pnas.1102001108

Perrier, X., Jenny, C., Bakry, F., Karamura, D., Kitavi, M., Dubois, C., Hervouet, C., Philippson, G., & De Langhe, E. (2019). East African diploid and triploid bananas: A genetic complex transported from South-East Asia. *Annals of Botany*, *123*(1), 19–36. https://doi.org/10.1093/aob/mcy156

Raboin, L.-M., Carreel, F., Noyer, J.-L., Baurens, F.-C., Horry, J.-P., Bakry, F., Montcel, H. T. D., Ganry, J., Lanaud, C., & Lagoda, P. J. (2005). Diploid ancestors of triploid export banana cultivars: Molecular identification of 2 n restitution gamete donors and n gamete donors. *Molecular Breeding*, *16*, 333–341. https://doi.org/10.1007/s11032-005-2452-7

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, C. P., Daniells, J., Dowiya, B., Effa Effa, B., Gueco, L., … 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

Sanchez, D., Sadoun, S. B., Mary-Huard, T., Allier, A., Moreau, L., & Charcosset, A. (2023). Improving the use of plant genetic resources to sustain breeding programs' efficiency. *Proceedings of the National Academy of Sciences, USA*, *120*(14), e2205780119. https://doi.org/10.1073/pnas.2205780119

Sanya, L. N., Sseguya, H., Kyazze, F. B., Diiro, G. M., & Nakazi, F. (2020). The role of variety attributes in the uptake of new hybrid bananas among smallholder rural farmers in central Uganda. *Agriculture & Food Security*, *9*, 1–13. https://doi.org/10.1186/s40066-020-00257-7

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 in Plant Science*, *13*, 969220. https://doi.org/10.3389/fpls/2022/969220

Sathiamoorthy, S., Uma, S., & Selvarajan, R. (2001). Banana research and development programme in India and highlights of NRCB-INIBAP collaborative projects. *Advancing Banana and Plantain Research & Development in Asia and the Pacific*, *10*, 103–106.

Serrote, C. M. L., Reiniger, L. R. S., Silva, K. B., dos Santos Rabaiolli, S. M., & Stefanel, C. M. (2020). Determining the polymorphism information content of a molecular marker. *Gene*, *726*, 144175. https://doi.org/10.1016/j.gene.2019.144175

Silva, S. O., Souza-Junior, M. T., Alves, E. J., Silveira, J. R. S., & Lima, M. B. (2001). Banana breeding program at Embrapa. *Crop Breeding and Applied Biotechnology*, *1*(4), 399–436. https://doi.org/10.13082/1984-7033.v01n04a09

Simmonds, N. (1962). Where our bananas come from. *New Scientist (Reed Business Information)*, *16*, 36–39.

Simmonds, N. W., & Shepherd, K. (1955). The taxonomy and origins of the cultivated bananas. *Botanical Journal of the Linnean Society*, *55*(359), 302–312. https://doi.org/10.1111/j.1095-8339.1955.tb00015.x

Swennen, R. (1990). Limits of morphotaxonomy: Names and synonyms of plantain in Africa and elsewhere. In R. Jarret (Ed.), *The identification of genetic diversity in the genus Musa. Proceedings of an International Workshop* (pp. 171–210). INIBAP.

Swennen, R., & Vuylsteke, D. (1993). Breeding black sigatoka resistant plantains with a wild banana. *Tropical Agriculture*, *70*(1), 74–77.

Tenkouano, A., Lamien, N., Agogbua, J., Amah, D., Swennen, R., Traoré, S., Thiemele, D., Aby, N., Kobenan, K., & Gnonhouri, G. (2019). Promising high-yielding tetraploid plantain-bred hybrids in West Africa. *International Journal of Agronomy*, *2019*, 3873198. https://doi.org/10.1155/2019/3873198

Tenkouano, A., Vuylsteke, D., Okoro, J., Makumbi, D., Swennen, R., & Ortiz, R. (2003). Registration of TMB2×5105-1 and TMB2×9128-3 diploid banana hybrids with good combining ability, partial resistance to black Sigatoka and resistance to nematodes. *HortScience*, *38*, 468–472. https://doi.org/10.21273/HORTSCI.38.3.468

Vuylsteke, D., Ortiz, R., Ferris, R. S. B., & Crouch, J. H. (2010). Plantain improvement. In J. Janick (Ed.), *Plant breeding reviews* (Vol. 14, pp. 267–320). John Wiley & Sons, Inc.

Wickham, H. (2009). Getting started with qplot. *ggplot2: elegant graphics for data analysis* (pp. 9–26). Springer. https://doi.org/10.1007/978-0-387-98141-3_2

Wright, S. (1968). *Evolution and the genetics of populations, 4 vols*. University of Chicago Press.

Yao, M., Chen, L., & Liang, Y. (2008). Genetic diversity among tea cultivars from China, Japan and Kenya revealed by ISSR markers and its implication for parental selection in tea breeding programmes. *Plant Breeding*, *127*(2), 166–172. https://doi.org/10.1111/j.1439-0523.2007.01448.x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Akech, V., Bengtsson, T., Ortiz, R., Swennen, R., Uwimana, B., Ferreira, C. F., Amah, D., Amorim, E. P., Blisset, E., Van den Houwe, I., Arinaitwe, I. K., Nice, L., Bwesigye, P., Tanksley, S., Uma, S., Suthanthiram, B., Saraswathi, M. S., Mduma, H., & Brown, A. (2024). Genetic diversity and population structure in banana (*Musa* spp.) breeding germplasm. *The Plant Genome*, e20497. https://doi.org/10.1002/tpg2.20497