Journal Pre-proofs

Analysis of genetic diversity and agronomic variation in banana sub-populations for genomic selection under drought stress in Southern Benin

Luther Fort Mbo Nkoulou, Lauriane Archange Tchinda Ninla, David Cros, Guillaume Martin, Zenabou Ndiang, Jordan Houegban, Hermine Bille Ngalle, Joseph Martin Bell, Enoch G. Achigan-Dako

PII: S0378-1119(23)00051-3
DOI: https://doi.org/10.1016/j.gene.2023.147210
Reference: GENE 147210

To appear in: *Gene Gene*

Received Date: 3 October 2022
Revised Date: 12 January 2023
Accepted Date: 13 January 2023

Please cite this article as: L. Fort Mbo Nkoulou, L. Archange Tchinda Ninla, D. Cros, G. Martin, Z. Ndiang, J. Houegban, H. Bille Ngalle, J. Martin Bell, E.G. Achigan-Dako, Analysis of genetic diversity and agronomic variation in banana sub-populations for genomic selection under drought stress in Southern Benin, *Gene Gene* (2023), doi: https://doi.org/10.1016/j.gene.2023.147210

Analysis of genetic diversity and agronomic variation in banana sub-populations for genomic selection under drought stress in Southern Benin

Luther Fort Mbo Nkoulou[1,2,6*], Lauriane Archange Tchinda Ninla[3], David Cros[4,5], Guillaume Martin[4,5], Zenabou Ndiang[3], Jordan Houegban[1], Hermine Bille Ngalle[2], Joseph Martin Bell[2], Enoch G. Achigan-Dako[1*]

[1] Unit of Genetics, Biotechnology, and Seed Science (GBioS), Laboratory of Phytotechnics, Physiology, Genetics and Plant Breeding (PAGEV), University of Abomey-Calavi, Abomey-Calavi, School of Plant Sciences, Cotonou, Republic of Benin

[2] Unit of genetics and plant Breeding (UGAP), Department of Plant Biology, Faculty of Science, University of Yaoundé 1, Po. Box 812, Yaoundé, Cameroon

[3] Department of Plant Biology and Physiology, Faculty of Science, University of Douala, Po. Box 24157, Douala, Cameroon

[4] Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Unité Mixte de Recherche (UMR), Amélioration Génétique et Adaptation des Plantes méditerranéennes et tropicales (AGAP) Institut, F-34398 Montpellier, France

[5] Unité Mixte de Recherche (UMR), Amélioration Génétique et Adaptation des Plantes méditerranéennes et tropicales (AGAP) Institut, Univ. Montpellier, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), Institut Agro, F-34398 Montpellier, France

[6] Institute of Agricultural Research for Development, Mbalmayo Agricultural Research Centre (CRAM) Mbalmayo, Cameroon

*** Correspondence:**

Corresponding Author:
1. Enoch G. Achigan-Dako (e.adako@gmail.com) 01 P.O. BOX 526 Cotonou, Republic of Benin, Tel. (+229) 95 39 3283;
2. Luther Fort Mbo Nkoulou (joachimnkoulou7.jn@gmail.com), 01 P.O. BOX Cotonou, Republic of Benin, Tel. (+229) 90 50 63 92

Analysis of genetic diversity and agronomic variation in banana sub-populations for genomic selection under drought stress in Southern Benin

Luther Fort Mbo Nkoulou[1,2,6*], Lauriane Archange Tchinda Ninla[3], David Cros[4,5], Guillaume Martin[4,5], Zenabou Ndiang[3], Jordan Houegban[1], Hermine Bille Ngalle[2], Joseph Martin Bell[2], Enoch G. Achigan-Dako[1*]

[1] Unit of Genetics, Biotechnology, and Seed Science (GBioS), Laboratory of Phytotechnics, Physiology, Genetics and Plant Breeding (PAGEV), University of Abomey-Calavi, Abomey-Calavi, School of Plant Sciences, Cotonou, Republic of Benin

[2] Unit of genetics and plant Breeding (UGAP), Department of Plant Biology, Faculty of Science, University of Yaoundé 1, Po. Box 812, Yaoundé, Cameroon

[3] Department of Plant Biology and Physiology, Faculty of Science, University of Douala, Po. Box 24157, Douala, Cameroon

[4] Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Unité Mixte de Recherche (UMR), Amélioration Génétique et Adaptation des Plantes méditerranéennes et tropicales (AGAP) Institut, F-34398 Montpellier, France

[5] Unité Mixte de Recherche (UMR), Amélioration Génétique et Adaptation des Plantes méditerranéennes et tropicales (AGAP) Institut, Univ. Montpellier, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), Institut Agro, F-34398 Montpellier, France

[6] Institute of Agricultural Research for Development, Mbalmayo Agricultural Research Centre (CRAM) Mbalmayo, Cameroon

*** Correspondence:**

Corresponding Author:
1. Enoch G. Achigan-Dako (e.adako@gmail.com) 01 P.O. BOX 526 Cotonou, Republic of Benin, Tel. (+229) 95 39 3283;
2. Luther Fort Mbo Nkoulou (joachimnkoulou7.jn@gmail.com), 01 P.O. BOX Cotonou, Republic of Benin, Tel. (+229) 90 50 63 92

Abstract

In the perspective of investigating genomic selection (GS) among *Musa* genotypes in West and Central Africa, banana accessions were phenotyped under natural drought stress in Benin and genotyped using genotyping by sequencing. Sixty-one (61) accessions grouped into three major genomic groups AAA, AAB and ABB and those without genomic affiliation information were used. Variation within the population was determined by phenotypic variables while population structure and clustering analysis were carried out to understand the genetic diversity at the molecular level. Among the genomic groups evaluated, the group AAB showed the best performance for fruit weight at maturity, (3.41 ± 1.99 kg) and for plant height (198.46 ± 12.66 cm). At the accession level, HD 117 S1 and NIA 27 showed the best plant height (263.16 ± 20.98 cm) and the best fruit weight at maturity (9.43 ± 0.0 kg) respectively. Phenotypic data did not reveal clear genetic diversity among accessions; however, the genetic diversity was conspicuous at the molecular level using 5000 markers. The affiliations of local accessions in genomic groups were determined for the first time based on the phenotypic and molecular data obtained in this study. The knowledge generated allows the possibility to apply GS in banana.

Keywords: Banana accessions, Drought, Single nucleotide polymorphism, Genetic diversity, Genomic selection.

1. Introduction

Bananas are giant herbs belonging to the Musaceae family and *Musa* genus, which is widespread throughout the inter-tropical zones (Charrier et al., 1997; Péréfarres et al., 2007). In Benin, banana and plantain are among the most produced, consumed, and traded commodities (Chabi et al., 2018). Also used as a staple food, bananas are grown in more than 135 countries (Drenth and Kema, 2021). In addition, bananas are rich in nutrients such as vitamins, trace elements, and carotenes which make its consumption popular throughout the world (Amorim et al., 2009). Water deficit (WD) in the soil is the abiotic threat that most limits the growth and yield of the plant (Ekanayake et al., 1994; Kissel et al., 2015; van Asten et al., 2011). The need to control threats such as drought that drastically reduce agricultural production has led research to investigate new approaches in plant breeding. The majority of small farms depends on rain-fed agriculture and therefore are strongly impacted by water shortage (Kissel et al., 2015). Drought is a major constraint to banana production, and its adverse effect is increasing with climate change. Drought susceptibility was reported as the primary constraint to dessert banana and plantain production in Benin (Fanou et al., 2018).

To face the drought constraint in banana production, the solution adopted in plant breeding is the creation of tolerant varieties through breeding cycles that can exceed 15 years. The selection can be done directly by screening drought tolerance among accessions or created varieties (Kissel et al., 2015; Mboula, 2014). Two main species, *Musa acuminata* (A genome) and *Musa balbisiana* (B genome), are involved in cultivated banana. The hybridization within and between these two species resulted in a variable composition of the A and B genomes in cultivars and a conventional classification based on this global proportion (e.g. AA, AAA, AB, AAB and ABB) was proposed (Baurens et al., 2019; Cenci et al., 2021; Simmonds and Shepherd, 1955). In facts, the majority of banana cultivars containing *M. balbisiana* (B genome) are more drought tolerant than those solely based on *M. acuminata* (A genome) (Ravi et al., 2013). In conventional banana breeding programs, this genetic advantage is exploited to incorporate better characteristics into elite germplasm. The study of genetic diversity within the *Musa* genus initially considered phenotypic traits and subsequently molecular markers with the advent of modern tools in breeding programs. Both at the morphological and molecular level, these studies have always reported that there is a great genetic diversity among the species of the genus *Musa* (Jarret and Litz, 1986; Nsabimana and Van

Staden, 2007; Ude et al., 2002). This diversity represents an asset in the creation of varieties resistant to abiotic and biotic stresses. Originally, domestication of the *Musa* species was carried out by farmers who selected plant for parthenocarpic, vigour and high-yielding potential. These traits are mostly found in the triploid cultivars which are the most cultivated. The genomic affiliation that explains the number of copies of the parental genomes *Musa balbisiana* and *Musa acuminata* in the genome of the individuals contributes to the diversity observed. The AAA, AAB and ABB groups are the widely cultivated triploids. Thus, genetic diversity between genomic groups is greater than within a genomic group (Karamura and Mgenzi, 2004). Agronomic traits commonly measured to assess drought performance of banana cultivars are pseudo-stem size and fruit yield. If the conventional research has made it possible to obtain cultivated varieties, it nevertheless remains confronted with challenges such as the cost of phenotyping, the complexity of the polyploid banana genome, the parthenocarpy, male and / or female sterility, limited genetic variability. Overcoming these challenges has motivated the exploration of new approaches in banana breeding against drought. Nansamba et al. (2020) reported that conventional breeding should combine more molecular and biotechnological tools to facilitate the creation of drought-tolerant banana varieties through the conventional route. Since its evocation by Meuwissen et al. (2001), genomic selection (GS) represents in breeding a promising molecular tool to circumvent the limits of conventional selection. In practice, the accuracy with which the phenotypes of candidates are predicted is influenced by factors such as the genotypic and phenotypic diversity within the training population (TP), the size of the TP and the quality and quantity of markers used (Heslot et al., 2013; Ozimati et al., 2019). Taking these factors into account gives a better estimate of the potential of GS (Beil et al., 2017). In the framework of a germplasm population, the knowledge of diversity is a preliminary step in the formation of the TP.

This study gathers agronomic and genetic information that will allow of using GS as a tool for the improvement of triploid bananas for drought in West and Central Africa. Some of these bananas are found in Benin in collections or in the fields where they are cultivated by farmers. The understanding of the variation for drought-related traits among accessions, the genetic diversity within the population, and genomic affiliation of accessions are information that will contribute to the construction of a GS sub-population from the Benin population. To this end, the following questions will guide breeders in this perspective: what is the genetic organization of the banana population used in Benin that can contribute to the construction of a GS training population? Can

drought-related traits and Single Nucleotide Polymorphism (SNP) markers help to understand this genetic organization? What are the genomic affiliations of local accessions and their influence on genetic diversity within the population?

The objective of this study is to quantify the genetic and phenotypic diversity in banana germplasm, in order to evaluate the feasibility of constructing a part of phenotypic database for drought-related traits that can be used for GS. The variability among growth and yield-related traits in banana GS sub-population as well as the genetic diversity is assessed. Based on the structure of the population and the use groups allocated at harvest, the affiliation of the accessions in the three main genomic groups is proposed.

2. Materials and methods

2.1. Plant materials

The banana population consisted of 61 banana accessions, of which 36 accessions were collected in the collection of the Unit of Genetics, Biotechnology, and Seed Science (GBioS) and 25 accessions were obtained from the International Transit Center (ITC) in Belgium. The local accessions were propagated following the method of planting from stem fragments (PIF) (Tomekpe et al., 2011), while the ITC accessions were obtained by *in vitro* culture in Belgium and acclimated at GBioS. The ploidy level and genomic groups of the local accessions were unknown, whereas the ITC accessions were mostly triploids with genomic groups AAA, AAB and ABB. The list of accessions and their characteristics are presented in **Table 1**.

2.2. Field trial and experimentation

2.2.1. Field location and management

The trial was carried out in the experimental site of the GBioS unit. This site is located in the locality of Zinvié, Department of Atlantic, and Commune of Abomey Calavi in Rep. of Benin. Geographic coordinates are 6° 37’ 0” North, 2° 21’ 0” East. Over the year, the average temperature is 28°C and the rainfall ranges from 1,000 to 1,300 mm (**Table 2**). Banana plants were planted on March 2020 for local accessions and on April 2020 for genotypes obtained from ITC, using an alpha lattice design with two replications. Each genotype was represented by 6 plants, 3 per replication with spacing of 3m between rows and 3m between columns. The trial was conducted

using the technical itinerary described by GBioS. In order to maintain the plants under drought conditions, one third of the water requirements as reported by Doto et al. (2013) were supplied to each plant during the dry seasons. To prevent other abiotic and biotic stresses from reducing the performance of the accessions for the measured traits, the field was rigorously monitored and maintained throughout the trial based on the technical itinerary for banana production in Benin developed by the GbioS unit.

Growth-related data were collected during two seasons at the flowering time and yield-related data were collected at the fruit maturity. The mobile application *Fieldbook* was used to record data in the field (Rife and Poland, 2014). Several growth and yield-related traits suggested by Nansamba et al., (2021) and Ravi et al. (2013) as indicators that could be used to assess bananas for drought stress were measured during the trial. In total, five yield parameters and six growth-related traits were measured during the experiment. The description of the measured parameters and the collection procedure are reported in **Table 3**.

2.2.2. Soil moisture content

The average moisture content of the soil was determined using the oven dry method (Mkhabela and Bullock, 2012) at the Integrated Soil and Crop Management Research Unit, Faculty of Agronomic Sciences, University of Abomey-Calavi. The average soil moisture content estimated over the three sample was 10.75. Three soil samples were taken at depths of 0-20 cm, 20-40 cm and 40-60 cm respectively. The moisture content in each different depth is reported in **Table 4**. The weight of moist soil in the pot was measured using an electronic balance. The pots with containing samples were oven dried at 105°C for 24 hours. Each of the 3 samples was replicated 4 times. The weight of the dried samples was measured for each replicate, which allowed the soil moisture content to be determined on the average of the three samples according to the following formula:

$$\text{Soil moisture (\%)} = \frac{\text{weight of moist soil in pot} - \text{weight dry soil in pot}}{\text{weight of moist soil in pot} - \text{weight of pot}} * 100$$

2.3. Phenotypic diversity analysis

2.3.1 The analysis of variance

The data collected were analyzed using the R software (Version 4.1.2). The means on each individual were estimated by taking the average values obtained in the two replications. The analysis of variance (ANOVA) as well as the comparison of the adjusted means were calculated with the function *PBIB.test* of agricolae package (De Mendiburu, 2014). The estimation method was the restricted maximum likelihood (REML) while the least significant difference (lsd) test was used to compare treatments at probability level of 5%. The normality of the variables was checked by the *shapiro.test* function and graphically by the *qqplot* and *ggdensity* functions. If the ANOVA of the transformed and the original data gave similar conclusions, the original data was maintained. Only the NFF was transformed by the square root. For the ANOVA, two independent variables namely group and accessions were used.

2.3.2. Correlation and path analyses

If the correlations allow to quantify the magnitude and direction of the components that influence the expression of a main trait, the direct and indirect contribution of these components on this trait can be determined through path analysis (Baye et al., 2020). Thus, the interactions between traits were determined by correlation tests and path analysis using the library variability (Popat et al., 2020). The path analysis was done considering fruit weight at maturity as the dependent variable, because the yield is the major trait of interest regarding the drought effect in banana (Ravi et al., 2013). In addition, the package psych (Revelle, 2017) was used to visualize the correlation matrix from correlation analysis in order to identify traits (phenotypic information) that can be used to study the diversity within the population.

2.3.3. Principal components analysis

The principal components analysis (PCA) was performed on the phenotypic data and the projection of the graphs were done using packages FactoMineR (Lê et al., 2008) and factoextra (Kassambara et al., 2017). The projection of the dependent variables as well as the accessions was applied on the first two components, based on the fact that they represent the greatest observed variability.

2.4. Genetic diversity analysis

2.4.1. Genotyping and Variant calling

Banana leaf samples after sterilization with 70% alcohol were collected from the nursery and stored in genotyping kits. Genotyping took place at Bioscience for eastern and central Africa-International Livestock Research Institute (BecA-ILRI), SEQART AFRICA service in Nairobi, Kenya. Genotyping by sequencing using DartSeq™ technology was applied as reported by Elshire et al. (2011), the main steps being the DNA extraction, DNA quality check, enzyme restriction with Pstl and Msel, Polymerase Chain Reaction (PCR), pooling, purification, quantification and Illumina sequencing. The quality, concentration and purity of the samples were checked by running 1µl of each sample in 0.8% agarose gel, after incubation for 1h in a digestion buffer.

In order to perform variant calling that respects the polyploid nature of the genotyped accessions, the fastq data obtained from the genotyping was analyzed with the VcfHunter toolbox (Garsmeur et al., 2018). The variant calling was performed on the *M. acuminata* reference genome V4 (Belser et al., 2021). VcfHunter tools processed the data in three steps which are variant calling to generate the VCF (process_reseq tool), prefiltering of the VCF (VcfPrefilter tool) and final filtering (vcfFilter tool). The prefiltering was done to identify and remove SNPs position resulting from sequencing error. Further filtering was done with vcfR package (Knaus and Grünwald, 2017) to select SNPs on 11 chromosomes and remove SNPs on chloroplast and mitochondrial genome, and to select those with a sequencing depth greater than 13. Only bi-allelic sites with at most 20% missing data were retained. The missing data in the final SNPs file were imputed with the *missForest* function of R, using the Random Forest method (Stekhoven, 2015). The expected heterozygosity and the frequency of the minor allele (MAF) in the SNPs obtained were determined using the *Hs* and *minorAllele* functions of the adegenet library respectively (Jombart, 2008). The coverage of the markers on the chromosomes was explored and the visualization of this coverage was done by making a physical map using the function *lmv.linkage.plot* of the package LinkageMapView (Ouellette et al., 2018).

2.4.2. Structure and clustering analyses

The genetic diversity of banana GS sub-population was assessed by the structure and hierarchical clustering analysis. To study the structure of the genotyped banana population, a random subset of 5000 SNPs markers was generated from the obtained SNPs after variant calling. This subset was used in the STRUCTURE software version 2.3.4 (Porras-Hurtado et al., 2013). The initial number of populations was fixed at K = 10, each genotype was burned 10000 times using Monte Chain

replicates (MCMC) sampling procedure for the inference of accessions in sub-populations (van Ravenzwaaij et al., 2018). Moreover, for each value of K, the analysis was done on 10 iterations. The results of the population structure were harvested online in the Structure Harvester platform and the optimal number of populations was defined based on the best value of Delta K (ΔK) following the Evanno method (Evanno et al., 2005). The Genetic distances by Nei distance method (Nei, 1972) and genetic variation in the original population were calculated respectively using the *Gdist* and *Fst* functions of the NAM package (Xavier et al., 2015). A clustering analysis was performed with the *hclust* function following the ward.D method using stats library. The number of optimal clusters was determined with the *find.clusters* function of the adegenet library (Jombart, 2008). The number of axes to be retained in the PCA was 200 and the maximum number of clusters was 20.

3. Results

3.1. Phenotypic diversity

3.1.1. Trait variation

Overall, a significant difference was observed between accessions for all measured variables (**S1** Table). At the level of genomic groups, a significant difference was noted only for fruit circumference (FC), leaf senescence ratio (RLS) and the number of functional leaves (NFL) (**S1** Table). Observations of the fruits at harvest allowed us to classify the accessions into three main cultivar types, namely dessert bananas, cooking bananas and plantains. All accessions of genomic group AAA produced dessert banana fruits, AAB produced plantain fruits while ABB produced cooking banana fruits. Under water stress conditions, pseudo-stem size or plant height (PH) and yield are the most evaluated agronomic traits in general. Then, accessions HD 117 S1 had the best plant size (263.16 ± 20.98 cm) and NIA 27 the best fruit weight at maturity (FWM) (9.43 ± 0 kg). Likewise, the genomic group AAB (Plantain cultivars) showed the highest performance for fruit weight at maturity (3.41 ± 1.99 kg) and for plant height (198.46 ± 12.66 cm). The variations of parameters among genomic group are visualized in **Figure 1**, while the summary table of the average performance of the accessions is shown in the supplement material (**S2** Table). Because of the presence of multicollinearity, only the analysis of variance with one factor was performed on the variables with genomic group and accessions. A low coefficient of variation (CV) was

observed among accessions for all variables measured, but a greatest CV was noted for FWM (34.47), number of suckers at flowering (NSF) (25.41) and leaf surface (LS) (25.36). Likewise, all variables showed a CV greater than 20% among genomic groups except the number of functional leaves. FWM (68.15%) showed the greatest CV and NFL (19.74%) recorded the lowest one.

3.1.2. Correlation and path analyses

The study of interactions between variables showed strong phenotypic correlations within and between growth and yield-related traits (**Figure 2**). Within growth-related traits, RLS, which expresses the number of dead leaves over the number of living leaves, was negatively correlated with all other traits. The highest positive correlation was observed between the plant circumference (PC) and PH (0.79) while the lowest one was observed between NSF and LS (0.33). No negative correlation was observed within yield-related traits, the highest positive correlation was observed between the number of hands per fruit (NHF) and the number of fingers per fruit (NFF) (0.77) while the lowest one was observed between NHF and fruit circumference (0.13). Regarding both groups of traits (yield and growth) together, PC and NFF (0.35) showed the highest correlation, and NSF and the fruit length (FL) (0.02) yielded the lowest correlation. NSF was also negatively correlated with all yield-related traits.

The understanding of the influence of variable on fruit weight through a path analysis showed direct positive effects of PC, RLS, NHF, FC and FL on FWM at the genotypic level, and the direct positive effects of PH, LS, RLS, NHF, FL and FC on FWM at the phenotypic level **(S3 Table)**. Direct negative effects of PH, NFL, LS, NSF and NFF were observed on FWM at the genotypic level, and PC, NFL, RLS and NFF on FWM at the phenotypic level. PC (1.57) and NFF (-0.83) showed the largest direct positive and negative effects at the genotypic level respectively, while FL (0.55) and PC (0.19) showed the largest direct positive and negative effects at the phenotypic level respectively. The phenotypic and genotypic residual effects were 0.58 and 0.61 respectively.

3.1.3. Principal components analysis

The principal components analysis performed on the entire dataset showed that the first two dimensions explained 56.5% of the variation, the first dimension explaining 36.2% and the second dimension 20.3% (**Figure 3**). Considering the two main axes together, the most discriminated variables were the plant circumference and the fruit length while NIA 27 and HD 72 were the most

discriminated accessions. At the individual level, the NIA 27 accession contributed significantly to the first dimension and the local accessions HD 45B and HD 72 contributed most to the second dimension. At the level of variables, agronomic traits including the plant circumference and the plant height contributed mostly to dimension 1 while yield-related traits including the number of hands per fruit, the number of hands per fruit and Fruit weight at maturity contributed mostly to dimension 2. The PCA also showed a random distribution of accessions on both axes, independent of their genomic groups and origin. Overall, the projection of individuals on the two axes did not reveal any structuring of the population.

3.2. Genetic diversity

3.2.1. Variant calling

Genotyping by sequencing performed on the population produced after the first phase of VcfHunter 81,672 bi-allelic SNPs with a percentage of missing data of 40.27%. Filtering to exclude off-chromosome variants produced 73,187 SNPs with 22.13% missing data, while filtering to reduce redundant SNPs produced 36,812 bi-allelic SNPs with 22.38% missing data. The highest number of markers was observed on chromosome 4 (4671 SNPs) while chromosome 2 had the lowest coverage (2461 SNPs). The statistics on the distribution of SNPs on the chromosomes are reported in **Table 5**. The physical map of the SNPs on banana chromosomes is visualized in **Figure 4**. In this proportion of 5000 used to conduct the genetic diversity analysis of the population, the estimated expected heterozygosity was 0.129 while the average observed heterozygosity was 0.115. The minor allele frequency ranged from 0.0095 to 0.33, for an average of 0.24.

3.2.2. Population structure

In the analysis of genetic diversity using SNPs markers, only the 35 genotyped accessions were used first. In this analysis the influence of genomic groups and cultivar types on genetic diversity was observed in order to infer later the non-genotyped accessions into genomic groups and genetic sub-populations. The analysis of the population structure revealed two sub-populations based on the best value of K determined by the Evanno method (**Figure 5**). The inference of individuals in the different sub-populations was influenced by their genomic affiliation. The sub-population 1 was constituted mainly by individuals belonging to the ABB group with two individuals belonging

to the AAB and AAA groups. Sub-population 2 was represented by individuals belonging to genomic groups AAB and AAA with the presence of one individual from group ABB. In addition, the few local accessions were found in both sub-populations. The inference ancestry of genotypes and the result of Evanno test to determine the optimal number of populations are reported in **Table 6** and **S4 Table** respectively. In these sub-populations, 62.3% (20 individuals) of the population was inferred to population 1 and 37.7% (15 individuals) to population 2. The allele-frequency divergence among populations was 0.056 while the expected heterozygosity between individuals of the same sub-population was 0.10 in population 1 and 0.11 in population 2. The mean genetic variation (Fst) was 0.10 between sub-population 1 and sub-population 2.

3.2.3. Hierarchical clustering in the population

The discriminant analysis of principal components (DAPC) revealed 4 clusters. Cluster 1 had the highest number of genotypes (20) while the other three clusters had 5 genotypes each (**Figure 6**). The hierarchical clustering was greatly influenced by the genomic constitution of the genotypes. Cluster 1 consisted of individuals from the ABB genomic group, cluster 2 had individuals from the AAA genomic group, and clusters 3 and 4 consisted mostly of accessions from the genomic group AAB. Within the main population, the genetic distance among individuals calculated by the Nei distance method varied between 0.0061 and 0.179 with an average of 0.074 and the expected heterozygosity was 0.129. Estimation of genetic differentiation between pairs of clusters via pairwise Fst analysis showed values ranged between 0.26 and 0.48. Cluster 3 and 4 had the lowest Fst value (0.26) while cluster 1 and 2 had the highest Fst value (0.48).

4. Discussion

4.1. Phenotypic diversity in banana GS sub-population

4.1.1. Traits variation and accessions behavior

In Banana, size and yield are the most important agronomic traits evaluated to identify genotypes performances under drought conditions (Nansamba et al., 2021; Ravi et al., 2013). Ravi et al. (2013) established the behavior of banana genotypes to drought based on many studies. Cavendish (AAA) accessions are reported to be susceptible; this can lead to the complete crop failure in “Grande Naine” (highly susceptible). In AAB, susceptible cultivars such as “Pome” may form

fruits but are unable to fill them completely, while moderate tolerant and tolerant varieties normally produce fruits. The majority of ABB are reported to be drought tolerant (Ravi et al., 2013). In some moderate tolerant ("Pisang Awak" ABB), there is a reduction in the number of hands and fruit length while others maintain yield stability. Of the accessions evaluated, only three individuals from group AAA produced the fruit. Group ABB had the best fruit weight at maturity (3.41 kg) while AAA (1.9 kg) yielded the lowest one. These results are different from those obtained by Uwimana et al. (2020) studying the effect of seasonal drought. They obtained a fruit weight at maturity of 10.87 kg with the ABB group ("Cachaco" accession) and 20.84 kg with the AAA group ("Nakitengwa" accession). This difference may result from genetic variation within groups in terms of drought tolerance, and from the fact that the two studies used different and limited numbers of accessions per group. Consideration of physiological and biochemical indicators may help to provide a comprehensive understanding of the biological changes that differentiate susceptible and resistant accessions to water stress.

The effect of genomic group remains however important as seems to confirm the analysis of variance, which showed significant differences in fruit weight at maturity between genomic groups. To quantify the variability within a plant population, the coefficient of variation is the most commonly used parameter. The larger CV is, the more dispersed the individuals are within the population (Kotzamanides et al., 2009). In the present study the majority of the variables had a low CV, however the leaf surface, the number of suckers at flowering and the fruit weight at maturity had a CV > 25% explaining a possible variation of the population for these parameters. These coefficients of variation are mostly higher than those obtained by Sirisena and Senanayake (2000) and Ortiz and Vuylsteke (1998), who obtained a CV of 20% for the fruit weight at maturity and 10% for PH for the first authors, 27.7% for the fruit weight at maturity and 6% for the plant height for the second authors. According to Ortiz and Vuylsteke (1998), low CVs may result from measurements taken with minimum error. Moreover, in the case of Ortiz and Vuylsteke (1998), the space between the plants of the same plot was relatively smaller (2m) than the one we used. Many works have reported that when the density increases, the CV also increases and vice versa (Daynard and Muldoon, 1983; Hamblin et al., 1978; Kotzamanides et al., 2009).

4.1.2. Phenotypic interaction among traits

A positive correlation was observed between all traits evaluated except the senescence rate which was negatively correlated with almost all others traits. This result is similar to the result of Pinar et al. (2021) who observed positive correlation for the stem height, number of hands, number of fruits, bunch weight, fruit weight and fruit length. However, the negative correlation observed for senescence rate with the rest of the trait explains the behavior of the plant under water stress conditions. According to Bananuka et al. (1999), sensitivity in banana is manifested by changes in growth through the reduction of leaf size and increase in leaf senescence resulting in limited photosynthetic activity. Path analysis is an important factor in improving desirable traits in plants. It informs about the direct and indirect role of traits on the expression of the dependent variable, making it possible to link the causes and effects between the traits (Baye et al., 2020). The highest positive indirect effects were observed for the plant height through the plant circumference and for the leaf surface through the plant height at the phenotypic level, next for the plant height through the plant circumference and for the leaf surface through the plant circumference at the genotypic level. This suggests that these traits could be used with other traits that had a direct positive effect in improving banana for bunch weight. Residual effects in path analysis are an indicator of the role of the traits involved in the analysis on the variability of the dependent trait (Baye et al., 2020). The phenotypic and genotypic residual effects were 0.58 and 0.61, respectively, indicating that the traits involved in the path analysis explained 42% and 39% of the variability in bunch weight at the phenotypic and genotypic levels respectively. The plant height, the number of functional leaves, the leaf surface, the number of suckers at flowering and the number of fingers per fruit had a negative direct effect at the genotypic level while the plant circumference, the number of functional leaves, the ratio of leaf senescence and the number of fingers per fruit had negative direct effect at the phenotypic level on fruit weight at maturity. However, all of these traits except the ratio of leaf senescence are positively correlated with bunch weight, suggesting that indirect effects are responsible for the positive correlation observed. Among trait with negative direct effect, some had a positive indirect effect. The negative direct effects were compensated by the positive indirect effects.

4.1.3. Principal components analysis

The principal components analysis is used to explain the differences observed in a dataset and thus to understand the possible relationships between the variables associated with that dataset (Ramli

et al., 2010). In the present study, the principal components analysis was not able to clearly discern the individuals in the population into groups. The projection of accessions in the two main axes did not group the accessions according to their genomic affiliations or origin, but revealed a wide dispersion of accessions, indicating that the traits contributing to these axes are not strongly influenced by genomic affiliations. This result was also obtained by Ekanayake et al. (1995) and by Nyine et al. (2017). In contrast, Osuji et al. (1997) obtained better grouping of triploid bananas based on inflorescence and growth-related traits. According to Nyine et al. (2017), the phenomenon explaining this difference could be due to differences in the ability to access the carbon source. The random distribution of accessions in the PCA can also be explained by the proportion of the parental genomes *M. acuminata* (A) and *M balbisiana* (B) carried by the genomes of the accessions. Indeed, recent studies have pointed out that within the accessions of the groups, the proportion of B genomes can vary among individuals belonging to the group (Baurens et al., 2019; Cenci et al., 2021). Likewise, accessions of group AAB can have a high proportion of B genome compared to individuals of the ABB group. This makes it difficult to group accessions according to their genomic affiliation as we expected in the PCA.

4.2. Genotypic variation in the banana GS sub-population

4.2.1. Genotyping and SNP distribution

Knowledge of the genetic diversity in a species of food importance is crucial to identify the different loci that are implicated in resistance to abiotic and biotic threats and in yield (Alemu et al., 2020). Several works using molecular tools have reported the high genetic diversity in *Musa* spp (Jarret and Litz, 1986; Karamura and Mgenzi, 2004; Nsabimana and Van Staden, 2007; Ude et al., 2002). In the context of genomic selection, knowledge of genetic diversity via population structure allows for better construction of the training population and thus increases prediction accuracy. In the present study the 5000 markers used to establish the genetic diversity of the population appeared to be sufficiently informative. Variant calling with the VcfHunter platform yielded 81, 672 pre-filtered and 36, 812 bi-allelic SNPs after multiple filtering. This marker size constitutes an important molecular dataset for performing genomic study in the population. The MAF was greater than 0.05, indicating that the markers used were credible to perform a genetic diversity study. Indeed, the marker set with a MAF less than 0.05 is generally disqualifying for a genetic diversity study (Luo et al., 2019). Heterozygosity is one of the quantitative measures of

the degree of polymorphism of a marker (Shete et al., 2000). The expected heterozygosity obtained in the original population was 0.129 and the average observed heterozygosity was 0.115. The low observed heterozygosity compared to the expected one explains the probable mixture of 2 populations. In some cases, the occurrence of error when sampling a population may also influence the ratio between both heterozygosities. This result indicates that the population used in the study is not in Hardy-Weinberg equilibrium.

4.2.2. Population genetic background

The hierarchical clustering analysis revealed 4 clusters within the population, explaining a significant diversity. The distribution of clusters was strongly influenced by the genomic affiliation of individuals, which shows that these accessions have followed a different genetic evolution, manifested by the copy number of *M. accuminata* and *M. balbisiana* present in their genome. Indeed, accessions with the same genomic group share the same parental genomic information that is read by the markers, and thus these accessions are found in the same clusters. Understanding population structure is an important step in considering potential association analyses as this structure impacts the prediction accuracies (Luo et al., 2019; Varshney et al., 2017). Varshney et al. (2017) reported a reduction in selection accuracy in the structured population. Luo et al. (2019) reported that a training population based on simple stratified sampling was better in a structured rice population, showing that the optimization of the training population depends on the type of population present. In this study, two populations were identified using Evanno method to identify the best K value. The distribution of individuals within the population was strongly influenced by their genomic group, confirming the result obtained in the hierarchical clustering analysis. However, the appearance of certain genomic groups in the two populations can be explained by the exchange of genes that occurred during the evolution of the genus *Musa*. Nyine et al. (2018) found that some half-sib families were close to individuals with whom they shared parents. The population structure studied here is consistent with the expectations, as the germplasm is made up of genotypes from different origins that are therefore genetically distant. When working with germplasm structuring is to be expected as a consequence of the great genetic diversity of *Musa* spp. which has been proven in several studies (Karamura and Mgenzi, 2004; Li et al., 2013; Nsabimana and Van Staden, 2007; Simmonds and Weatherup, 1990). Genetic variation within a population can be revealed by estimating Fst between sub-populations. According to Luo et al.,

(2019), an Fst greater than 0.15 is considered significant for discriminating individuals within a population indicating the presence of sub-populations. In the genetic diversity analysis, the average genetic variation (Fst) was 0.10 between sub-population 1 and sub-population 2, showing a little genetic differentiation in populations. According to Eltaher et al. (2018), large gene pool exchanges can lead to low genetic differentiation between populations. However, the SNPs used here allowed to understand the population structure through cluster and structure analyses, which led to the construction of a good genomic selection population based on these analyses. The results of the population differentiation revealed two sub-populations in the structure analysis and 4 clusters in the DAPC. However, all the individuals belonging to the same cluster in DAPC were also grouped within the same sub-population. In STRUCTURE, k = 10 was the number of possibilities of choosing the best k value but the k = 2 was the optimum value revealed in the analysis instead of 4 like showed DAPC. This difference explains that, either there was not define population or STRUCTURE misrepresented the true number of k as reported by Tehseen et al., (2021).

4.2.3. Inference of genomic groups to local accession

Based on the results obtained in the clustering and structure analysis, it is possible that some accessions that we could not genotype could be inferred in one of the populations and consequently in the three main genomic groups. Studies based on structure analyses or clustering analyses allowed inferring accessions of plants for which there was insufficient information in their origin, population and other different classification groups (Şakiroğlu et al., 2010; Uba et al., 2021). In addition, it is reported that in Benin the majority of cultivars grown in the south of the country are either dessert bananas or plantains (Chabi et al., 2018). In the world food and trade, cultivars generally called plantains are triploids and belong to the AAB genomic group, those considered as bananas (or dessert bananas) almost all belong to the AAA genomic group, while other cultivars called cooking bananas belong to the ABB group (Florent et al., 2015; Happi Emaga et al., 2007; Strosse et al., 2006). The observations of the fruit type (**S5 Table**) made at harvest and the results of structure analysis allowed the classification of local accessions into different genomic groups and sub-populations (**Table 1**; **S6 Table**).

4.2.4. *Drought tolerance improvement in banana*

This study was conducted to lay the foundation for genetic improvement of banana plants through the use of GS. As population type can greatly influence the efficacy of GS (Varshney et al. 2017; Luo et al. 2019), the results on the genetic diversity of the population assessed in this study will allow for better organization of TP during GS application. Knowledge of the genetic diversity of the East African banana population has allowed the application of GS in this population (Nyine et al., 2018, 2017). Moreover, the phenotypic performance of accessions evaluated under natural water stress conditions provides a good phenotypic database that can be used for genetic improvement of banana plants for drought. Mbo Nkoulou et al. (2022) reported that the effectiveness of GS in the improvement of banana plants for drought requires a good evaluation of the TP under water stress. The present study produced growth and yield indicators for 61 banana accessions that will then be used to predict the genetic values of each accession and select the best ones using GS. The phenotypic and molecular information allowed to find the genomic affiliations of local accessions. This represents an important advance to involve these accessions in GS. Indeed, it was reported that triploid genomic groups (AAB, AAA and ABB) could influence the GS efficiency of triploid banana plants for drought and black sigatoka disease (Mbo Nkoulou et al., 2022). Therefore, considering the improvement of banana plants for these two stresses via GS implies the knowledge of the genomic affiliations of the accessions we wish to select.

5. Conclusion

The present study allowed understanding the phenotypic and genetic variations among banana accessions in a water deficit environment. ANOVA revealed significant differences among accessions for the traits evaluated, and the genomic group AAB and ABB showed the best plant height and fruit weight performance. The positive correlation observed between growth and yield variables allows to understand that it is possible to select banana accessions with good growth and high yield. The path analysis showed the important role of indirect effects of traits measured on fruit weight at maturity. However, the 5000 SNP markers selected after genotyping proved to be informative enough to highlight the genetic diversity of the genotyped accessions in the population. The information from this genetic diversity analysis and phenotyping allowed the allocation of genomic groups for the accessions where they were unknown. Therefore, the phenotypic and molecular information obtained in this study opens a pathway towards the genetic improvement of polyploid bananas for drought using genomic selection tool.

Acknowledgment

This work was realized with the support of MESO@LR-Platform at the University of Montpellier.

Formatting of funding sources

We gratefully acknowledge the European Union and the African Union for supporting our study through the GENES Intra-Africa Academic Mobility project.

Declaration of Competing Interest

The authors have no declared conflict of interest

Author contributions

Conceptualization: Luther Fort Mbo Nkoulou, Enoch G. Achigan-Dako. **Data curation**: Luther Fort Mbo Nkoulou, Tchinda Ninla Lauriane Archange, Jordan Houegban. **Formal analysis**: Luther Fort Mbo Nkoulou, Tchinda Ninla Lauriane, David Cros, Guillaume Martin. **Funding acquisition** : Luther Fort Mbo Nkoulou, Enoch G. Achigan-Dako, Hermine Bille Ngalle. **Investigation** : Luther Fort Mbo Nkoulou, G. Achigan-Dako. **Methodology**: Luther Fort Mbo Nkoulou, Enoch G. Achigan-Dako. **Project administration**: Enoch G. Achigan-Dako. **Supervision**: Enoch G. Achigan-Dako, David Cros, Hermine Bille Ngalle, Ndiang Zenabou, Joseph Bell. **Validation**: Enoch G. Achigan-Dako, David Cros. **Writing – original draft**: Luther Fort Mbo Nkoulou, Tchinda Ninla Lauriane Archange. **Writing – review & editing**: Enoch G. Achigan-Dako. Luther Fort Mbo Nkoulou, David Cros, Tchinda Ninla Lauriane Archange, Ndiang Zenabou, Hermine Bille Ngalle.

Conflicts of interest

The authors have no declared conflict of interest

Supplement material

S1 Table. ANOVA for the effect of genotypes and genomic groups on the variation of variables measured (DOCX).

S2 Table. Phenotypic performances of banana GS sub-population and standard deviation for growth and yield-related traits (XLSX)

S3 Table. Direct and indirect effects of variables on fruit weight at the genotypic (A) and phenotypic (B) levels. Direct effects are shown in bold (DOCX).

S4 Table. Inference of the optimal K number in banana genomic population using SNPs with K statistics (DOCX).

S5 Illustration. Phenotypic observation of some banana accession fruits (PPTX).

S6 Table. Inference of non-genotyped accessions in genomic groups and sub-population (DOCX).

References

Alemu, A., Feyissa, T., Letta, T., Abeyo, B., 2020. Genetic diversity and population structure analysis based on the high density SNP markers in Ethiopian durum wheat (*Triticum turgidum ssp. durum*). BMC genetics 21, 1–12. https://doi.org/10.1186/s12863-020-0825-x.

Amorim, E.P., Vilarinhos, A.D., Cohen, K.O., Amorim, V.B., Santos-Serejo, J.A. dos, Pestana, K.N., Santos, V.J. dos, Paes, N.S., Monte, D.C., Reis, R.V. dos, others, 2009. Genetic diversity of carotenoid-rich bananas evaluated by Diversity Arrays Technology (DArT). Genetics and Molecular Biology 32, 96–103.

Bananuka, J.A., Rubaihayo, P.R., Tenywa, M.M., 1999. Reactions of *Musa* genotypes to drought stress. Afr. Crop Sci. J 7, 333–339.

Baurens, F.-C., Martin, G., Hervouet, C., Salmon, F., Yohomé, D., Ricci, S., Rouard, M., Habas, R., Lemainque, A., Yahiaoui, N., 2019. Recombination and large structural variations shape interspecific edible bananas genomes. Molecular biology and evolution 36, 97–111. https://doi.org/10.1093/molbev/msy199.

Baye, A., Berihun, B., Bantayehu, M., Derebe, B., 2020. Genotypic and phenotypic correlation and path coefficient analysis for yield and yield-related traits in advanced bread wheat (*Triticum aestivum* L.) lines. Cogent Food & Agriculture 6, 1752603. https://doi.org/10.1080/23311932.2020.1752603.

Beil, C.T., Manmathan, H.K., Anderson, V.A., Morgounov, A., Haley, S.D., 2017. Population Structure and Genetic Diversity Analysis of Germplasm from the Winter Wheat Eastern European Regional Yield Trial (WWEERYT). Crop Science 57, 812–820. https://doi.org/10.2135/cropsci2016.08.0639.

Belser, C., Baurens, F.-C., Noel, B., Martin, G., Cruaud, C., Istace, B., Yahiaoui, N., Labadie, K., Hřibová, E., Doležel, J., 2021. Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. Communications biology 4, 1–12. https://doi.org/10.1038/s42003-021-02559-3.

Brisson, N., Ozier-Lafontaine, H., Dorel, M., 1998. Effects of soil management and water regime on banana growth between planting and flowering. simulation using the stics model. Acta Hortic. 229–238. https://doi.org/10.17660/ActaHortic.1998.490.23.

Cenci, A., Sardos, J., Hueber, Y., Martin, G., Breton, C., Roux, N., Swennen, R., Carpentier, S.C., Rouard, M., 2021. Unravelling the complex story of intergenomic recombination in ABB allotriploid bananas. Annals of Botany 127, 7–20. https://doi.org/10.1093/aob/mcaa032.

Chabi, M.C., Dassou, A.G., Dossou-Aminon, I., Ogouchoro, D., Aman, B.O., Dansi, A., 2018. Banana and plantain production systems in Benin: ethnobotanical investigation, varietal diversity, pests, and implications for better production. Journal of Ethnobiology and Ethnomedicine 14, 1–18. https://doi.org/10.1186/s13002-018-0280-1.

Charrier, A., Jacquot, M., Hamon, S., Nicolas, D., 1997. L'amélioration des plantes tropicales, ORSTOM. ed. CIRAD, Paris, 632.

Daynard, T.B., Muldoon, J.F., 1983. Plant-to-plant variability of maize plants grown at different densities. Canadian Journal of Plant Science 63, 45–59. https://doi.org/10.4141/cjps83-005.

De Mendiburu, F., 2014. Agricolae: statistical procedures for agricultural research. R package version 1, 1–4.

Doto, C., Ahamidé, B., Agbossou, K., 2013. Impact de l'irrigation goutte à goutte sur les données phénologiques et le rendement du bananier plantain (*Musa paradisiaca*) planté sur un sol ferralitique au Sud-Bénin. Bulletin de la Recherche Agronomique du Bénin 36–47.

Drenth, A., Kema, G., 2021. The vulnerability of bananas to globally emerging disease threats. Phytopathology PHYTO-07. https://doi.org/10.1094/PHYTO-07-20-0311-RVW.

Ekanayake, I., Ortiz, R., Vuylsteke, D., 1995. Physiological factors in drought tolerance of various *Musa* genotypes. International lnstilute of Tropical Agriculture (IITA) RESEARCH 7–10.

Ekanayake, I.J., Ortiz, R., Vuylsteke, D.R., 1994. Influence of Leaf Age, Soil Moisture, VPD and Time of Day on Leaf Conductance of Various *Musa* Genotypes in a Humid Forest-Moist

Savanna Transition Site. Annals of Botany 74, 173–178. https://doi.org/10.1006/anbo.1994.1106.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. PLoS One 6, e19379. https://doi.org/10.1371/journal.pone.0019379.

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 F3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing. Frontiers in genetics 9, 76. https://doi.org/10.3389/fgene.2018.00076.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Fanou, M., Ewedje, E.-E.B., Adeoti, K., Djedatin, G.L., Affokpon, A., Farid, B.-M., Toukourou, F., 2018. Diversity of local varieties of banana and plantain cultivated in Benin. International Journal of Biodiversity and Conservation 10, 497–509. https://doi.org/10.5897/IJBC2018.1232.

Florent, A., Loh, A.M., Thomas, H., 2015. Nutritive value of three varieties of banana and plantain blossoms from Cameroon. Journal of Agricultural Sciences 5, 52–61. https://doi.org/10.15580/GJAS.2015.2.012115009.

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., Jenkins, J., Martin, G., Charron, C., Hervouet, C., Costet, L., Yahiaoui, N., Healey, A., Sims, D., Cherukuri, Y., Sreedasyam, A., Kilian, A., Chan, A., Van Sluys, M.-A., Swaminathan, K., Town, C., Bergès, H., Simmons, B., Glaszmann, J.C., van der Vossen, E., Henry, R., Schmutz, J., D'Hont, A., 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nat Commun 9, 2638. https://doi.org/10.1038/s41467-018-05051-5.

Hamblin, J., Knight, R., Atkinson, M.J., 1978. The influence of systematic micro-environmental variation on individual plant yield within selection plots. Euphytica 27, 497–503. https://doi.org/10.1007/BF00043175.

Happi Emaga, T., Andrianaivo, R.H., Wathelet, B., Tchango, J.T., Paquot, M., 2007. Effects of the stage of maturation and varieties on the chemical composition of banana and plantain peels. Food Chemistry 103, 590–600. https://doi.org/10.1016/j.foodchem.2006.09.006.

Heslot, N., Rutkoski, J., Poland, J., Jannink, J.-L., Sorrells, M.E., 2013. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PloS one 8, e74612. https://doi.org/10.1371/journal.pone.0074612.

Jarret, R., Litz, R., 1986. Isozymes as genetic markers in bananas and plantains. Euphytica 35, 539–549. https://doi.org/doi.org/10.1007/BF00021863.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129.

Karamura, D., Mgenzi, B., 2004. On farm conservation of *Musa* diversity in the great lakes region of East Africa. African Crop Science Journal 12, 75–83. https://doi.org/10.4314/acsj.v12i1.27665.

Kassambara, A., Mundt, F., others, 2017. Factoextra: extract and visualize the results of multivariate data analyses. R package version 1, 337–354.

Kissel, E., van Asten, P., Swennen, R., Lorenzen, J., Carpentier, S.C., 2015. Transpiration efficiency versus growth: Exploring the banana biodiversity for drought tolerance. Scientia Horticulturae 185, 175–182. https://doi.org/10.1016/j.scienta.2015.01.035.

Knaus, B.J., Grünwald, N.J., 2017. vcfr: a package to manipulate and visualize variant call format data in R. Molecular ecology resources 17, 44–53. https://doi.org/10.1111/1755-0998.12549.

Kotzamanides, S.T., Lithourgidis, A.S., Roupakias, D.G., 2009. Plant density effect on the individual plant to plant yield variability expressed as coefficient of variation in barley. Spanish journal of agricultural research 607–610.

Lê, S., Josse, J., Husson, F., 2008. FactoMineR: A Package for Multivariate Analysis. Journal of Statistical Software 25, 1–18. https://doi.org/10.18637/jss.v025.i01.

Li, L.-F., Wang, H.-Y., Zhang, C., Wang, X.-F., Shi, F.-X., Chen, W.-N., Ge, X.-J., 2013. Origins and domestication of cultivated banana inferred from chloroplast and nuclear genes. PLoS One 8. https://doi.org/doi.org/10.1371/journal.pone.0080502.

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.

Mbo Nkoulou, L.F., Ngalle, H.B., Cros, D., Adje, C.O., Fassinou, N.V., Bell, J., Achigan-Dako, E.G., 2022. Perspective for genomic-enabled prediction against black sigatoka disease and

drought stress in polyploid species. Front. Plant Sci. 13. https://doi.org/10.3389/fpls.2022.953133.

Mboula, S., 2014. Ecophysiology of dwarf plantain hybrids in peri-urban areas of Cameroon (PhD Thesis). UCL-Université Catholique de Louvain, Belgium.

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157, 1819–1829. https://doi.org/10.1093/genetics/157.4.1819.

Mkhabela, M.S., Bullock, P.R., 2012. Performance of the FAO AquaCrop model for wheat grain yield and soil moisture simulation in Western Canada. Agricultural Water Management 110, 16–24. https://doi.org/10.1016/j.agwat.2012.03.009.

Nansamba, M., Sibiya, J., Tumuhimbise, R., Karamura, D., Kubiriba, J., Karamura, E., 2020. Breeding banana (*Musa spp.*) for drought tolerance: A review. Plant Breeding 139, 685–696. https://doi.org/10.1111/pbr.12812.

Nansamba, M., Sibiya, J., Tumuhimbise, R., Ocimati, W., Kikulwe, E., Karamura, D., Karamura, E., 2021. Assessing Drought Effects on Banana Production and On-Farm Coping Strategies by Farmers–A Study in the Cattle Corridor of Uganda. Research square 31. https://doi.org/10.21203/rs.3.rs-867469/v1.

Nei, M., 1972. Genetic distance between populations. The American Naturalist 106, 283–292. https://doi.org/10.1086/282771.

Nsabimana, A., Van Staden, J., 2007. Assessment of genetic diversity of Highland bananas from the National Banana Germplasm Collection at Rubona, Rwanda using RAPD markers. Scientia Horticulturae 113, 293–299. https://doi.org/10.1016/j.scienta.2007.04.010.

Nyine, M., Uwimana, B., Blavet, N., Hřibová, E., Vanrespaille, H., Batte, M., Akech, V., Brown, A., Lorenzen, J., Swennen, R., Doležel, J., 2018. Genomic Prediction in a Multiploid Crop: Genotype by Environment Interaction and Allele Dosage Effects on Predictive Ability in Banana. Plant Genome 11, 170090. https://doi.org/10.3835/plantgenome2017.10.0090.

Nyine, M., Uwimana, B., Swennen, R., Batte, M., Brown, A., Christelová, P., Hřibová, E., Lorenzen, J., Doležel, J., 2017. Trait variation and genetic diversity in a banana genomic selection training population. PLoS One 12, e0178734. https://doi.org/doi.org/10.1371/journal.pone.0178734.

Ortiz, R., Vuylsteke, D., 1998. Quantitative variation and phenotypic correlations in banana and plantain. Scientia Horticulturae 72, 239–253. https://doi.org/10.1016/S0304-4238(97)00137-4.

Osuji, J.O., Okoli, B.E., Vuylsteke, D., Ortiz, R., 1997. Multivariate pattern of quantitative trait variation in triploid banana and plantain cultivars. Scientia Horticulturae 71, 197–202. https://doi.org/10.1016/S0304-4238(97)00101-5.

Ouellette, L.A., Reid, R.W., Jr, S.G.B., Brouwer, C.R., 2018. LinkageMapView - Rendering High Resolution Linkage and QTL Maps. Bioinformatics 34, 306–307. https://doi.org/10.1093/bioinformatics/btx576.

Ozimati, A., Kawuki, R., Esuma, W., Kayondo, S.I., Pariyo, A., Wolfe, M., Jannink, J.-L., 2019. Genetic variation and trait correlations in an East African cassava breeding population for genomic selection. Crop science 59, 460–473. https://doi.org/10.2135/cropsci2018.01.0060.

Péréfarres, F., Acina, I.N., Teycheney, P.-Y., 2007. Prévalence et diversité du virus de la mosaïque en tirets du bananier (BSV) et du virus X du bananier (BVX) en Guadeloupe, in : Rencontres de Virologie Végétale. Presented at the 11 èmes Rencontres de virologie végétale, Aussois, 28 janvier au 1er février 2007, CIRAD, Aussois (France), p. 1.

Popat, R., Patel, R., Parmar, D., 2020. variability: Genetic Variability Analysis for Plant Breeding Research. R Package Version 01, https://cran.r-project.org/web/packages/variability/variability.pdf 7.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., Lareu, M.V., 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. Frontiers in genetics 4, 98. https://doi.org/10.3389/fgene.2013.00098.

Ramli, S., Ismail, N., Alkarkhi, A.F.M., Easa, A.M., 2010. The Use of Principal Components and Cluster Analysis to Differentiate Banana Peel Flours Based on Their Starch and Dietary Fibre Components. Trop Life Sci Res 21, 91–100.

Ravi, I., Uma, S., Vaganan, M., Mustaffa, M., 2013. Phenotyping bananas for drought resistance. Frontiers in Physiology 4. https://doi.org/doi.org/10.3389/fphys.2013.00009.

Revelle, W.R., 2017. psych: Procedures for personality and psychological research 97.

Rife, T.W., Poland, J.A., 2014. Field book: an open-source application for field data collection on android. Crop Science 54, 1624–1627. https://doi.org/10.2135/cropsci2013.08.0579.

Şakiroğlu, M., Doyle, J.J., Charles Brummer, E., 2010. Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa* L.) accessions using SSR markers. Theoretical and applied genetics 121, 403–415. https://doi.org/10.1007/s00122-010-1319-4.

Shete, S., Tiwari, H., Elston, R.C., 2000. On estimating the heterozygosity and polymorphism information content value. Theoretical population biology 57, 265–271. https://doi.org/10.1006/tpbi.2000.1452.

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

Simmonds, N.W., Weatherup, S.T.C., 1990. Numerical taxonomy of the wild bananas (*Musa*). New Phytologist 115, 567–571. https://doi.org/10.1111/j.1469-8137.1990.tb00485.x.

Sirisena, J.A., Senanayake, S.G.J.N., 2000. Estimation of variability parameters within 'Mysore' banana clones and their implication for crop improvement. Scientia Horticulturae 84, 49–66. https://doi.org/10.1016/S0304-4238(99)00094-1.

Stekhoven, D.J., 2015. missForest: Nonparametric missing value imputation using random forest. R package version 1.4, Astrophysics Source Code Library 1, ascl-1505.

Strosse, H., Schoofs, H., Panis, B., Andre, E., Reyniers, K., Swennen, R., 2006. Development of embryogenic cell suspensions from shoot meristematic tissue in bananas and plantains (*Musa spp.*). Plant Science 170, 104–112. https://doi.org/10.1016/j.plantsci.2005.08.007.

Tehseen, M.M., Istipliler, D., Kehel, Z., Sansaloni, C.P., da Silva Lopes, M., Kurtulus, E., Muazzam, S., Nazari, K., 2021. Genetic Diversity and Population Structure Analysis of *Triticum aestivum* L. Landrace Panel from Afghanistan. Genes 12, 340. https://doi.org/10.3390/genes12030340.

Tomekpe, K., Kwa, M., Dzomeku, B.M., Ganry, J., 2011. CARBAP and innovation on the plantain banana in Western and Central Africa. International Journal of Agricultural Sustainability 9, 264–273. https://doi.org/10.3763/ijas.2010.0565.

Uba, C.U., Oselebe, H.O., Tesfaye, A.A., Abtew, W.G., 2021. Genetic diversity and population structure analysis of bambara groundnut (*Vigna subterranea* L) landraces using DArT SNP markers. PloS one 16, e0253600. https://doi.org/10.1371/journal.pone.0253600.

Ude, G., Pillay, M., Nwakanma, D., Tenkouano, A., 2002. Genetic Diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP Markers. Theor Appl Genet 104, 1246–1252. https://doi.org/10.1007/s00122-002-0914-4.

Uwimana, B., Zorrilla-Fontanesi, Y., van Wesemael, J., Mduma, H., Brown, A., Carpentier, S., Swennen, R., 2020. Effect of seasonal drought on the agronomic performance of four banana genotypes (*Musa spp*.) in the East African Highlands. Agronomy 11, 4. https://doi.org/10.3390/agronomy11010004.

van Asten, P.J.A., Fermont, A.M., Taulya, G., 2011. Drought is a major yield loss factor for rainfed East African highland banana. Agricultural Water Management 98, 541–552. https://doi.org/10.1016/j.agwat.2010.10.005.

van Ravenzwaaij, D., Cassey, P., Brown, S.D., 2018. A simple introduction to Markov Chain Monte–Carlo sampling. Psychon Bull Rev 25, 143–154. https://doi.org/10.3758/s13423-016-1015-8.

Varshney, R.K., Roorkiwal, M., Sorrells, M.E., 2017. Genomic selection for crop improvement: new molecular breeding strategies for crop improvement. Springer, Cornell University Ithaca, NY, USA.

Xavier, A., Xu, S., Muir, W., Rainey, K., 2015. NAM: association studies in multiple populations. Bioinformatics 31, 3862–3864.

List of figures

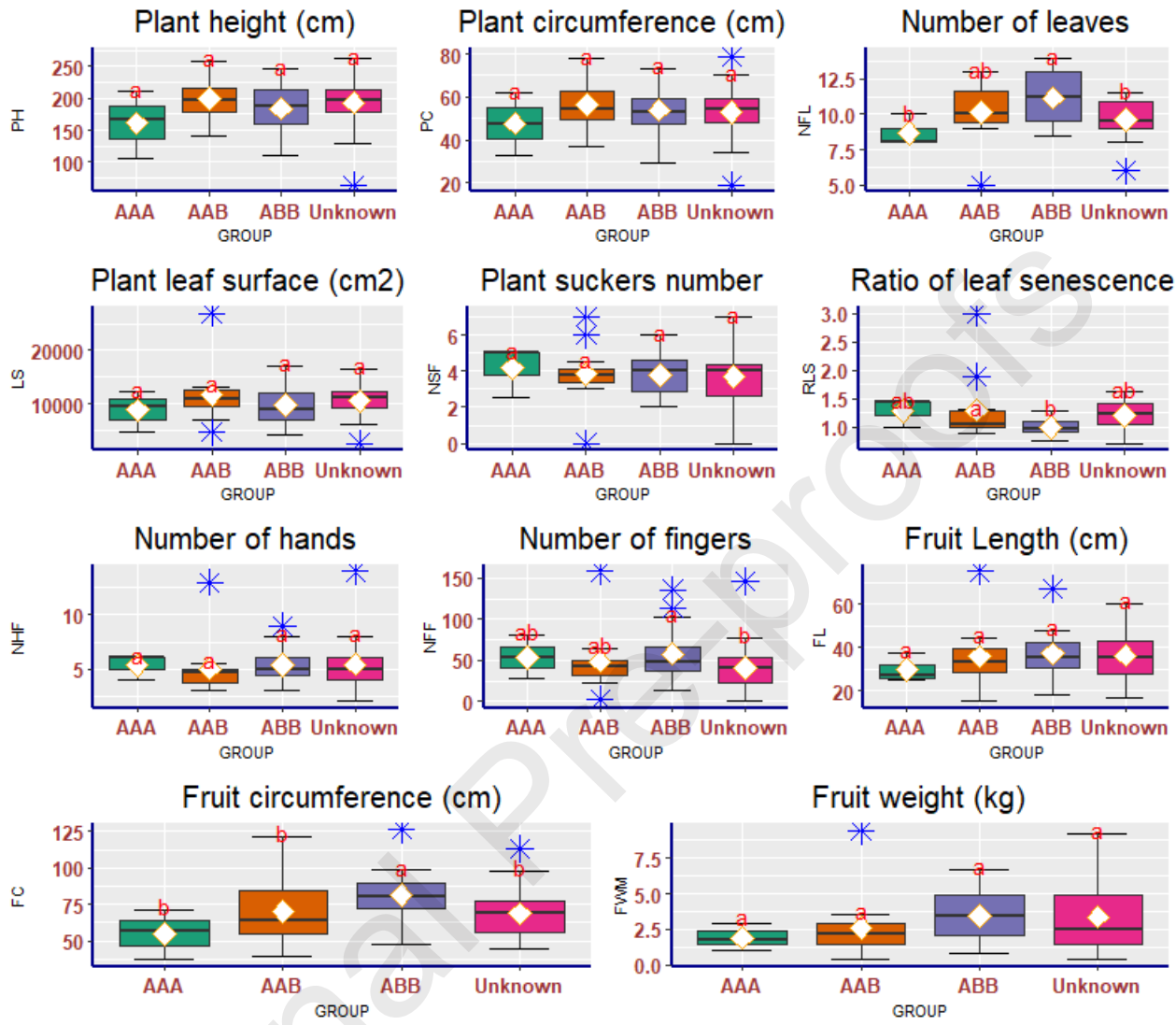

Fig. 1. Variation of growth and yield-related traits among triploid genomic groups. The white squares indicate the average values and the blue asterisks represent the outliers. AAA = Dessert banana cultivars, AAB = Plantain cultivars, ABB = Cooking banana cultivars, Unknown = Cultivars with unknown genomic group but including all possible cultivar types

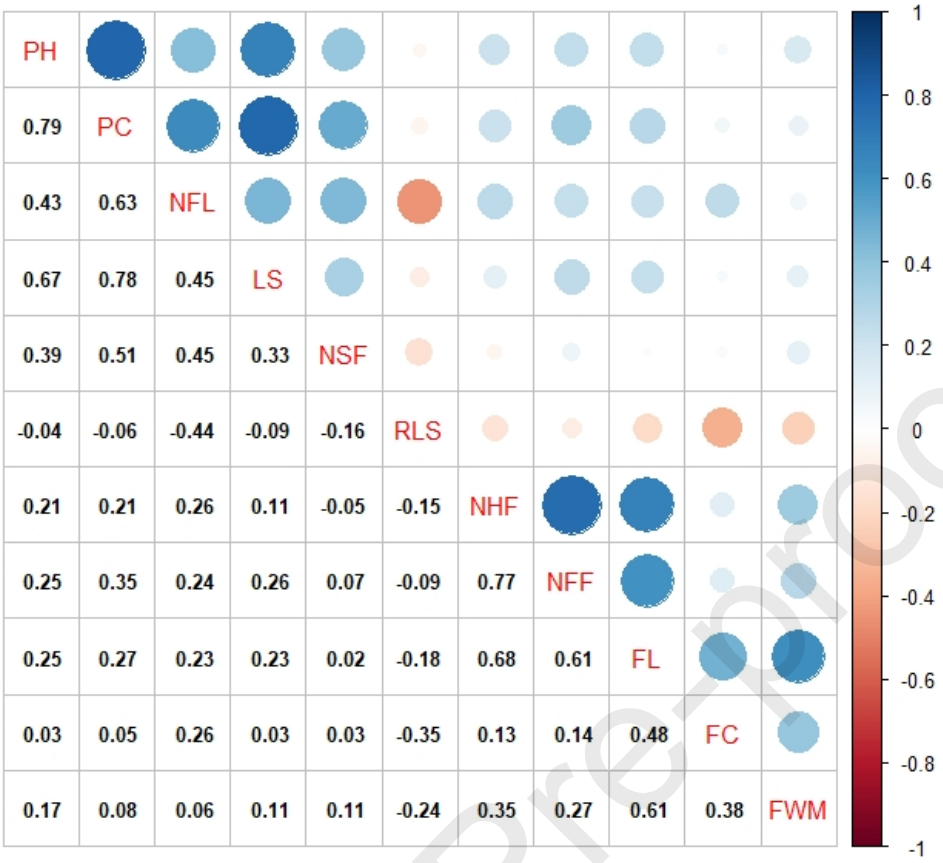

Fig. 2. Phenotypic correlation among traits evaluated under drought stress conditions. The blue color indicates positive correlation while the red color indicates negative correlation. The darker the blue the stronger the positive correlation. The darker the red, the stronger the negative correlation. PH = Plant Height, PC = Pseudostem Circumference, NFL = Number of functional leaves, LS = Leaf surface, RLS= Rate of leaf senescence, NHF = number of hands per fruit, NFF = number of fingers per fruit, FL = Fruit length, FC = Fruit circumference, FWM = Fruit weight at maturity, NSF = Number of suckers at Flowering

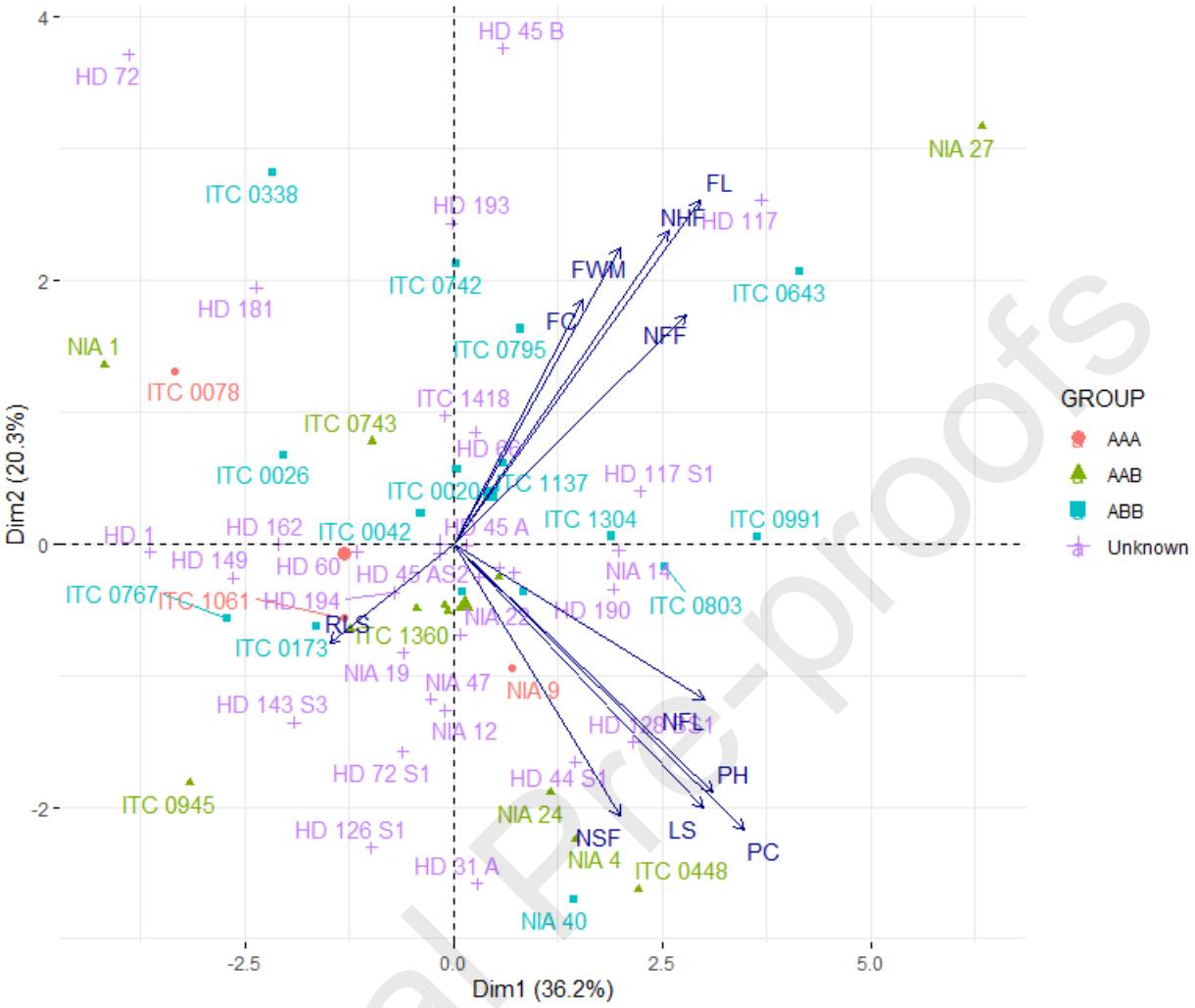

Fig. 3. Principal components analysis projection on the first two axes for accessions (dots) and traits (arrows). Dim1 = Dimension 1, Dim2 = Dimension 2, PH = Plant Height, PC = Pseudostem Circumference, NFL = Number of functional leaves LS = Leaf surface, RLS= Rate of leaf senescence, NHF = number of hands per fruit, NFF = number of fingers per fruit, FL = Fruit length, FC = Fruit circumference, FWM = Fruit weight at maturity, NSF = Number of suckers at Flowering

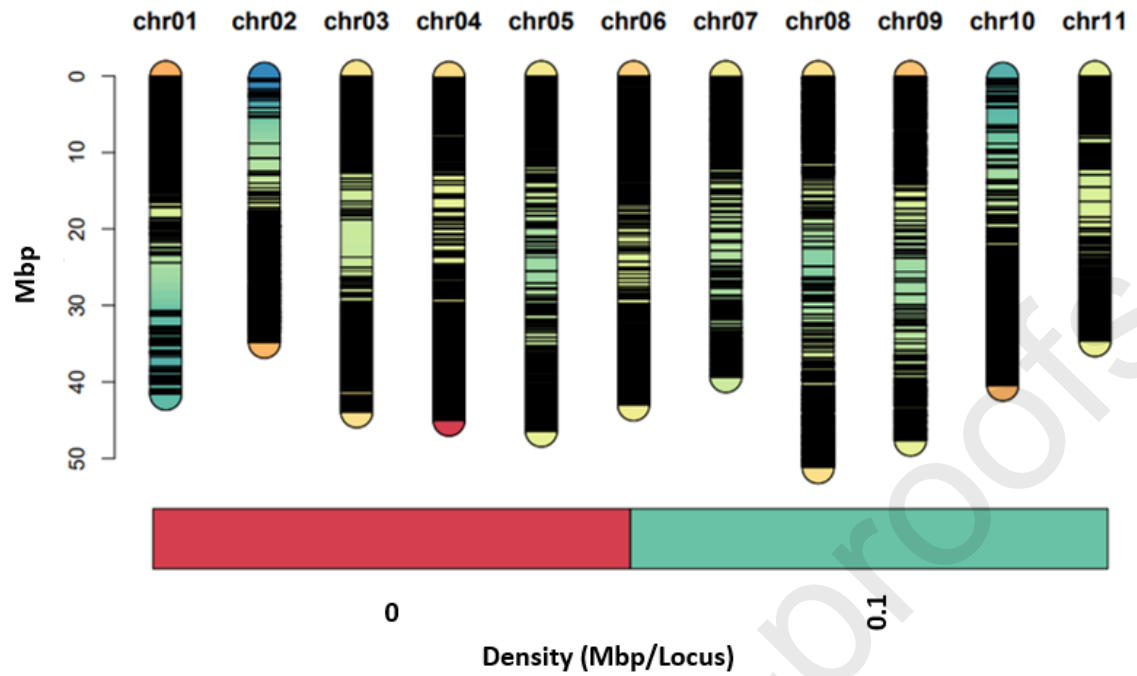

Fig. 4. Physical map of SNPs markers on the 11 banana chromosomes. Markers positions were converted in Mbp by dividing the initial position by 1,000,000. A black bar represents SNP markers, SNP density increase from red to green color.

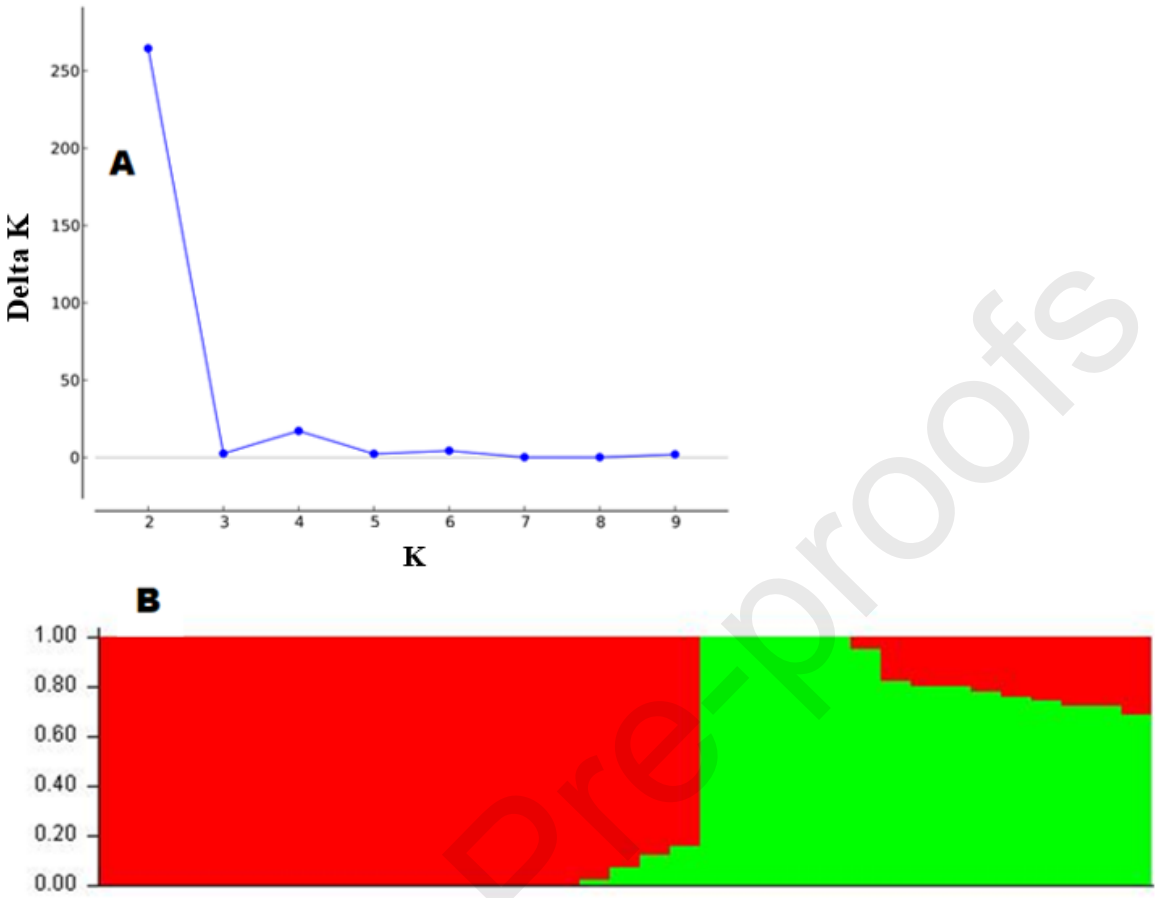

Fig. 5. Population structure using 5000 SNPs in STRUCTURE software. **A**: delta K for different number of sub-population or genetic group. **B**: Plot for different sub-populations at K = 2, each color representing one sub-population. The red color represents the sub-population 1 and the green color represents the sub-population 2.

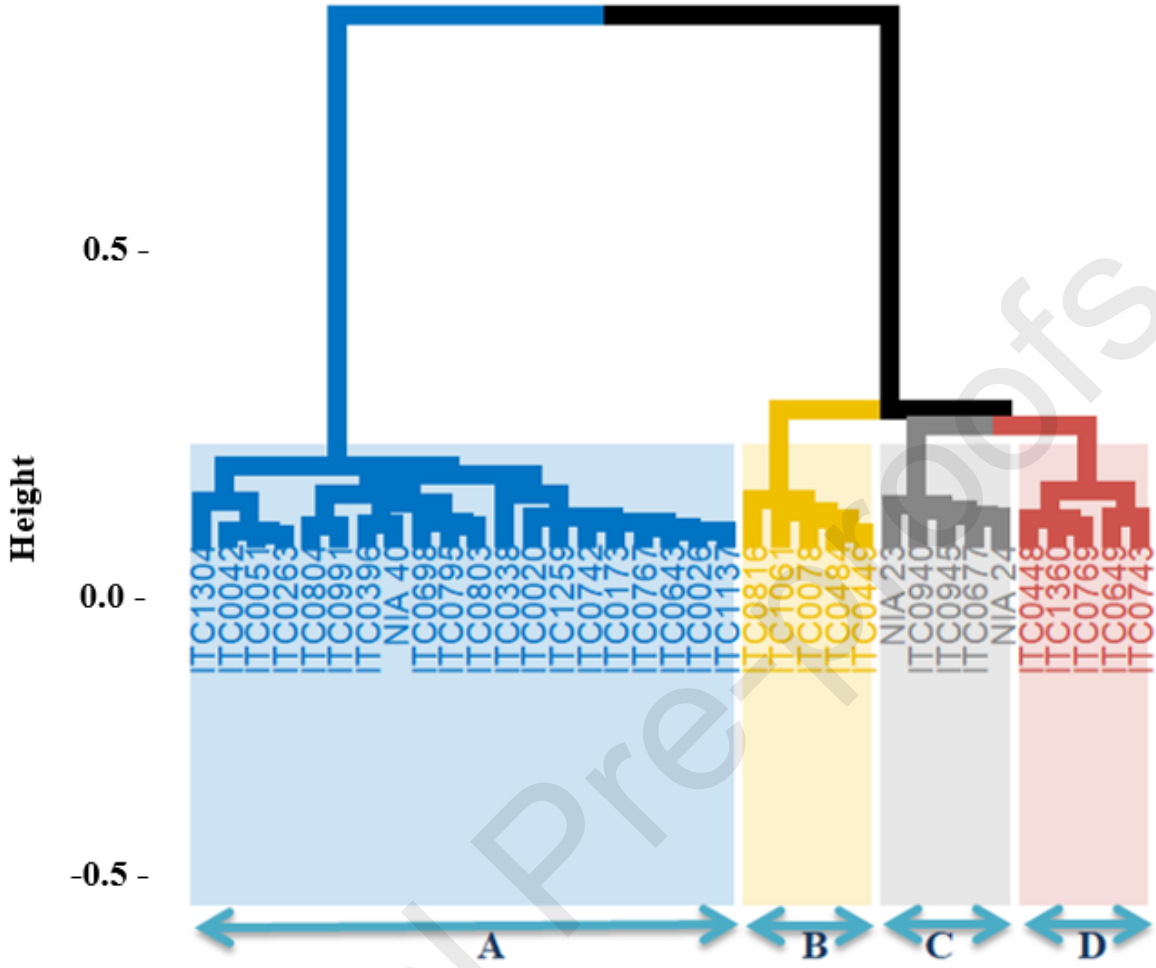

Fig. 6. Hierarchical phylogenic dendrogram of banana accessions visualized with factoextra library using SNPs markers. **A** = Cluster 1; **B** = Cluster 2; **C** = Cluster 3; **D** = Cluster 4

List of tables

Table 1.

List of banana accession used for the study. Use groups were allocated based on the fruit type observed after the phenotyping.

Accession name	Origin	Use group	Genomic group status	Inferred Genomic group
HD 1	Local	Dessert banana	unknown	AAA

HD 117	Local	Cooking banana	unknown	ABB
HD 117 S1	Local	Cooking banana	unknown	ABB
HD 126 S1	Local	Dessert banana	unknown	AAA
HD 128 BS1	Local	Cooking banana	unknown	ABB
HD 13	Local	Dessert banana	unknown	AAA
HD 143 S3	Local	Cooking banana	unknown	ABB
HD 149	Local	Cooking banana	unknown	ABB
HD 162	Local	Dessert banana	unknown	AAA
HD 181	Local	Dessert banana	unknown	AAA
HD 190	Local	Cooking banana	unknown	ABB
HD 193	Local	Cooking banana	unknown	ABB
HD 194	Local	Dessert banana	unknown	AAA
HD 31 A	Local	Dessert banana	unknown	AAA
HD 39	Local	Plantain	unknown	AAB
HD 44 S1	Local	Cooking banana	unknown	ABB

HD 45 A	Local	Dessert banana	unknown	AAA
HD 45 AS2	Local	Dessert banana	unknown	AAA
HD 45 B	Local	Cooking banana	unknown	ABB
HD 55	Local	Dessert banana	unknown	AAA
HD 60	Local	Dessert banana	unknown	AAA
HD 66	Local	Cooking banana	unknown	ABB
HD 72	Local	Plantain	unknown	AAB
HD 72 S1	Local	Plantain	unknown	AAB
ITC 0020	ITC	Cooking banana	ABB	ABB
ITC 0026	ITC	Cooking banana	ABB	ABB
ITC 0042	ITC	Cooking banana	ABB	ABB
ITC 0078	ITC	Dessert banana	AAA	AAA
ITC 0173	ITC	Cooking banana	ABB	ABB
ITC 0338	ITC	Cooking banana	ABB	ABB
ITC 0396	ITC	Cooking banana	ABB	ABB
ITC 0448	ITC	Plantain	AAB	AAB
ITC 0449	ITC	Plantain	AAB	AAB

ITC 0643	ITC	Cooking banana	ABB	ABB
ITC 0742	ITC	Dessert banana	ABB	ABB
ITC 0743	ITC	Plantain	AAB	AAB
ITC 0767	ITC	Cooking banana	ABB	ABB
ITC 0769	ITC	Plantain	AAB	AAB
ITC 0795	ITC	Cooking banana	ABB	ABB
ITC 0803	ITC	Plantain	ABB	ABB
ITC 0940	ITC	Plantain	AAB	AAB
ITC 0945	ITC	Cooking banana	AAB	AAB
ITC 0991	ITC	Cooking banana	ABB	ABB
ITC 1061	ITC	Dessert banana	AAA	AAA
ITC 1137	ITC	Cooking banana	ABB	ABB
ITC 1259	ITC	Cooking banana	ABB	ABB
ITC 1304	ITC	Cooking banana	ABB	ABB
ITC 1360	ITC	Plantain	AAB	AAB
ITC 1418	ITC	Cooking banana	unknown	ABB
NIA 1	INRAB	Plantain	AAB	AAB

NIA 12	INRAB	Dessert banana	unknown	AAA
NIA 14	INRAB	Cooking banana	unknown	ABB
NIA 19	INRAB	Cooking banana	unknown	ABB
NIA 22	INRAB	Dessert banana	unknown	AAA
NIA 23	INRAB	Plantain	AAB	AAB
NIA 24	INRAB	Plantain	AAB	AAB
NIA 27	INRAB	Plantain	AAAB	AAAB
NIA 4	INRAB	Plantain	AAB	AAB
NIA 40	INRAB	Cooking banana	ABB	ABB
NIA 47	INRAB	Cooking banana	unknown	ABB
NIA 9	INRAB	Dessert banana	AAA	AAA

Table 2.

Average climate conditions in Abomey-Calavi. 6° 37' 0" North, 2° 21' 0" (https://fr.climatedata.org/afrique/benin/atlantique)

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Average temperature (°C)	27.1	27.7	28.1	27.8	27.1	25.7	25	24.6	25.2	26	27	27.3
Average minimum temperature (°C)	24.7	25.6	26.2	26	25.2	24.1	23.5	23.1	23.5	24.2	25.1	25
Maximum temperature (°C)	30.9	31.3	31.3	30.7	29.7	28	27.3	27	27.8	28.9	29.9	30.7
Precipitation (mm)	31	45	74	101	179	226	137	110	152	159	86	42

Humidity (%)	76%	79%	81%	82%	84%	87%	85%	85%	86%	86%	83%	80%
Rainy days	7	8	13	14	19	19	17	15	18	19	16	9
Sunlight Hours (h)	8.4	8.1	8.1	8.1	7.5	6.5	6	4.7	5.7	6.6	7.5	8.3

Table 3.

Description of traits collected during the experimentation at Zinvié. Data were collected from November 2020 to May 2022

Trait	**Codes**	**Description**	**Measurement**
		Growth-related traits	
Plant Height (cm)	(PH)	size of the pseudo stem	From the substrate to the last formed leaf, using a measuring tap
Pseudostem Circumference (cm)	(PC)	Circumference of the pseudo stem at the time of the measurement	At 10 cm above substrate, using a measuring tap
Number of functional leaves	(NFL)	Number of leaves with 50% of the leaf surface still green	Observe and count the number of leaves with 50% of their green area
Leaf surface (cm^2)	(LS)	Surface of functional leaves	Leaf length * Leaf width *0.83 (Brisson et al., 1998)
Rate of leaf senescence	(RLS)	Number of dead leaves on the total number per month	Divide the number of dead leaves over the number of living leaves (respecting the NFL criterion)

Number of suckers at Flowering	(NSF)	Number of shoots per plant at the time of flowering	Count the number of suckers at flowering
		Yield-related traits	
number of hands per fruit	(NHF)	Number of consumable hands per fruit diet	count the number of hands
number of fingers per fruit	(NFF)	Number of consumable fingers per fruit	count the number of fingers
Fruit length (cm)	(FL),	Length of the mature banana bunch	Use measuring tap
Fruit circumference (cm)	(FC)	Circumference of the mature banana diet	Use a measuring tap
Fruit weight (kg)	(FWM)	Weight of the mature diet	Weight the fruit on a scale

Table 4.

Soil moisture parameters of the trial location at Zinvié (2-1 and 2-2: sample at 0-20 cm depth; 3-1 and 3-2: soil sample at 20-40 cm depth; 1-1 and 1-2: Sample at 40-60 cm depth)

Sample code	**Pot (g)**	**Pot + Sample (g)**	**Pot + Sample 105°C (g)**	**Moisture**
1-1	17.9	48.3	44.5	12.50
1-2	18.8	42.3	39.5	11.91
2-1	17.9	58.8	54.3	11.00
2-2	17.6	55.7	51.5	11.02

3-1	18.0	43.3	41.1	8.69
3-2	17.7	44.7	42.4	8.51
1-1	18.2	45.3	41.8	12.91
1-2	16.2	51.5	47.0	12.74
2-1	18.9	59.4	54.8	11.35
2-2	17.9	51.0	47.2	11.48
3-1	17.7	52.4	49.4	8.64
3-2	17.6	43.1	41.0	8.23

Table 5.

Statistics on 36,812 SNPs obtained on 11 chromosomes of banana after filtering

Chromosome names	**Number of SNPs**	**Start position**	**End position**	**Length (Mb)**
chr01	2668	136938	41618066	41.481128
chr02	2461	272735	34775788	34.503053
chr03	3695	46174	43885966	43.839792
chr04	4671	164938	44955327	44.790389
chr05	3173	111883	46372321	46.260438
chr06	3898	62903	43032859	42.969956
chr07	2525	78282	39340982	39.262700
chr08	3931	125715	51165327	51.039612
chr09	3533	109735	47649375	47.539640
chr10	2825	290604	40434605	40.144001
chr11	2666	102772	34581619	34.478847

Table 6.

Accession inference ancestry in different populations after hierarchical clustering and structure analysis

Accession name	Inferred in Population 1	Inferred in Population 2	Population	Cluster	Group
ITC0649	0.177	0.823	2	4	AAB
ITC0078	0.000	1.000	2	2	AAA
ITC1360	0.272	0.728	2	4	AAB
ITC0448	0.312	0.688	2	4	AAB
ITC0484	0.000	1.000	2	2	AAA
ITC0816	0.000	1.000	2	2	AAA
ITC0940	0.253	0.747	2	3	AAB
ITC0945	0.200	0.800	2	3	AAB
ITC0677	0.195	0.805	2	3	ABB
ITC1061	0.000	1.000	2	2	AAA
ITC0743	0.273	0.727	2	4	AAB
ITC0769	0.242	0.758	2	4	AAB
ITC0449	0.000	1.000	2	2	AAB
NIA 24	0.219	0.781	2	3	AAB
NIA 23	0.047	0.953	2	3	AAB
ITC0643	1.000	0.000	1	1	ABB
ITC0042	1.000	0.000	1	1	ABB
ITC0051	0.974	0.026	1	1	ABB
ITC0698	1.000	0.000	1	1	ABB
ITC1259	1.000	0.000	1	1	ABB
ITC0396	1.000	0.000	1	1	ABB

ITC0795	1.000	0.000	1	1	ABB
ITC0803	1.000	0.000	1	1	ABB
ITC0804	0.998	0.002	1	1	ABB
ITC0020	1.000	0.000	1	1	ABB
ITC0026	1.000	0.000	1	1	ABB
ITC0991	0.926	0.074	1	1	ABB
ITC0742	1.000	0.000	1	1	ABB
ITC1137	1.000	0.000	1	1	ABB
ITC0173	1.000	0.000	1	1	ABB
ITC0263	1.000	0.000	1	1	AAA
ITC0767	1.000	0.000	1	1	ABB
ITC1304	0.878	0.122	1	1	AAB
NIA 40	1.000	0.000	1	1	ABB
ITC0338	0.842	0.158	1	1	ABB

Abbreviations List

AA: A genomic group for banana diploid individuals with 2 copies of *Musa aciminata* genome

AAA: A genomic group for banana triploid individuals with 3 copies of *Musa aciminata* genome

AB: A genomic group for banana diploid individuals with 1 copy of *Musa aciminata* genome and 1 copy of *Musa balbisiana* genome

AAB: A genomic group for banana triploid individuals with 2 copies of *Musa aciminata* genome and 1 copy of *Musa balbisiana* genome

ABB: A genomic group for banana triploid individuals with 1 copy of *Musa aciminata* genome and 2 copies of *Musa balbisiana* genome

bp: base pair(s)

Mbp: Million base pairs

Msel: Isoschizomer of RspRS II

PstI: Type II restriction endonuclease

SNP: Single Nucleotide Polymorphism

CRediT author statement

Luther Fort Mbo Nkoulou: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing**. Tchinda Ninla Lauriane Archange**: Data curation, Formal analysis, Writing – original draft, Writing – review & editing**. David Cros**: Formal analysis, Supervision, Validation, Writing – review & editing**. Guillaume Martin**: Formal analysis**. Ndiang Zenabou**: Supervision, Writing – review & editing**. Jordan Houegban**: Data curation**. Hermine Bille Ngalle**: Funding acquisition, Supervision**. Joseph Martin Bell**: Supervision**. Enoch G. Achigan-Dako**: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Validation, Writing – review & editing

Declaration of interests

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Highlights

1. The Benin banana population represents a sub-population for genomic selection
2. Benin banana accessions are affiliated with the triploid genomic groups
3. The VcfHunter platform is useful for the development of SNPs in polyploid bananas
4. Banana collection in Benin is enriched with 35 additional accessions