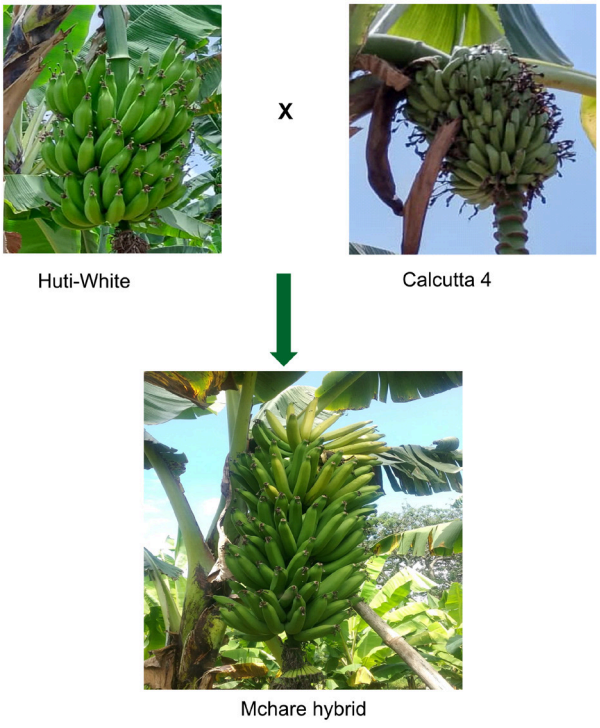
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2024:94
Faculty of Landscape Architecture, Horticulture and Crop Production Science

Genetic improvement of bananas for East Africa

Violet Akech



Genetic improvement of bananas for East Africa

Violet Akech

Faculty of Landscape Agriculture, Horticulture and Crop Production Science

Department of Plant Breeding

Alnarp

SLU

SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS

Alnarp 2024

Acta Universitatis Agriculturae Sueciae
2024:94

Cover: Mchare banana landrace (left hand) that was crossed with wild diploid (right hand) to produce the high-yielding and Fusarium wilt resistant hybrid (T2070-1) (Violet Akech, 2024, Licence under IITA)

ISSN 1652-6880
ISBN (print version) 978-91-8046-421-5
ISBN (electronic version) 978-91-8046-429-1
https://doi.org/10.54612/a.4617i0mre4

Swedish University of Agricultural Sciences, Department of Plant Breeding, Alnarp, Sweden

Print: SLU Grafisk service, Alnarp 2024

Genetic improvement of bananas for East Africa

Abstract

Banana (*Musa* spp.) is an important food and income security crop for millions of people in the global tropical and subtropical South. However, its improvement through crossbreeding is challenging because of the limited number of offspring for evaluation, unsynchronized data, and a long, expensive and tedious breeding process that leads to low rates of genetic gains. This thesis aimed to determine how much genetic diversity exists for Musa breeding and explore the potential of accelerating genetic gain through optimization of the Mchare breeding pipeline, considering three approaches; 1) evaluate the crop's genetic diversity and population structure, 2) use the clonal model to determine the level of genetic variation and its decomposition into additive, dominance and epistasis and 3) integrating genomic information to improve estimates of genetic parameters for key traits in banana trials.

The genetic diversity was low, especially among the diploids and tetraploids. Musa accessions with the potential to be used for parental improvement were identified. The Additive variation was significant ($p \leq 0.001$) and was the strongest genetic variance for all yield-related traits. Therefore, focusing on parental improvement and selection of the best parents is recommended for genetic gains in Mchare banana populations. The highest positive correlations were observed among the yield-related traits. The correlation between the plant crop and the first ratoon crop for all yield-related traits was significant, positive and of high magnitude ($0.62 > r < 0.81$, $p \leq 0.001$). The estimated narrow-sense heritability (h^2) was > 0.50 for all yield-related traits. The hybrid, H-BLUP (best linear unbiased predictor) model had the highest prediction accuracies for yield-related traits, while the pedigree-based, P-BLUP, model had the highest prediction accuracies for agronomic traits. The implications of this research extend to the identification of available genetic diversity, and enhancement of the accuracy of estimating genetic parameters using marker-based models. The research aids the breeder in the amount of genetic gain that can be anticipated and also paves the way for further genomic exploration in diploid banana breeding programs.

Keywords: Additive variance, clonal model, G-BLUP, genetic diversity, H-BLUP, heritability, Musa, non-additive variance, P-BLUP, prediction accuracy

Genetic improvement of bananas for East Africa

Abstract

Banan (*Musa* spp.) är en viktig gröda både för livsmedelsförsörjningen och som inkomstkälla för miljontals människor i de tropiska och subtropiska delarna av världen. Banan förädling är dock utmanande på grund av det begränsade antalet avkommor som kan utvärderas, osynkroniserade data samt en lång, kostsam och arbetsintensiv förädlingsprocess som resulterar i små genetiska framsteg. Syftet med den här avhandlingen var att undersöka den genetiska diversiteten som finns att tillgå för *Musa*-förädling och att utforska möjligheterna att påskynda genetiska framsteg genom optimering av förädlingsprocessen för Mchare, med tre angreppssätt: 1) utvärdera grödans genetiska diversitet och populationsstruktur, 2) använda den klonala modellen för att fastställa nivån av genetisk variation och dess uppdelning i additiv genetisk varians, dominans och epistasi, och 3) integrera genomisk information för att förbättra uppskattningar av genetiska parametrar för nyckelegenskaper i bananförsök.

Den genetiska diversiteten var låg, särskilt bland diploider och tetraploider. *Musa*-accessioner med potential att användas för förbättring av föräldralinjer identifierades. Den additiva genetiska variationen var signifikant ($p \leq 0{,}001$) och hade den starkaste genetiska variansen för alla avkastningsrelaterade egenskaper. Därför rekommenderas att fokus läggs på ah förbättra och raljå ut de bästa föräldrarna för att uppnå genetiska framsteg i Mchare-bananpopulationer.

De högsta positiva korrelationerna observerades bland de avkastningsrelaterade egenskaperna. Korrelationen mellan moderplantan och det första rotskottet var signifikant, positiv och av hög magnitud ($0{,}62 > r < 0{,}81$, $p \leq 0{,}001$) för alla avkastningsrelaterade egenskaper. Den uppskattade arvbarheten (h^2) var > 0,50 för alla avkastningsrelaterade egenskaper. H-BLUP (best linear unbiased predictor) hade de högsta prediktionsnoggrannheterna för avkastningsrelaterade egenskaper, medan P-BLUP-modellen hade de högsta prediktionsnoggrannheterna för agronomiska egenskaper. Implikationerna av denna forskning omfattar identifiering av tillgänglig genetisk diversitet, förståelse för de kvantitativa egenskapernas arkitektur och förbättrad noggrannhet i uppskattningen av genetiska parametrar med hjälp av markörbaserade modeller samt en bedömning av de genetiska framsteg som

en förädlare kan förvänta sig. Forskningen banar dessutom väg för vidare genomisk användning inom diploid bananförädling.

Dedication

In loving memory of my late mom, Jane Nantongo, who would have been super proud of this achievement.

To the apples of my eyes, Brielle and Adriel, on whose happiness and love I drew my strength and resilience, and in lieu of the attention I could have given you through this journey.

To any future offspring who needed to be delayed because of this dissertation.

"He has made everything beautiful in its time. He has also set eternity in the human heart; yet no one can fathom what God has done from beginning to end"

Ecclesiastes 3:11

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. **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., and Brown, A. (2024). Genetic Diversity and Population Structure in Banana (Musa spp.) Breeding Germplasm. The Plant Genome, 17(2). DOI: 10.1002/tpg2.20497.

II. **Akech, V**., Bayo, S., Bengtsson, T., Ortiz, R., Swennen, R., Vuylsteke, M., and Brown, A. (2024). Estimation of Seasonal, Additive and Non-additive Genetic Components in Diploid Banana (Musa spp.). Theoretical and Applied Genetics. (Submitted)

III. **Akech, V**., Bayo, S., Batte, M., Bengtsson, T., Nyine, M., Ortiz, R., Swennen, R., Uwimana, B., Vuylsteke, M., and Brown, A. (2024). Integrating genomic additive relationship matrices improves the efficiency in diploid banana breeding. (Manuscript)

Paper I is reproduced with the permission of the publishers.

The contribution of Violet Akech to the papers included in this thesis was as follows:

I. Planned the experiment with supervisors, performed the data analysis, wrote the manuscript together with co-authors and was the corresponding author for the paper.

II. Designed the study, planned the experiment and data collection, performed the data analysis, wrote the manuscript together with co-authors and paper and was the corresponding author for the paper.

III. Designed the study, planned and performed the data analysis and wrote the manuscript together with co-authors.

Related work that Violet Akech has been part of but not included in this thesis:

I. Nyine, M., Uwimana, B., Blavet, N., Hřibová, E., Vanrespaille, H., Batte, M., **Akech, V**., Brown, A., Lorenzen, J., & Swennen, R. (2018). Genomic prediction in a multiploid crop: genotype by environment interaction and allele dosage effects on predictive ability in banana. *The Plant Genome*, 11(2), https://doi.org/10.3835/plantgenome2017.10.0090

II. Batte, M., Nyine, M., Uwimana, B., Swennen, R., **Akech, V**., Brown, A., Hovmalm, H. P., Geleta, M., & Ortiz, R. (2020). Significant progressive heterobeltiosis in banana crossbreeding. *BMC Plant Biology*, 20, 1–12, https://doi.org/10.1186/s12870-020-02667-y

III. 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

IV. Uwimana, B., Nakato, G. V., Kanaabi, R., Nasuuna, C., Mwanje, G., Mahuku, G. S., **Akech, V**., Vuylsteke, M., Swennen, R., & Shah, T. (2024). Identification of the Loci Associated with Resistance to Banana *Xanthomonas* Wilt (*Xanthomonas vasicola* pv. *musacearum*) Using DArTSeq Markers and Continuous Mapping. *Horticulturae*, 10(1), 87, https://doi.org/10.3390/horticulturae10010087

List of tables

List of figures

Abbreviations

AMOVA	Analysis of molecular variance
BLUP	Best linear unbiased prediction
DAPC	Discriminant analysis of principal component
DArTseq	Diversity array technology sequencing
EAHB	East African highland bananas
G-BLUP	Genomic best linear unbiased prediction
H-BLUP	Hybrid best linear unbiased prediction
LMM	Linear mixed model
P-BLUP	Pedigree best linear unbiased prediction
PCoA	Principal coordinate analysis
SNP	Single nucleotide polymorphism

1. Introduction

Bananas and plantains (*Musa* spp.) are the one of the world's favourite fruit, grown in the tropical and sub-tropical regions. With an estimated global production of 140 million metric tons, 85% of the world's banana crop is grown by small-scale farmers for food and as a product for sale at local and regional markets (De Buck & Swennen, 2016; Evans *et al*., 2020; Ochola *et al*., 2013). Banana produces fruit throughout the year (Lescot, 2020; Nayar, 2010), making it a critical crop for more than 70 million people in Africa and a staple for over 30 million people in the Great Lakes Region of East Africa (that includes Burundi, Rwanda, north-eastern Democratic Republic of the Congo, Uganda and north-western Kenya and Tanzania). The East African highland bananas (EAHB) and plantains make up approximately 70% of all bananas grown on the continent (Adeniji *et al*., 2010; Pillay *et al*., 2001). The EAHBs dominated by the triploid Matooke (AAA subgroup) in Uganda and the diploid Mchare (AA subgroup) in Tanzania, represent a unique set of *Musa* germplasm mainly found in the East and Central Africa (ECA) high land region (Tushemereirwe *et al*., 2015) that are cooked before consumption.

The growth in domestic demand volume over the next 30 years, for plantain and banana in Sub-Saharan Africa has been estimated at 40 million metric tons and 24 million metric tons, respectively (IFPRI, 2022). To satisfy this demand and world demands, only 10% of these demands may be met by increasing the area under cultivation; the remaining 90% must be met by increasing yields per unit area (harvests per hectare) (77%) and cropping intensity (yields per year) (13%) (IFPRI, 2022). Current yields in metric tons per hectare per year in East Africa are still low (Uganda: 5.5 t ha^{-1} $year^{-1}$, Tanzania: 8.4 t ha^{-1} $year^{-1}$, Rwanda: 8 t ha^{-1} $year^{-1}$, Kenya: 4.5 to 10 t ha^{-1} $year^{-1}$ and Burundi: 5 t ha^{-1} $year^{-1}$) (BMGF 2014; Gaidashova *et al*., 2010; Kamira

et al., 2013;), compared to potential yields of 70 metric tons per hectare per year (Ndabamenye *et al*., 2012; Stevens *et al*., 2020; Wairegi *et al*., 2010). The low yield is attributed to the banana pathogens, pests, and abiotic constraints including, but not limited to, declining soil fertility and factors associated with climate change such as drought and heat stress (Uwimana *et al*., 2020; van Asten *et al*., 2011). Genetic improvement has proved to be the most sustainable approach to increase yield by developing improved hybrids showing stable high yields, resilience to biotic and abiotic stresses, and meeting the consumers' quality preferences.

These constraints are the major focus of the banana breeding programs with the general objective of developing improved hybrids with high and stable yield and quality to meet the market demands of consumers, using genetic improvement as the most sustainable approach. However, improving bananas is challenging since it is a clonally propagated crop and asexual behaviour is often a barrier to cross-breeding and the most important cultivars are sterile and hardly produce seeds even when hand-pollinated (Batte *et al*., 2019; Pillay & Tripathi, 2008; Ray, 2002). The polyploidy nature of the crop hampers the breeding process due to reduced fertility and hence, poor seed set (Batte *et al*., 2019; Ssebuliba *et al*., 2005). The breeding process is also slow and expensive due to the growth habit and space requirements of bananas; it takes up to three years from seed to seed and requires 6 to 9 m^2 per plant. A narrow genetic base in the bananas further compounds the problem. Furthermore, the perennial nature of the crop creates a very heterogeneous field of plants at different developmental stages, making evaluations and selections difficult for breeders (Tixier *et al*., 2004).

This thesis aimed to lay the foundation towards the improving the core breeding clones used as diploid and tetraploid parents in *Musa* breeding schemes and to aid in the optimization of the breeding pipeline of the IITA Mchare breeding program of Tanzania. The focus on the Mchare banana for the second part of this thesis is because the hybrids used in this study, developed from the diploid x diploid crosses are confirmed to be diploids. This makes them more useful in understanding banana genetic architecture at a more basic ploidy level. To achieve this, three research approaches were explored: (i) determining the genetic diversity and population structure among banana germplasm used as parental material in five major banana-breeding programs using DArTseq single nucleotide polymorphism (SNP)

markers; (ii) using the clonal model and the pedigree-based BLUP procedure to determine the level of genetic variation and its decomposition into additive and non-additive (dominance and epistasis) components, and estimate the breeding values to enhance selection efficiency for key traits of interest and (iii) integrating genomic information into the clonal model to improve the partitioning of the variance components, yielding more accurate estimates of heritability and genetic effects to improve efficiency in diploid banana breeding.

2. Background

2.1 Origin and domestication of banana

Bananas and plantains collectively known as bananas are large perennial herbaceous monocotyledonous plants. The wild species of bananas are diploids ($2n = 2x$) with 14, 18, 20 or 22 chromosomes (Deepthi, 2016). The cultivated bananas with 22 ($2n = 2x$), 33 ($2n = 3x$) or 44 ($2n = 4x$) chromosomes arose from intra- and interspecific hybridization between two wild diploid species that belong to section *Eumusa* of the genus *Musa* (Čížková *et al.*, 2015; Deepthi, 2016; Simmonds & Shepherd, 1955), namely *M. balbisiana* Colla that provided the B genome and *M. acuminata* Colla that donated the A genome. Recent research, however, have indicated hybridizations with unidentified ancestral contributors from uncharacterized gene pools (Jeensae *et al.*, 2021; Martin *et al*., 2020; Sardos *et al*., 2022). A wide range of genomic combinations and ploidy levels were produced because of this hybridization *n* and 2*n* gametes of the wild species, as well as diploid (AA, AB), triploid (AAA, AAB, ABB) and tetraploid (AAAB, AABB, ABBB) cultivars (Manzo-Sánchez *et al*., 2015; Nayar, 2010). It is from these combinations that we have, for instance, East African highland cooking (Matooke AAA and Mchare AA), East African highland beer bananas (AAA), dessert bananas (AAA and AAB) and Plantains (AAB).

The earliest domestication of bananas (*Musa* spp.) is considered to have been in Southeast Asia and the Pacific regions (Castillo & Fuller, 2012) from where the crop was introduced to other regions of the world (Simmonds, 1962). Archeobotanical, cultural, and linguistic evidence combined with genetic results, suggest that bananas came into Africa between 2,000 and 6,000 years ago through the Indian Ocean islands such as Madagascar,

Zanzibar, Comoros, and Pemba by various waves of migration of the Austronesian populations from Southeast Asia (Blench, 2009; De Langhe, 2009; De Langhe *et al*., 1994; Mbida Mindzie *et al*., 2001; Perrier *et al*., 2019). Evidence of banana cultivation from the archeobotanical studies, and the indicated number of cultivars and their socioeconomic importance in the humid tropics of West and Central Africa elucidates that plantains have been cultivated for more than two millennia (De Langhe, 1964; Mbida Mindzie *et al*., 2001). It has been proposed that Bornean migrants who travelled over the coastal regions of South Arabia may have carried plantains from India or Sri Lanka to Africa (Fuller & Madella, 2009; Lejju *et al*., 2006).

2.2 Banana classification and diversity

Bananas and plantains belong to the genus *Musa*, the order Zingiberales and the family *Musaceae*. The genus *Musa* with about 70 confirmed species was previously divided into five sections: *Australimusa* ($2n = 2x = 20$), *Callimusa* ($2n = 2x = 20$), *Eumusa* ($2n = 2x = 22$), *Rhodochlamys* ($2n = 2x = 22$), and *Ingentimusa* ($2n = 2x = 14$) (Swennen and Vuylsteke, 2001; Daniells *et al*., 2001; Wong *et al*., 2002). However, recent revision by Häkkinen (2013) and evidence from molecular studies (Hřibová *et al*., 2011) recognize only section *Callimusa*, which combines *Australimusa*, *Callimusa* and *Ingentimusa*, and section *Musa*, which combines *Eumusa* and *Rhodochlamys*. Most cultivated bananas arose from section *Eumusa*, which is the largest containing 11 species namely, *M. acuminata, M. balbisiana, M. schizocarpa, M. basjoo, M. itinerans, M. flaviflora, M. sikkimensis, M. cheesmani, M. nagensium, M. halabanensis* and *M. ochracea*. *Musa acuminata* is the most diverse with several subspecies *including burmannica, siamea, malaccensis, truncata, microcarpa, zebrina, errans and banksii* (Perrier *et al*., 2011).

A distinct group of about 70 cultivars endemic to the East African region referred to as the East African Highland Bananas (EAHB) arose by large variability caused by somatic mutations or possibly epigenetics. These include both the triploid East African highland cooking and beer bananas (AAA) and the diploid Mchare bananas (AA) making up approximately 64% of all bananas grown in Africa (Lescot, 2020). Hence, East Africa is a secondary centre of banana diversity (De Langhe, 2009; Kitavi *et al*., 2016). Events of natural mutations over a long period may have led to a

diversification of African plantain as suggested by evidence from a large number of African plantain cultivars that have never been recorded elsewhere (Adheka *et al*., 2018a; Adheka *et al*., 2018b; Adheka *et al*., 2018c; Blench, 2009; De Langhe, 2007). These mutations gave rise to secondary plantain cultivars (De Langhe, 1964; De Langhe & de Maret, 2004), consequently making West and Central Africa a centre of secondary diversity for the plantain (Swennen, 1990).

2.3 Importance and production of banana

Bananas are valuable for food security as a major staple in dietary needs and as a source of supplemental income for hundreds of millions of people in the global tropical and subtropical South (Aurore *et al*., 2009; Kilimo, 2012; Ploetz & Evans, 2015). In East, West, and Central Africa, bananas provide about 25% of food energy requirements for about 90 million people (Sharrock *et al*., 2001). The non-seasonality of the banana crop guarantees the production of fruits throughout the year (Calberto *et al*., 2015; Lescot, 2020; Nayar, 2010), serving as a food security crop and ensuring regular income flow for the farmers. The dessert type (Cavendish, AAA) eaten when ripe yellow, is the most consumed banana in temperate countries and is grown for export, thus being a source of income for the producing countries (FAO, 2014). Pome, Silk, Mysore and Sukali Ndizi (AAB bananas) are also consumed as dessert bananas in other regions. Plantains are AAB bananas with high starch content and high provitamin A content (Amah *et al*., 2019a; Amah *et al*., 2019b), with the fruit remaining firm even after ripening. They are mostly eaten after roasting and or as plantain chips. The EAHB, include the cooking types ‘Matooke’ (AAA) and ‘Mchare’ (AA) and brewing types ‘mbidde’. The former types is cooked when fresh green in different forms while the latter, very astringent due to high tannin content, is allowed to ripen before the juice is squeezed out of the pulp and fermented to make beer, hence the name beer banana.

Global banana production increased by 15.4%, from 98.8 million metric tons in 2008 to 114 million metric tons in 2017 (Evans *et al*., 2020) and is currently reported to be at 140 million metric tons (Lescot, 2020). Asia is the leading banana-producing continent, accounting for 55.9% of the total production, followed by Africa (24.6%), America (16.1%), the Pacific islands (1.7%), Oceania (1.3%) and Europe (0.4%) (FAO, 2022). India is the

highest producer of ABB cooking bananas, which are starchy fruits and sometimes cooked when ripe.

In Africa, 3.2 million farming households depend on plantains and 2.5 million farming households depend on bananas. They produced 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), respectively (Lescot, 2020). The Great Lakes Region of East Africa is the leading producer of bananas (comprising highland bananas, ABB and other cooking types) in Africa accounting for 88% of the total production with Uganda as the largest producing country in Africa with 3.1 million metric tons of harvest followed by Tanzania with 2.7 million metric tons (Lescot, 2020).

2.4 Production challenges of banana

Banana yields in East Africa are currently only 9% of their potential yield, despite the crop's significance and high demand (Ndabamenye *et al*., 2012; Stevens *et al*., 2020; Wairegi *et al*., 2010). Banana production is affected by both biotic and abiotic factors causing considerable fruit yield loss and reduced plantation life span (Viljoen *et al*., 2017). The most economically important pests are the banana weevil (*Cosmopolites sordidus*) and nematodes (*Radopholus similis* and *Helicotylenchus multicinctus*) (Gold *et al*., 2004; Ocan *et al*., 2008; Speijer *et al*., 1999), while the devastating diseases include banana bacterial wilt caused by *Xanthomonas vasicola* pv. *musacearum* (Biruma *et al*., 2007; Nakato *et al*., 2019; Tinzaara *et al*., 2016), black leaf streak disease (also known as black Sigatoka) caused by the fungus *Mycosphaerella fijiensis* Morelet (Barekye *et al*., 2009; Marin *et al*., 2003), Fusarium wilt caused by *Fusarium oxysporum* f. sp. *cubense* (Arinaitwe *et al*., 2019) and banana bunchy top disease, caused by the banana bunchy top virus and transmitted by aphids *Pentalonia nigronervosa* (Boloy *et al*., 2014; Ocimati *et al*., 2021; Shimwela *et al*., 2022).

Among the abiotic stresses, declining soil fertility (Wairegi *et al*., 2010), drought and heat (Stevens *et al*., 2020; Uwimana *et al*., 2020; Vantyghem *et al*., 2023; Varma & Bebber, 2019) are the most prevalent. The development of improved banana hybrids with host plant resistance and tolerance to these biotic and abiotic constraints, respectively, is the most economical strategy to improve banana production.

2.5 Banana breeding programs and strategies

The inter- and intraspecific hybridization events between the wild progenitors gave rise to hybrids that had lost many of the wild characteristics and had attributes attractive to humans, such as high yield, plant vigour, seedlessness and palatability of fruits (Simmonds, 1962). Later there were deliberate efforts and strategies to develop improved banana hybrids that could overcome the production constraints that bananas were facing, initially in 1922 in Trinidad, from where the first breeding program was then expanded to Jamaica in 1924. The Fundación Hondureña de Investigación Agrícola (FHIA), established in Honduras in 1984, was the first breeding program to successfully release farmer-acceptable improved hybrids.

Apart from FHIA, there are a few globally recognized cross-breeding initiatives. One such initiative is the International Institute of Tropical Agriculture (IITA) in Nigeria, where research on bananas and plantains began in 1976 but practical breeding only started in 1987 (Ray, 2002) when IITA accepted the responsibility of breeding plantain and cooking bananas for Africa with resistance to black leaf streak. This led to the establishment of the Plantain and Banana Improvement Program (PBIP) in 1991 (Swennen Vuylsteke, 1993; Vuylsteke *et al*., 1993a). The first breeding strategy was crossing plantains with wild bananas leading to the first milestone of the registration of 14 improved tropical *Musa* plantain hybrids with resistance to black leaf streak in 1993 (Vuylsteke *et al*., 1993b). Breeding efforts in Uganda to enhance the EAHB were initiated by the late Dirk Vuylsteke (Vuylsteke, 2001). A collaboration between IITA and the National Agricultural Research Organization (NARO) was initiated in the 1990s using some of the banana improved diploid parents that IITA had developed at Onne station in Nigeria (Vuylsteke *et al*. 1993a).

Banana breeding in Tanzania started in 2011 by IITA with a focus on improving Mchare diploid bananas for yield, host plant resistance to pests (weevils and nematodes) and pathogens (Fusarium wilt) and acceptable consumer preferences (Brown *et al*., 2017; Madalla *et al*., 2023). In 2021, IITA started a collaboration with the Tanzania Agricultural Research Institute (TARI) where hybrids developed by NARO and IITA, called NARITAs, were given to TARI to run performance trials (Madalla *et al*., 2022a), followed by a handover of four Fusarium wilt-resistant Mchare hybrids in 2023 for multi-locational evaluation by TARI (IITA, 2023). The

primary objectives of IITA and the national program partners are to develop hybrids with consumer-preferred qualities in terms of color, texture, taste, 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 (Brown *et al*., 2017; Nowakunda *et al*., 2023; Ortiz *et al*., 1995; Swennen & Vuylsteke, 1993; Tenkouano *et al*., 2019; Tenkouano & Swennen, 2004; Vuylsteke *et al*., 2010; Vuylsteke, Ortiz, *et al*., 1993).

The initial steps for conventional genetic improvement of bananas and plantains by IITA and NARO have involved crossing the triploid cultivars to be improved, with diploid breeding clones that are disease resistant to produce tetraploid fertile hybrids. The tetraploid selections are thereafter crossed with diploid cultivars, breeding clones or accessions to produce secondary triploids that are sterile (Figure 1). This strategy has led to the development of primary tetraploids and secondary triploids from triploid French plantains Obino L'Ewai, Bobby Tannap, and Mbi Egome by the Plantain breeding program of IITA (Ortiz *et al*., 1995; Tenkouano *et al*., 2019; Tenkouano & Swennen, 2004; Vuylsteke *et al*., 1993a). The same strategy was used in East Africa by IITA and NARO, to develop improved tetraploids (AAAA) from EAHBs (AAA- landraces) by crossing to the wild-seeded fertile male parent, Calcutta 4 (AA). The improved tetraploids developed such as '365K-1', '1201K-1', '917K-2', '660K-1', '1438K-1', and '222K-1' were fundamental in the development of the 27 high yielding NARITA triploid banana hybrids with host plant resistance to the black leaf streak (Brown *et al*., 2017; Madalla *et al*., 2022a; Madalla *et al*., 2022b; Tushemereirwe *et al*., 2015). Four of these NARITA hybrids were released in Tanzania as TARIBAN 1 (NARITA 4), TARIBAN 2 (NARITA 7), TARIBAN 3 (NARITA 18), and TARIBAN 4 (NARITA 23). They are liked by the farmers due to their resistance to black leaf streak, high yield (18 to 35 metric tons ha^{-1} $year^{-1}$) and acceptable consumer qualities (Madalla *et al*., 2022a; Madalla *et al*. 2022b).

To select and advance the best-performing offspring as improved hybrids or improved parental breeding material, bananas are evaluated in a series of field trials such as early evaluation trials (EET) and preliminary yield trials (PYT) (Brown *et al*., 2017; Ortiz, 2015; Ortiz & Vuylsteke, 1994). The hybrids are initially set up in an EET, where host plant resistance to black leaf streak and bunch characteristics are evaluated throughout at

least two crop cycles (plant crop and first ratoon crop). Following this trial, a PYT is planted with the selected hybrids, and the yield potential as well as

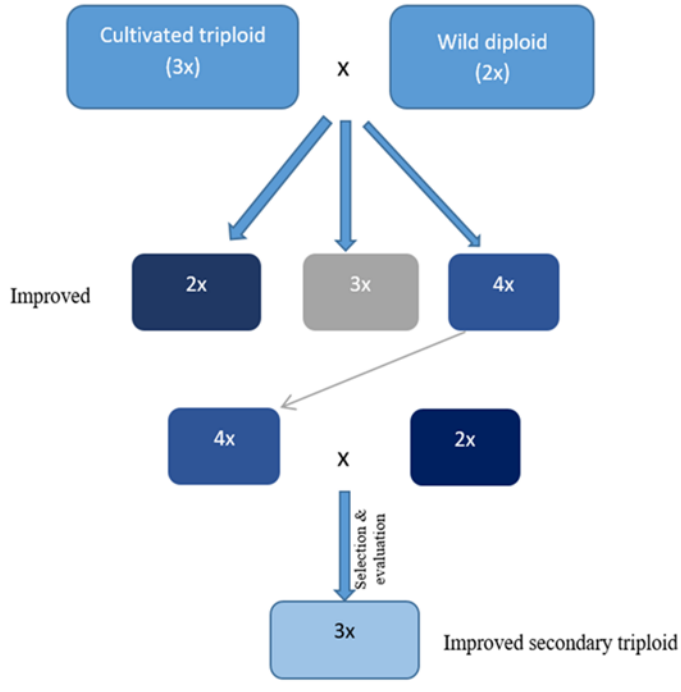


Figure 1. The general crossbreeding scheme for banana and plantain improvement of the International Institute of Tropical Agriculture (IITA) in sub-Saharan Africa (Akech, 2023).

agronomic and quality attributes are assessed for two crop cycles. In the Advanced Yield Trials (AYT), which are multi-site and off-station and sometimes referred to as Multi-Environment Trials (MET), selections from PYT are assessed for stability in a variety of environmental conditions (Madalla *et al*., 2022b). This process by IITA and NARO of Uganda and now also used by TARI is summarized in Figure 2.

In Brazil, the Empresa Brasiliera de Pesquisa Agropecuaria (Embrapa) was established in 1982 with the focus on improving the Pome and Silk, 'AAB' bananas for resistance to *Fusarium*, black Sigatoka, insects and parasitic nematodes, reducing plant height and length of plant cycle and increasing yield. The Embrapa banana breeding program employs two strategies for the development of banana hybrids; hybridization between

improved diploid and triploid parents and the use of tetraploid hybrids as female parents and improved diploids as the male parent (Amorim *et al*., 2011a; Bakry *et al*., 2009). The program is based mainly on the production of AAAB tetraploids by crossing improved diploids (AA) with Prata (Pome) and Maca (Silk) triploid bananas (AAB), aiming for a final product of tetraploid sweet varieties such as Pacovan Ken: AAAB, Preciosa: AAAB and Tropical: AAAB.

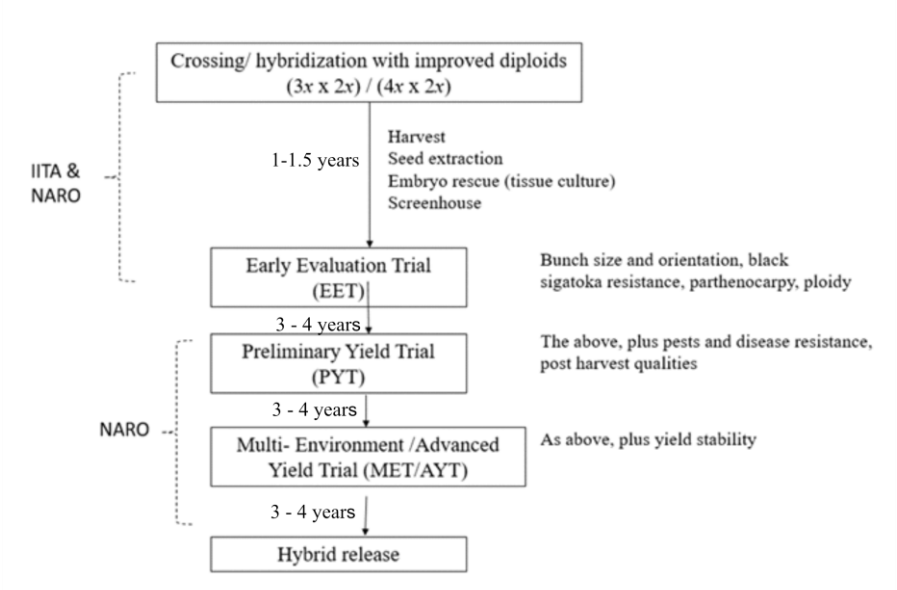


Figure 2. Typical flow diagram of the banana evaluation and selection process used by IITA and NARO (Adopted from Akech, 2023).

Such improved diploids include 'SH3263' of low stature, productive and resistant to *Radopholus similis* and the 'SH3362' with low stature, productive and resistant to Race 4 of fusarium (Amorim *et al*., 2011a; Ribeiro *et al*., 2018; Silva *et al*., 2001).

The National Research Centre for Bananas (NRCB) of India incorporates both conventional and new breeding techniques in the improvement of bananas. Potential tetraploids such as 'BRS-01' Pome and 'BRS-02' hybrids from crosses between the triploids 'Agniswar' and 'Vanna' with a diploid 'Pisang Lilin' have been developed. Secondary triploids with tolerance to leaf spot disease were produced from the primary tetraploid hybrids by crossing with *acuminata* diploids (AA), 'Matti',

'Tongat' and 'Sanna Chenkadali'. Doubled-diploids through *in vitro* polyploidization of Indian diploid clones have resulted in the synthesis of doubled-diploids (Bakry *et al*., 2009; Sathiamoorthy *et al*., 2001).

Even though the agronomic attributes towards high yield and pest (weevils and nematodes) resistance are crosscutting for all regions and breeding programs, the relative importance of the diseases and quality attributes varies across the regions. For example, Fusarium wilt race 1 is of higher importance to Brazil, India and Tanzania breeding programs while black Sigatoka is of more importance to all the programs other than in Tanzania. Similarly, the fruit quality preferences for each program vary according to the end user's taste and need. The NRCB targets among other quality traits, total soluble solids (TSS) and acidity as quality traits for value-added products such as wine and juice (Sathiamoorthy *et al*., 2001) while Embrapa targets biofortification to increase Iron, Zinc and Vitamin A (Amorim *et al*., 2011b) in the hybrids. The IITA and NARO breeding programs have quality traits such as consumer-specific texture, taste, aroma and pulp color making their way to the product profile as they have been reported to affect consumer acceptability and adoption of improved banana hybrids (Madalla *et al*., 2023; Marimo *et al*., 2019; Nowakunda *et al*., 2022). The Katholieke Universiteit Leuven, Belgium maintains the world's largest banana germplasm collection, called the International Musa Germplasm Transit Centre – ITC, on behalf of the Alliance of Bioversity International and the International Centre for Tropical Agriculture (CIAT). This supports the activities of these major breeding programs as a source of germplasm (Ray, 2002; Lorenzen *et al*. 2010).

2.6 The Mchare banana and breeding program

Mchare are edible parthenocarpic diploid (AA) bananas that are a staple crop for millions of subsistence farmers in Tanzania, grown in the north-central and southern highlands of Tanzania and are preferred for their unique textural characteristics (Brown *et al*., 2017; Onyango *et al*., 2009; Perrier *et al*., 2019). Known as 'Mlali' in the Comoros Islands and 'Muraru' in Kenya (De Langhe *et al*., 2001; Perrier *et al*., 2019), Mchare is relatively unique as it is one of the few diploid (AA) bananas that are cultivated on a large scale yet they have been given less attention. The Mchare, as observed (Figure 3) and described by Onyango et al. (2009), have an erect leaf habit typical of

their diploid ploidy level, and can grow tall up to more than five meters. They have a cream-colored compound tepal with brown pigmentation. The fruits have short pedicels and gradually tapering points. The mature fruits remain green and shiny with thicker peels and a firm texture. Distinctively from the Matooke, they have purple blotches on the leaves of water suckers, just like the Cavendish and Gros Michel. The rachis has few neutral flowers for most Mchare cultivars, but some have persistent withered bracts on the rachis. The pseudostem color is medium green with a little rusty brown to black pigmentation.

Figure 3. Typical Mchare banana characteristics (A) erect leaves (B) pseudostem (C) rachis (D) male flower/bud (E) compound tepal with bract (F) pedicel (G) mature bunch (H) fruit cluster (I) fruits/fingers.

Their diploid nature makes them more useful in understanding banana genetic architecture at a more basic ploidy level than most triploid bananas such as Cavendish dessert (AAA), Matooke (AAA) and Plantain (AAB) bananas, making them an important model crop in the conventional improvement of bananas worldwide. Mchare is phenotypically diverse, but as in plantain and Matooke, these cultivated bananas have been reported to have a very narrow genetic base (Kitavi *et al*., 2016; Nyine *et al*., 2017; Ortiz, 1997; Perrier *et al*., 2011) and all Mchare are considered to be of the same clone set (Onyango *et al*., 2009). The Mchare subgroup includes cultivars such as 'Mshale', 'Huti (Shumba Nyelu)', 'Mlelembo', 'Kahuti', 'Huti green bell', and 'Akondro Mainty' among others (Perrier *et al*., 2019). Despite being genetically homogenous, they are adapted to wider ecological

environments ranging from sea level to elevations above 1,500 m.a.s.l (De Langhe *et al*., 2001).

Matooke and Mchare are referred to as East African Highland Cooking bananas (EAHBs) as a reference to their distribution across the highlands of East Africa, but they are not genetically related and the variability within these subgroups arose from somatic mutations of single clones, derivatives of *M. acuminata* (Kitavi *et al*., 2016; Perrier *et al*., 2019). Mchare cultivars are the sole surviving diploid subgroup that is reported to be the donor of unreduced gametes to many of the most widely grown and successful triploid dessert bananas such as Cavendish, Gros-Michel, Silk, and Prata (Hippolyte *et al*., 2012; Martin *et al*., 2020; Perrier *et al*., 2019; Raboin *et al*., 2005).

The breeding scheme of diploid Mchare bananas of Tanzania involves crossing diploid (2*x*) Mchare landrace cultivars as females and diploid (2*x*) male accessions as sources of resistance genes to pests and pathogens (Bayo *et al*., 2024). The focus is on developing hybrids improved for yield, host plant resistance to pathogens, especially Fusarium wilt, and acceptable consumer preferences (Brown *et al*., 2017; Madalla *et al*., 2023).

2.7 Banana breeding and evaluation challenges

The challenge of banana improvement through crossbreeding resides in the inherent polyploidy nature of the crop where the homologous chromosomes cannot pair up during synapsis of meiosis, as they are uneven, affecting meiosis, and production of viable gametes. This is characterized by reduced female and male fertility, poor seed set, and low seed germination rates, which makes it difficult to produce large offspring for evaluation in well-balanced trials (Aguilar, 2013; Batte *et al*., 2019; Ortiz, 2013; Pillay & Tripathi, 2008; Vuylsteke, 2000). The reported seed set rate for triploid Matooke ranges from 1.5 to 26 seeds per bunch pollinated (Batte *et al*., 2019; Ssebuliba *et al*., 2005) while for the Cavendish group, only 40 viable tetraploid embryos from the pollination of over 20,000 bunches were obtained (Aguilar, 2013). In triploid plantain, a seed rate of up to 219 seeds per bunch has been reported (Swennen & Vuylsteke, 1993).

The above is further complicated by the long growth cycle (two years from seed to seed) and the subsequent evaluation over a minimum of two crop cycles (mother plant, known as plant crop or cycle 1, and first ratoon,

known as cycle 2) (Brown *et al*., 2017). Banana plants being large, evaluation plots are equally large requiring 6 to 9m^2 per plant, making the banana improvement process slow, expensive and tedious (Batte *et al*., 2019). The perennial nature of banana creates a field consisting of a mix of individual plants, each at a different developmental stage, which is a challenge for a breeder to efficiently evaluate and select potential hybrids due to the overlapping generations and unsynchronized data among genotypes and even plants of the same genotype (Tixier *et al*., 2004). Progress in genetic banana improvement is also hindered by a low genetic variation among the edible cultivated bananas such as the EAHB and plantains that need to be improved for pest and pathogen resistance or other agronomically important characteristics (dwarfness, duration, yield, quality of fruit etc.) (Akech *et al*., 2024; Kitavi *et al*., 2016; Nyine *et al*., 2017; Ortiz, 1997).

2.8 Improving banana breeding efficiency to increase genetic gains

The development of high-yielding banana hybrids through breeding for host plant resistance to several pathogens and pests has become one of the strategies to revive the declining banana production (Madalla *et al*., 2023; Tushemereirwe *et al*., 2015). These hybrids have been developed through cross-breeding, which is slow, tedious, very expensive and relies on a limited genetic base. Improvement of the breeding pipeline through the use of best practices, continuous optimization of the breeding strategy and operations and more effective use of resources will lead to increased rates of genetic gain.

Based on the genetic gains equation (equation 1) of Moose and Mumm (2008), the most promising innovations for increasing the rate of genetic gain in banana breeding include the identification and increase of useful diversity and genetic variation (Sanchez et al., 2023) and reduction of the breeding and selection cycle.

$$\Delta G = \frac{h^2 \sigma_p i}{L}, \qquad (1)$$

Where:

ΔG is the genetic gain,

h^2 is the narrow-sense heritability,

σ_p is the phenotypic standard deviation in the population before selection,
i is the selection intensity, and
L is the length of the selection cycle.

A key element of plant breeding is accessible genetic variation, which offers a pool of alleles that can be used to improve crops. A higher accumulation of additive genes by making crosses with complementary genotypes is expected to increase genetic gains in populations (Batte *et al*., 2020; Stringer *et al*., 2011; Zhou & Mokwele, 2016). Genetic parameters such as breeding values, general combining ability (GCA) and specific combining ability (SCA) are useful tools for identifying parents of superior genetic merit, and for identifying the best cross-combinations based on the performance of the crosses (Fasahat *et al*., 2016; Viana & Matta, 2003). Implementation of parental improvement and selection criteria by determining the breeding values of parental material allows a breeder to manage genetic diversity, set the desired rate of genetic gain, and rapidly increase the frequency of favorable alleles in breeding populations (Cobb *et al*., 2019).

Given the long breeding and selection cycle observed in bananas as illustrated in Figure 2, reducing this time will result in accelerated selection efficiency and increased genetic gains as breeding products can be released much faster to consumers. There are several ways faster genetic gain can be achieved in bananas. For instance, the use of molecular tools such as DNA markers in innovative approaches such as genomic selection can help shorten the banana breeding and selection cycle. Superior parents and hybrids can be predicted using optimized prediction models at the nursery level saving time and costs of field evaluations (Nyine *et al*., 2018). If there exists a positive correlation between the different crop cycles (plant crop, first and second ratoon crop) for key traits, the length of the breeding cycle can be reduced further by selecting the highest performing hybrids at plant crop (cycle 1) enabling faster breeding decisions thus improve genetic gain. A correlation between key traits can also reduce the breeding cycle as significantly correlated traits can be improved simultaneously.

3. Aims and objectives

3.1 General objectives

The aim of this thesis was divided into two parts (i) a description of the genetic diversity and population structure of *Musa* germplasm available in the breeding programs representative of the four largest consumers of bananas worldwide and at the *Musa* conservation centre (ITC) and (ii) optimization of the breeding pipeline of the diploid Mchare banana breeding program of Tanzania to increase selection accuracy and improve genetic gains.

3.2 Specific objectives

Assess and document the level of genetic diversity available and the pattern of population structure among the *Musa* breeding materials from five breeding programs and *Musa* conservation centre; International Institute of Tropical Agriculture (IITA) in Nigeria for plantain, Tanzania for Mchare, IITA-NARO breeding programs for Matooke in Uganda, the Brazilian Agricultural Research Corporation (Embrapa) in Brazil, and the National Research Centre for Banana (NRCB) in India; and accessions from the *in vitro* banana gene bank at the International Transit Centre (ITC) of Alliance of Bioversity International and CIAT, and identify compatible new sources of germplasm harbouring valuable variation for improving resistance to biotic stresses (Paper I).

Determine the level of genetic variation and its decomposition into additive and non-additive (dominance and epistasis) using the clonal model and demonstrate how the pedigree-based BLUP procedure enhances selection efficiency for key traits of interest in diploid banana breeding (Paper II).

Investigate if integrating genomic information into the clonal model improves the partitioning of the variance components, yielding more accurate estimates of heritability and genetic effects to improve efficiency in diploid banana breeding (Paper III).

4. Materials and methods

4.1 *Musa* genetic diversity and population structure

This study used 856 *Musa* accessions, with diverse ploidy and genomic constitutions from five breeding programs and the ITC (Table 1). The material from the breeding programs was grown *in situ* in different field trials at their respective sources. Germplasm details and passport data are provided in Table 1 and Table S1 of Paper I.

Table 1. Source of the 856 *Musa* spp. accessions

Region	Source	Breeding materials	Name in the study	Number of accessions
East Africa	IITA & NARO-Uganda	Triploid Matooke	Matooke	111
	IITA-Tanzania	Diploid Mchare	Mchare	205
West Africa	IITA-Nigeria	Triploid Plantain	Plantain	34
Asia	NRCB-India	Triploid Silk/Pome, Dessert, plantain, and ABB cooking	SwS (Sweet and Starchy)	154
South America	Embrapa-Brazil	Triploid Prata (Pome)/Silk and plantain	Prata	154
International Transit Centre (ITC)	Conservation		ITC	205

The sampling procedure and treatment of samples are described in Paper I. All dried samples collected and pulverized with a geno grinder at the Nature Source Improved Plants facilities (NSIP) in Ithaca, USA, were sent to Diversity Array Technology (DArT), Australia, for DNA extraction and genotyping. DNA extraction and genotyping were performed by DArT and library construction using the DArT sequencing (DArTSeq) complexity reduction method was done at and by DArT (Kilian *et al*., 2016; Kilian *et al*., 2012; Sansaloni *et al*., 2010). Next-generation sequencing was performed using the sequencer HiSeq 2500 (Illumina) and DArTseq markers scoring was performed using DArTsoft14, which is an in-house marker-scoring pipeline (Kilian *et al*., 2016). The reads were aligned to version 4 of the "DH Pahang" banana reference genome (Belser *et al*., 2021). To reduce complexity in data analysis all accessions were called as diploids during genotype calling. For the downstream analysis for this study, the final SNP calls were coded as "0/0" for homozygous reference allele, "1/1" for homozygous alternate allele, and "0/1" for heterozygotes.

Data quality control, filtering and analysis for this study were done using the "dartR" package in R (Gruber *et al*., 2018) for the entire collection, and then the samples were grouped according to ploidy levels into diploids and tetraploids. The details of the retained marker sets for each grouping are reported in Paper I.

The SNP marker information using polymorphism information content (PIC), MAF and population diversity indices such as observed heterozygosity (Ho), expected heterozygosity/gene diversity (He), Shannon's information index or Shannon's diversity index (I), fixation index (FIS), and percentage of polymorphic loci (%Pl) were computed using "dartR" package in R (Gruber *et al*., 2018). The number of individuals for which these parameters are calculated is reduced from the original number of individuals retained after filtering, due to imputation of individuals with missing information during the analysis. The pairwise Nei's standard genetic distance (DS) was computed using the "dartR" package in R (Gruber *et al*., 2018) to determine the between-group differentiation

Population structure was assessed to determine genetic sub-population and admixture patterns, using the discriminant analysis of principle components (DAPC) as described in Paper I and by the Bayesian model-based clustering implemented in STRUCTURE (v2.3.4) (Pritchard *et al*., 2000). The analysis was computed using a burn-in-period of 50,000 and

100,000 Markov Chain Monte Carlo (MCMC) replications in ten independent runs from K=1 to K=10. The structure output was analysed using Structure Harvester (Earl & VonHoldt, 2012) which enabled the identification of the optimal number of sub-groups (K) using the method of Delta K (ΔK) (Evanno *et al*., 2005). ΔK values are calculated from the logarithm probability relative to standard deviation (ln P(D)). Admixture was decided at membership coefficient, $Q \geq 0.75$.

4.2 Estimation of trait variation, partitioning genetic variance and genetic parameters

Forty-eight clone hybrids from 12 full-sib families were used. They were generated by crossing eight diploid female Mchare cultivars with two wild diploids and one parthenocarpic male in a factorial design. The hybrids were confirmed to be diploids by flow cytometry using a Sysmex ploidy analyzer (CyFlow® - Sysmex, Software- CyViewTM 1.6, REF-CY-S-3039). The trial, referred to as the Early Evaluation Trial (EET 1) was set up at the IITA banana breeding statin in Arusha, situated at the Nelson Mandela African Institution of Science and Technology. The trial was laid out in a randomized complete block design with three blocks and each hybrid was represented by three plants per block. The area experiences a bimodal rainfall pattern, consisting of a long rainy season from late March to early June locally referred to as 'Masika'; a short rainy season from October to December locally referred to as 'Vuli'; a season with high temperatures and low relative humidity, locally referred to as 'Kiangazi' lasting from January to late February; a season lasting from July to August with low temperatures and no rainfall called 'Kipupwe' and 'Demani' lasting from late August to September, with high temperatures and high relative humidity. The plants received regular irrigation (rate of 4 L hr^{-1}) for two hours per day, three days a week to ensure each mat received at least 60 L per week. Agronomic management of the field trial is described in Paper II.

Data was recorded over two cycles (plant and first ratoon crop). Traits are summarized in Table 2 (traits without asterisk).

Table 2. Summary of traits considered in Paper II and III

Trait Category	Trait	Abbreviation	When and how measured
Fruit yield	Bunch weight in kg	BW	At harvest, weight of the bunch at full maturity
	Number of fruit clusters in the bunch	NH	Total count of number of clusters on the bunch at harvest
	Number of fingers in the bunch*	NF	Total count of number of fruits on the bunch at harvest
	Fruit external length in cm	FL	Measured at harvest as the length of the external arc of a fruit, without pedicel
	Fruit circumference in cm	FC	At harvest measured midway the length of ten random fingers of a bunch
	Fruit weight in g	FW	At harvest measured as weight of each of ten random fingers from a bunch
	Yield per year (kg)	Yield	A function of bunch weight, and crop cycle from planting to harvest of the plant crop
Agronomic	Plant circumference	PG	At flowering, as girth of the pseudostem at 100 cm from soil surface
	Plant height in cm	PH	Distance from the collar to the first leaves, at flowering
	Plant stature (PG to PH ratio)	PS	Derived as a ratio (of PG/PH)
	Number of functional leaves	NFL	At flowering, count of leaves that have 50% or more of their surface photosynthetic active tissue
	Plant height of tallest sucker in cm*	PHTS	At flowering of plant crop, distance from the collar to the first leaves
	Number of leaves of tallest sucker*	NLTS	At flowering of plant crop, count of leaves that have 50% or more of their surface photosynthetic active tissue
	Number of suckers at flowering	NSF	Total count of number of suckers at flowering of plant crop
Cycling	Cycling time	CT	Number of days between the flowering of the plant crop (cycle 1) to the flowering of the first ratoon crop (cycle 2)
	Days from planting to flowering	DPF	Difference between planting and flowering
	Days from flowering to harvesting	DFM	Difference between flowering and harvesting
	Days from planting to harvest*	DPH	Difference between planting and harvesting

*Indicates traits that were not included in Paper II

4.3 Genomic additive relationship matrices to improve efficiency in diploid breeding

The trial established as an EET 2, was conducted at the same location and was managed in the same way as in Paper II. In a factorial mating scheme, four diploid female cultivars were hand-pollinated with pollen from five male parents producing 59 hybrids from 14 full-sib families. The factorial crossing design utilized for this investigation along with the number of families per cross is presented in Paper II. Using a randomized complete block design, the 59 hybrids and their nine parental genotypes were established in three blocks, with two plants of each genotype per block. The data was collected in the same way as in Paper II for all traits (Table 2) and as described in Paper III. The 2,792 informative SNP markers used to create the genomic relationship matrix in this study were obtained by extraction of deoxyribonucleic acid (DNA) from two grams of cigar leaf samples. For each genotype, 50 µl of genomic DNA at a concentration of 30 ng/µl was sent to Diversity Array Technology (DArT) Pty Ltd, Canberra, Australia, for sequencing (DArTseq). The DNA extraction protocol, DArT genotyping and processing of the genetic markers, including the alignment of the read on the banana reference genome is described in Paper III. Marker data control as well as filtering was performed using the R package 'dartR' (Gruber *et al*., 2018).

4.4 Data analysis for Paper II and III

Linear mixed models (LMM) using the software ASReml-R v. 4.2 (Gilmour *et al*., 2015) were fitted to estimate variance components, heritability, BLUP values, GCA and SCA.

4.4.1 The LMM and relationship matrices

In Paper II, only a pedigree-based additive (A) relationship matrix constructed from the pedigree information of the hybrids was fitted to the data to estimate trait variation and genetic parameters. The traits were categorized into different groups based on the effects fitted for each trait

giving each trait category a specific LMM as provided in Table 2 and equations two to five in Paper II. For example, group 1 contained traits with data collected for cycle one only, hence no cycle interaction; group 2, traits with data collected over two cycles hence cycle interaction with genotype and group 3, traits analysed for two cycles but with multiple measurements (pseudo-replications). In general, the LMMs were constructed for each trait with respect to each cycle, season variation and their interaction with genotype as:

$$Trait_{bfijklmpy} = Block_b + Genotype_i + Cycle_j + Genotype{:}Cycle_{ij} + YearF{:}SeasonF_{fl} + YearP{:}SeasonP_{py} + Cycle{:}YearF{:}SeasonF_{jfl} + Cycle{:}Plot_{jm} + \epsilon_{bfijklmpy}, \quad (2)$$

where the trait has been measured on the kth plant of the ith genotype at cycle j, planted in season p of the year y and flowering in season f in year l. $Block_b$ and $Cycle_j$ are the only fixed terms in the model accounting for the variation between the three blocks and the two cycles, respectively. In all the LMM used in the analyses, the *Genotype* is the genetic effect of interest; the *Genotype:Cycle* accounts for the cycle-specific variation between genotypes. The variation due to seasonal fluctuations across the year during planting and flowering is represented by *YearP: SeasonP* and *YearF: SeasonF* terms, and ϵ represents the random error.

While different LMMs were fitted for each trait group as described in Paper II, the general pedigree-based LMM was constructed as:

$$y = X\beta + Z_1 b + Z_2 c + Z_3 a + Z_4 mf + Z_5 mf.c + \varepsilon, \quad (3)$$

where y refers to the vector of phenotypic observations, β is the vector of fixed effects (e.g. block, cycle), and X is the associated design matrix; b is the vector of "non-seasonal" random (design) effects (e.g. cycle.plot, genotype.cycle), ~ MVN(0, Iσb2), c is the vector of "seasonal" random effects (e.g. year and season of planting, year and season of flowering), ~ MVN(0, Iσc2), a is the vector of random additive genetic effects, ~ MVN(0, Aσa2), with A the pedigree-based relationship matrix among genotypes; mf is the vector of the random male by female interaction or family effects, mf ~ MVN(0, $I\sigma^2_{mf}$); mf.c is the vector of random clonal within family effects,

$mf.c$ ~ MVN(0, I$\sigma^2_{mf.c}$); Zi are the design matrices for the corresponding random effects; and ε is the vector of random residual effects, ~ MVN(0, I σ^2_e).

For Paper III, apart from the pedigree-based relationship matrix (referred to as the P-BLUP model), a genomic additive numerator relationship matrix (A_G) (referred to as the G-BLUP model) containing the observed identity by descent (IBD) proportions for each pair of individuals was calculated based on the set of 2,792 markers, following the formula described by VanRaden (2008). In addition, a third model that combines the A-matrix and the A_G matrix, known as the hybrid genomic relationship matrix or H-matrix (referred to as the H-BLUP model) was employed to estimate genetic parameters. The construction of these models is described in Paper II and Paper III. Linear mixed models (LMMs) were fitted using the A, A_G, and H relationship matrices.

4.4.2 Estimation of trait variation, genetic parameters, prediction accuracies and correlations

Variance components were estimated using the restricted maximum likelihood (REML) method (Patterson & Thompson, 1971). The clonal model was used to partition the genetic component 'Genotype' further into additive, family, and clonal components with the corresponding variance components tested against the reduced model using likelihood ratio tests (LRT). Based on the estimated variance components, BLUPs for additive, dominance, clonal random effects, broad-sense (H^2) and narrow-sense heritability (h^2) were estimated. BLUPs for the parental lines are twice the GCA of the parents and BLUPs for the families are twice the breeding values of the clones and can be interpreted as the SCA. Broad-sense (H^2) and narrow-sense heritability (h^2) were estimated according to Visscher et al. (2008) as the ratio of genetic variance component to the total phenotypic variance and the ratio of additive genetic variance component to the total phenotypic variance, respectively. To compare the reliability (ρ2) and the prediction accuracy (ρ) of the estimated breeding values (EBVs) by the three models the heritability method according to Cullis et al. (2006) was used. This is because this method relies on the correlation between true and predicted genetic values, and not on the proportions of the variance components.

$$\rho^2 = 1 - \left(\frac{PEV}{\sigma_a^2}\right) \quad (4)$$

$$\rho = \sqrt{\rho^2} \quad (5)$$

Correlation analysis among traits were estimated using BLUP values while correlation between the crop cycles (cycle 1 and cycle 2) was estimated using BLUE value because the cycle effect was treated as a fixed effect because the cycles occur during different seasons of time among the different genotypes and even for plants of the same genotype

5. Results and discussion

5.1 Genetic diversity and differentiation for Musa breeding

5.1.1 Genetic diversity

Population genetics parameters such as observed heterozygosity (Ho), expected heterozygosity (He) or gene diversity, fixation index (FIS), percentage of polymorphic loci (%Pl) and number of private alleles were estimated for the populations from the five breeding programs and ITC. The summary of these population genetics estimates for the populations from the six sources is presented in Table 3. The mean Ho value was 0.19 and it was higher than the mean value of He value of 0.17 in all populations. Ho was highest in the SwS population (0.28) and lowest in the ITC (0.09) population. A similar trend was observed for He. The highest number of private alleles (6,968) was between Mchare and Plantain while the lowest number was between the Matooke and Mchare (90). The pairwise comparison of private alleles between the populations from the six sources is presented in Supplementary Table S2.

Population genetics statistics were estimated for the diploid and tetraploid subsets using 19,678 and 6,143 SNP markers retained after applying filtering threshold on the subsets. Among the diploids, the average Ho was 0.14, with the highest Ho (0.16) among the Matooke breeding materials and lowest Ho from Prata and SwS (0.10). An opposite trend was observed for the He with the highest He for Prata and SwS (0.16) and the lowest He for Matooke (0.13). Among the tetraploids, the average Ho was 0.34, with the highest Ho

for tetraploids of SwS (0.37) and the lowest for Prata and Matooke (0.30). Plantain-derived tetraploids had the highest He (0.24) whereas the lowest He was observed in tetraploids of Prata and Matooke (0.19).

One of the measures of genetic diversity within populations is heterozygosity (Nei, 1978; Nei & Roychoudhury, 1974). Gene diversity encompasses two key components: 1) the number of alleles present and 2) the distribution (abundance or evenness) of these alleles among individuals within a population. These factors collectively contribute to the expected level of heterozygosity. However, if a population exhibits an overrepresentation of homozygous individuals carrying different alleles, this scenario results in a lower observed heterozygosity compared to what might be anticipated based on Hardy-Weinberg equilibrium (HWE).

The findings of previous research on the genetic diversity of bananas cannot be compared with ours, since each study differs in terms of the selected population, the type and number of markers employed, and data processing methods utilized. However, general patterns in genetic diversity parameters can be compared. The Ho and He, or gene diversity in this study were low, with the lowest Ho observed among accessions from the ITC. Samples from ITC selected for this study were mostly wild diploids and cultivars from the center of origin and not a comprehensive collection at ITC and this could explain their low measures of diversity as compared to breeding programs, where both gene flow and selection are causing higher Ho than in ITC. Such a low average Ho and He (0.12 and 0.13, respectively) have also been reported in banana sub-populations in southern Benin using DArTseq SNP markers (Mbo Nkoulou et al., 2023) and in *M. balbisiana* populations of Vietnam (Ho = 0.16 and He = 0.19) using microsatellite markers (Mertens et al., 2021). The low genetic variation in the studied populations indicates the limited genetic diversity in bananas and the need to identify or increase variation if genetic gain is to be realized in banana breeding (Sanchez et al., 2023). However, higher Ho than in our research has been reported in Musa germplasm diversity research (Christelová et al., 2017; Cyrille et al., 2022; Gardoce et al., 2023).

Table 3. Summary statistics of genetic diversity estimates based on single nucleotide polymorphic markers

Subset	Source	No. individuals	Ho	He	F_{IS}	%Pl
All accessions (No. Loci = 16,903; No. Acc = 845)	Prata	138	0.17	0.18	0.04	94
	SwS	148	0.28	0.21	-0.30	82
	Plantain	34	0.24	0.18	-0.30	69
	Mchare	200	0.17	0.17	0.01	99
	Matooke	106	0.20	0.15	-0.33	86
	ITC	201	0.09	0.14	0.35	97
Mean			0.19	0.17	-0.09	88
Standard deviation			0.07	0.03	0.27	11
Breeding programs (No. Loci = 12,455; No. Acc = 641)	Prata	138	0.18	0.19	0.04	97
	SwS	148	0.30	0.23	-0.30	88
	Plantain	34	0.25	0.20	-0.30	73
	Mchare	200	0.17	0.18	0.01	100
	Matooke	106	0.21	0.16	-0.33	88
Mean			0.22	0.19	-0.17	89
Standard deviation			0.05	0.03	0.18	11
Breeding programs; diploids (No. Loci = 19,678; No. Acc = 327)	Prata	70	0.10	0.16	0.40	96
	SwS	33	0.10	0.16	0.24	74
	Plantain	12	0.14	0.15	0.00	67
	Mchare	154	0.13	0.14	0.10	99
	Matooke	53	0.16	0.13	-0.18	61
Mean			0.14	0.16	0.11	79
Standard deviation			0.03	0.03	0.22	18
Breeding programs; tetraploids (No. Loci = 6,143; No. Acc = 108)	Prata	39	0.30	0.19	-0.59	77
	SwS	5	0.37	0.23	-0.56	48
	Plantain	18	0.34	0.24	-0.44	65
	Matooke	43	0.30	0.19	-0.59	77
Mean			0.34	0.26	-0.39	63
Standard deviation			0.04	0.08	0.24	17

No. Loci = Number of informative SNPs retained after filtering each subset, No. Acc = Number of accessions retained after filtering each subset, Ho = observed heterozygosity, He = expected heterozygosity or gene diversity, F_{IS} = fixation index, %Pl = percentage of polymorphic loci.

This could be because the accessions in the present study are specifically breeding materials (parents) that have been selected over time for specific key traits for use in the respective breeding programs, unlike the materials used in the above-cited research that included accessions not necessarily used as parents in these breeding programs and or accessions from other sections than *Eumusa*. Genetic diversity can also be indicated by the uniqueness of each population using an estimation of the number of private alleles, in terms of allele composition (Foulley & Ollivier, 2006; Höglund, 2009). The low number of private alleles between Matooke and Mchare could be because of the intra-hybridization evolution of both from the same or closely related *M. acuminata* wild species.

The Ho across all populations and within each population in Paper I was higher than the expected heterozygosity under the Hardy–Weinberg Equilibrium except for Prata, ITC and Mchare (where Ho = He). Higher Ho than He has also been reported by Cyrille et al. (2022) (Ho = 0.34; He = 0.29) and recently by Gardoce et al. (2023) (Ho = 0.25; He = 0.21) in a *Musa* germplasm collection in Côte d'Ivoire and the Philippines, respectively. The higher Ho than He could be due to the use of the same male diploids in the different breeding programs of IITA and selection for the different product profiles, hence non-adherence of the populations to some of the Hardy-Weinberg assumptions (Andrews, 2010), particularly lack of random mating, small population size and selection. It is known that mutations accumulate in clonal lineages and this promotes divergence between the two alleles for a specific locus, also known as the Meselson effect, thereby resulting in increased heterozygosity (Balloux *et al*., 2003). Therefore, the clonal propagation of the crop for most of the accessions (cultivated diploids) included in this study may also explain the higher Ho. The current finding is consistent with the findings of De Jesus et al. (2013) who reported higher Ho than He in cultivated versus wild diploids in an Embrapa *ex-situ* germplasm collection.

Higher Ho and He among tetraploids (0.34 versus 0.26) than among diploids (0.14 versus 0.16) and significantly lower F_{IS} within tetraploids compared to diploids, as here observed has also been reported in red clover (Osterman *et al*., 2021), cranberry (Mahy *et al*., 2000), knapweed (Hardy & Vekemans, 2001) and *Rorippa amphibia* (Brassicaceae) (Luttikhuizen *et al*., 2007), emphasizing the higher heterozygosity in tetraploids. The higher He or gene diversity within the tetraploid banana group is explained by the

higher probability of heterozygosity in tetraploids than in diploids under an outcrossing reproductive system. Hence the need for more efforts towards genetic improvement of diploids to increase diversity. Diploid parental improvement is important because *Musa* germplasm improvement continues to rely on diploid genetic resources as the source of useful variation, particularly for broadening the narrow genetic base of bananas and plantains (Ortiz & Swennen, 2014), and for the selection of promising parents in subsequent cycles of recurrent selection in breeding schemes (Ortiz, 2015).

The trend of higher Ho than He could also be explained by minimal inbreeding and this is consistent with the low and negative inbreeding coefficients for all populations as observed in Paper I. Mertens et al. (2021) also found negative F_{IS} in six populations of *M. balbisiana* using 18 SSR markers. The negative F_{IS} values observed could be attributed to the clonal propagation nature of bananas, as clonal reproduction might result in a heterozygosity excess because of low He due to a low number of genotypes present in a population while maintaining heterozygosity for a large proportion of the assessed loci (Pappert *et al*., 2000). Clonality was found to explain significant negative F_{IS} in the wild cherry population using microsatellite markers (Stoeckel *et al*., 2006).

5.1.2 Genetic differentiation

The genetic divergence between the populations from the five breeding programs (641 individuals retained after filtering) was determined by calculating the pairwise distances using Nei's standard genetic distance (DS), the pairwise genetic differentiation (FST) and analysis of molecular variance (AMOVA). The largest genetic distance was observed between SwS and Matooke breeding materials (DS = 0.067), while the lowest genetic distance (DS = 0.003) estimated was between Prata and the Mchare breeding materials (Table 4). Matooke breeding materials appeared to be consistently distant from all other germplasm sources. Among diploids, a similar trend was observed with the highest genetic distance between SwS and Matooke breeding materials (DS = 0.069) and the lowest between Prata and Mchare (DS = 0.005) (Table S3). A comparison among the tetraploids revealed the highest distance between breeding materials of SwS and Matooke (DS = 0.192) but the lowest distance between Prata and Plantain (DS = 0.042) (Table S3). The differentiation using AMOVA revealed that the total genetic

variation in the germplasm was mainly partitioned into the variation within the breeding programs and the variation among indicating that the majority of genetic variance exist within breeding programs (18.6%) rather than between them (13.5%). Genetic differentiation by FST described in detail in Paper I showed a similar trend.

Table 4. Pairwise distances among breeding programs using Nei's standard genetic distance (DS)

	Prata	SwS	Plantain	Mchare
SwS	0.020			
Plantain	0.018	0.030		
Mchare	0.003	0.029	0.018	
Matooke	0.022	0.067	0.031	0.015

The pairwise Nei's standard genetic distance (DS) between breeding programs, diploids and tetraploids was generally low. Even if DS was low, populations from SwS and Matooke breeding programs were more distant from other populations and each other. A similar low differentiation between populations was confirmed by the low estimated F_{ST} except between tetraploids where a higher differentiation was observed (Paper I). With a low DS and F_{ST} detected among the collections, the only potential reason for variation among them could be attributed to the different quality attributes that each breeding program aims to target. 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. The only significant and high genetic differentiation according to Wright (1968) and Luo et al. (2019), was observed between tetraploids from SwS and Matooke (F_{ST} =0.173). The unique alleles within each program were further demonstrated by higher contribution of variation within breeding programs than among the programs to the total variation from the AMOVA (Paper I). Higher variation within subpopulations than among subpopulations has been reported in bananas (Cyrille *et al*., 2022; Mertens *et al*., 2021), and white Guinea yam (Agre *et al*., 2021; Amponsah Adjei *et al*., 2023; Bhattacharjee *et al*., 2020), which are both clonally propagated crops and highly heterogenous; in red clover (Osterman *et al*., 2021) and *Camelina sativa* (Luo *et al*., 2019), both strictly outcrossing crops hence high gene exchange even within subpopulations,

and in inbred maize lines (Ayesiga *et al*., 2023) due to the accumulation of unique alleles from several promising lines crossed.

Genetic differentiation was lowest between populations from Prata and Mchare. This supports suggestions that Mchare bananas are the donor of unreduced gametes to many of the most widely grown and successful triploid dessert bananas such as Cavendish, Gros-Michel, Silk, and Prata (Christelová *et al*., 2017; Hippolyte *et al*., 2012; Martin *et al*., 2020; Perrier *et al*., 2011; Raboin *et al*., 2005).

5.2 Population structure in *Musa* spp. groups

The population structure of *Musa* genetic resources in this study, sub-grouped into breeding programs of diploids and tetraploids, respectively was analysed by three approaches; principal coordinate analysis (PCoA), discriminant analysis of principal components (DAPC) (reported in Paper I) and STRUCTURE. Among the breeding programs subset, the first two principal components (PCo) axes of the principal component analysis, which accounted for 51.6% of the variation among accessions, indicated that the 641 accessions could be grouped into five subpopulations, with some accessions in admixture (Paper I). Accounting for 89.4% and 51.0% of the variation among diploids and tetraploids, respectively, the first two PCo similarly suggested a grouping of five subpopulations among these subsets (Paper I).

The patterns detected using the K-means analysis method in the DAPC and using a 75% membership probability threshold (Q-value ≥ 75%) at delta K (ΔK) = 2, by STRUCTURE suggested the presence of two groups or clusters (Supplementary Figure S1). The cluster allocations for all accessions and accessions grouped according to ploidy (per diploid and tetraploid) accession from each breeding program according to STRUCTURE are summarized in Table 5 and Supplementary Table S4 and illustrated in Figure 3. Clustering of the accessions was observed to be based on genome group where K1 comprised 98.3% of accessions from the A-genome and 96.5% of accessions in K2 belong to the B-genome irrespective of their ploidy levels (Paper I). Noticeably, two *banksii* accessions (ITC0467 and ITC0806) clustered together with *balbisiana* accessions. The *Musa* germplasm in Paper

I was divided into two major subgroups, irrespective of the approach used, to infer population structure.

However, PCoA suggested the existence of five groups, which could mean that a genetic structure already exists or a common ancestor for accessions from different breeding programs (Luo *et al*., 2019; Meirmans, 2015).

Table 5. Summary of proportion of 641 accessions per breeding program based on STRUCTURE at ΔK = 2

		KI (He= 0.13; F_{ST}= 0.59)		K2 (He= 0.10; F_{ST}= 0.80)		Admixed	
	No. Ind.	Count	%	Count	%	Count	%
Prata	142	92	22.5	18	12.6	32	20.8
SwS	150	31	7.6	93	65.1	28	50.9
Plantain	35	19	4.7	01	0.1	15	9.0
Mchare	205	158	38.7	30	21.2	17	18.4
Matooke	109	108	26.5	01	0.1	00	0.9
Total	641	408	63.7	143	22.3	90	14.0

No. Ind. = Number of accessions per breeding program

Likewise, previous studies on Musa germplasms have also reported the presence of two clusters (De Jesus *et al*., 2013; Doloiras-Laraño *et al*., 2018; Gardoce *et al*., 2023; Onyango *et al*., 2010), which suggests that our collection was representative, to some extent, of the diversity in this crop. The representativeness of the available diversity in this germplasm is highly informative and very useful for its use in breeding for improved banana hybrids. The distribution of accessions in the same clusters irrespective of their source highlights relatedness between accessions from different regions, which can be attributed to the transfer and exchange of breeding material through breeding programs and the use of common diploids from the oldest banana breeding program from Honduras, by all the breeding programs.

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) and/or structural variations between the A and B genomes (Baurens *et al*., 2018), could account for the clustering of the accessions according to

genome group. Complex interspecific chromosome mosaic patterns in the ancestral groups that contributed to the cultivated banana genomes have also been suggested as the possible reasons for the genome clustering (Martin *et al*., 2020). This has also been observed and reported in other research in *Musa* (De Jesus *et al*., 2013; Doloiras-Larano *et al*., 2018; Gardoce *et al*., 2023; Onyango *et al*., 2010).

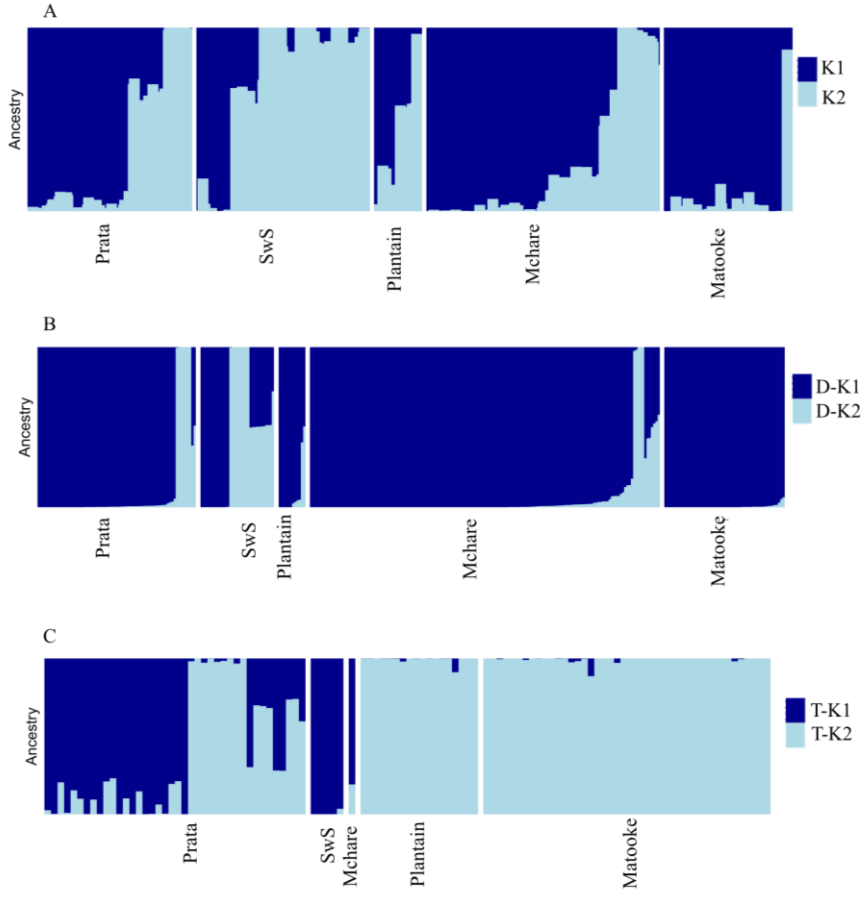


Figure 4. Population structure in the *Musa* germplasm based on single nucleotide polymorphisms. (A) 641 *Musa* accessions grouped by breeding program (B) 327 diploid and (C) 108 tetraploid *Musa* accessions.

5.3 Improving the efficiency of diploid banana 'Mchare' hybrid evaluation

5.3.1 Estimation of variation for key traits

Variance is often a measure of the extent of differences among individuals and is always related to a particular effect that has an impact on observations. In both Paper II and III, estimates of variance components for key traits, obtained from 12 and 14 full-sib early evaluation trials (EETs), are presented in Table 3 in Paper II and Table 4, Supplementary Tables S2 and S3 in Paper III. Both studies revealed significant ($p \leq 0.05$) and highly significant ($p \leq 0.001$) additive variation for all yield-related traits (see fruit yield Table 2), all agronomic traits except NFL in Paper II and NSF, PG, and NLTS in Paper III. The results from both these studies, obtained by the clonal model using the pedigree-based BLUP method indicate the existence of sufficient genetic variation and show that the strongest genetic variance component for most traits especially yield-related traits in the diploid Mchare breeding population is primarily the additive variance component. This implies that genetic improvement of diploid Mchare banana for yield is achievable as the availability and deciphering of the genetic variability among progeny populations is the basic premise for selection (Balzarini, 2000; Malikouski *et al*., 2021). Available genetic variation is expected to increase genetic gains in populations (Batte *et al*., 2020; Stringer *et al*., 2011; Zhou & Mokwele, 2016).

Even though seasonal variations in planting and flowering time were very small and non-significant for yield-related traits in Paper II, they were highly significant for most yield-related traits except for BW and NH by both the P-BLUP and the H-BLUP methods in Paper III. This indicates the environmental effect of seasons on these traits even though trials were irrigated year-round. In Paper II, we analysed the effect of the interaction of the genotypes with the cycle (Genotype:Cycle) and the interaction of the cycle with the year and season of flowering (Cycle:YearF:SeasonF). Results showed there was no significant genotype:cycle interaction for the traits evaluated, except for PH ($p \leq 0.001$) an indication that these traits are less influenced by cycle and are under genetic control. However, there was a significant ($p < 0.01$) source of variation for BW, NH and FC, among the yield-related traits, and for PG ($p < 0.001$) and PS ($p < 0.05$) among the

agronomic traits for the cycle:year:season interaction. The significant influence on these traits can be explained by seasonal climatic variations and the perennial nature of bananas with multiple production cycles on a single mat and or overlapping nature of these cycles in the same trial. Whether a plant is in the plant crop (cycle 1), first ratoon (cycle 2) or second ratoon crop (cycle 3) stage is determined by the time in season when that plant has flowered. This happens at different random times for the different genotypes and even for plants of the same genotypes and might explain the significant cycle:year:season interaction reported in this thesis.

5.3.2 Partitioning the genetic variance

The total genetic variance in a biological trait can be partitioned into additive and non-additive (dominance and epistatic) variances (Falconer & Mackay, 1996). The lack of well-designed experiments with a full-sib family structure and, a limited number of families and clones within a family can make it difficult to partition total genetic variance into these components in banana breeding. Particularly in complex traits that are governed by many genes with small effects and their interactions, partitioning the genetic variation into additive, dominance, and epistasis components helps prevent inflation and confounded estimates of the genetic parameters, such as narrow-sense heritability and breeding values (Falconer & Mackay, 1996; Gibson, 2012; Goddard, 2009; Henderson, 1984; Mrode, 2005; Muñoz *et al*., 2014; Wilson *et al*., 2010).

In Papers II and III, the clonal model was successful in partitioning and estimating the additive, dominance and epistasis variance. The likelihood ratio test was used to test if these effects were significant for the traits of interest. In the analysis in Paper II that used only the pedigree-based BLUP methods, the additive variance was significant ($p \leq 0.05$) and highly significant ($p \leq 0.001$) for all yield-related traits. A similar observation was made and reported in Paper III where all three models (P-BLUP, G-BLUP and H-BLUP) revealed significant ($p \leq 0.05$) and highly significant ($p \leq 0.001$) additive effects for all yield-related traits. The P-BLUP method did not detect significant dominance and epistasis variances for all traits in Paper II and for all traits except for NF ($p \leq 0.05$) and FW ($p \leq 0.001$) in Paper III. The application of the H-BLUP method in Paper III revealed highly significant ($p \leq 0.001$) dominance effects for NH, NF, FL and FW and highly

significant ($p \leq 0.001$) epistatic variance for NF, among yield-related traits (Table 4). In both studies, the additive variance component was the strongest genetic variance component for yield-related traits.

For bananas and other crops that are difficult to propagate and many clones are needed to capture all the genetic variance and enable selection within families or among clones (Baltunis *et al*., 2009; Isik *et al*., 2003), the large magnitude of additive genetic variation could save time and resources needed to produce many clones and replicates since focusing on parental improvement and selection of the best parents is feasible.

5.3.3 Estimation of genetic parameters

Estimation of additive genetic and non-additive genetic variances contributes to a better understanding of the genetic mechanism and architecture of traits of interest, which is important for designing breeding strategies by breeders. Secondly, estimates of genetic and phenotypic variances and covariance's are essential for the prediction of breeding values and the prediction of the expected genetic response of selection programs. Parameters that are of interest and were considered in this thesis are GCA and SCA, breeding values (as indicated by BLUP values), heritability (both broad- and narrow-sense), reliability and prediction accuracy of the estimated breeding values (EBVs) by the three models (Paper III). The GCA was higher than the SCA for all traits in Paper II, like an observation that was reported in plantain-banana hybrids for yield traits (Oselebe & Tenkouano, 2008; Tenkouano *et al*., 1998). This would imply that differences in the additive effects of the parents account for most of the variances among the progeny and that investigating intra- and inter-locus interaction is not as critical to the population's improvement. The breeding strategy for improving the Mchare diploid banana should therefore focus on parental improvement through the generation of elite diploid and tetraploid hybrids to be used as parents and secondary triploids. This is possible by cultivar development and selection through the accumulation of favourable alleles as recommended for the 4x – 2x and/or 3x – 2x breeding scheme to maximise genetic gain (Ortiz, 2013; Ortiz & Swennen, 2014).

Since BLUPs for the parents are twice their GCA and twice the breeding values of the clones, they have been used in this thesis to discuss the performance of the parents for backward selection and clones for forward

selection and hybrid evaluations. In EET 1 (Paper II), generally, clones had a higher range of BLUP values than their female parents for BW and yield per year. Two parents Akondro Mainty (female) and Guyod (male) were among the ten best-performing genotypes according to their BLUPs for BW, while the eight were hybrids (Figure 5a). These eight clones also were among the best performers for yield per year (Figure 5b). Among these best-performing clones, six were from crosses involving 'Huti-White' as the female parent. The better performance of hybrids from crosses involving 'Huti-White' was also observed and reported in Paper III, and 'Huti-White' was identified as the best-performing female parent while 'CV Rose' was the best-performing male parent for the yield-related traits, by the three methods.

The success of many yield-improvement programs hinges on the accurate identification and selection of parents from which to make crosses with complementary genotypes that can lead to the higher accumulation of additive genes (Stringer *et al*., 2011; Zhou & Mokwele, 2016). This is expected to generate superior breeding populations, and the selection of high-yielding genotypes from early segregating generations (Nyine *et al*., 2018) leading to an increase in genetic gains in such breeding populations.

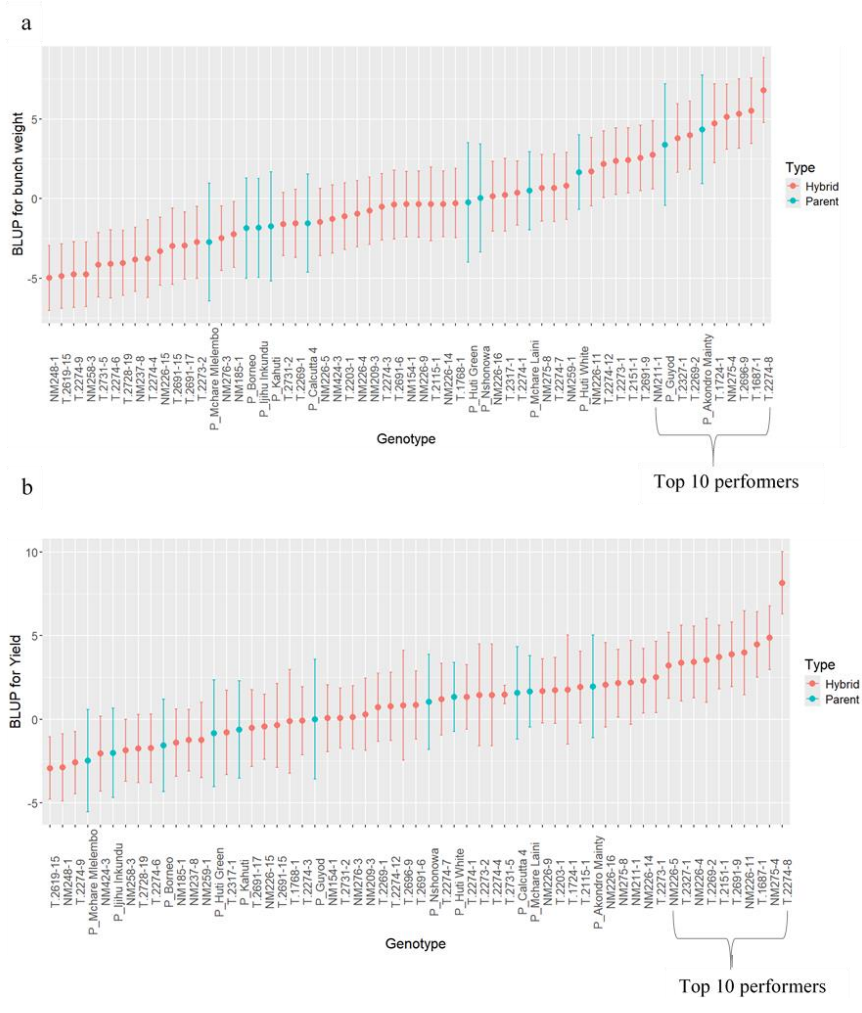


Figure 5. Ranking of genotypes according to their Best Linear Unbiased Predictor (BLUP) values for bunch weight (a) and yield per year (b).

5.3.4 Correlation among key traits and between production cycles

The results and discussions for the additive genetic correlation in this thesis are based on results from the P-BLUP (Paper II) and the H-BLUP method (Paper III). The results indicated the differential magnitude of correlation coefficients between different pairs of key traits ranging from 0.01 to 0.94 (Table 5 in Paper II) and 0.00 to 0.99 (Table 6 in Paper III). In both studies, the highest positive correlations were observed among the yield-related traits with a noticeable positive and significant correlation between BW and all fruit traits (FC, FL, FW, NF and NH) as well as between yield per year and the other yield-related traits. Similar observations were reported in other banana studies (Nyine *et al*., 2017; Uwimana *et al*., 2020). This means that selection for large fruits will also lead to heavy bunches and indirect selection for yield is possible using yield components as has been suggested for plantain germplasm (Tenkouano *et al*., 2002). Agronomic and yield-related traits had a weak correlation except for a positive and significant correlation between NH and PH (r= 0.73 in Paper II; r= 0.64 in Paper III), and between NH and PG (r= 0.64 in Paper II; r= 0.54 in Paper III). The correlation between PH and BW was predominantly weak and positive in both studies indicating an opportunity for selection of dwarf Mchare hybrids with heavy bunches. The correlation between cycling traits with yield-related traits was low and weak except for a high positive and significant correlation between DPF and NH in Paper II (r= 0.68), which greatly dropped to 0.13 in Paper III.

Breeders can evaluate the selective response and estimate or obtain indirect gains in other variables by using estimates of the genetic correlation between traits, and benefit from the multi-trait selection, a feature that can improve the efficiency of a breeding program. Another advantage of estimating the genetic correlation between traits is that the amount of phenotyping in early stages can be reduced especially for traits strongly affected by the environment as such traits can be indirectly selected from other variables measured more easily and accurately.

The correlation between production cycle one (plant crop) and cycle two (first ratoon crop) for all yield-related traits was significant, positive and of high magnitude ($0.62 > r < 0.81$, $p \leq 0.001$; Table 6).

Table 6. Correlation of yield traits between cycle 1 (plant crop) and cycle 2 (first ratoon crop)

Trait category	Trait	Correlation
Yield	Bunch weight	0.68***
	Number of hands	0.62***
	Fruit circumference	0.81***
	Fruit length	0.73**
	Fruit weight	0.72***

, * indicate significant at $p \leq 0.01$ and $p \leq 0.001$

This indicates that for yield-component traits such as BW, NH, FC, FL and FW, a hybrid would continue to perform as well as or better in the first ratoon crop if it has demonstrated good performance in the plant crop as illustrated in Figure 6. A similar high correlation between cycle 1 and cycle 2 for BW ($r = 0.56$) was initially reported by Swennen and De Langhe (1985) in False Horn plantain in Nigeria.

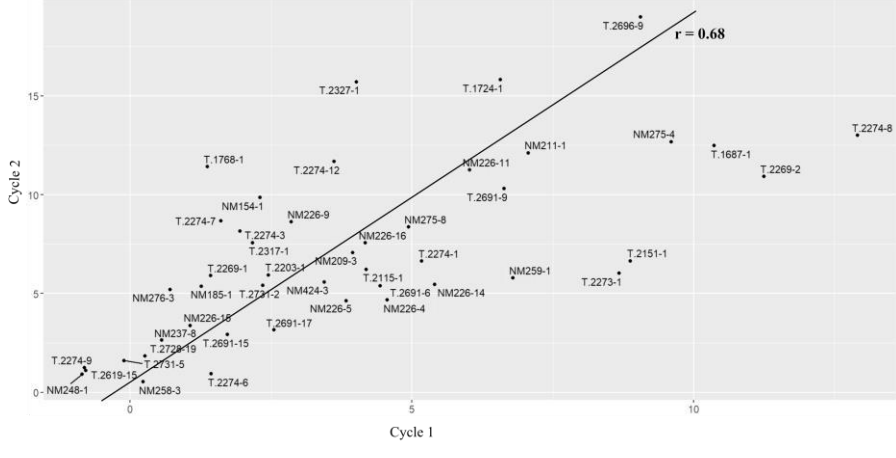


Figure 6. Bunch weight best linear unbiased prediction (BLUP) values for Mchare hybrids in cycle 1 and cycle 2.

In current practice, maximal yields in bananas are evaluated and selection is done using data from the first two cycles and sometimes up to cycle 3 (Batte *et al*., 2019; Madalla *et al*., 2022b; Tushemereirwe *et al*., 2015). This makes the already long breeding cycle even longer. The observed significant

and positive correlations of high magnitude (all r > 60), for yield trait in this study suggest that selection of the highest yielding hybrids is possible using data from the plant crop (cycle 1), thus reducing the banana breeding cycle by at least two years and boosting genetic gains.

5.3.5 Genetic relationship modelling improves efficiency in diploid banana breeding

In Paper III, we investigated how replacing the pedigree-based additive matrix (P-BLUP) with a genetic relationship matrix (G-BLUP) or a hybrid genomic relationship matrix (H-BLUP) in a linear mixed model (LMM) affects estimates of heritability and genetic effects by improving the partitioning of the genetic variance for traits of interest, into additive and non-additive variance.

Additive effects for all yield-related traits, PH, PS and PHTS agronomic traits were significant ($p \leq 0.05$) and highly significant ($p \leq 0.001$) by all three models. A larger magnitude of dominance variance was estimated by the G-BLUP and H-BLUP models than the P-BLUP models, while except for PHTS, NH and NF, the epistatic variance components estimated by all three models were close to zero for all traits. Despite all three models demonstrating the additive variance component as the strongest genetic variance component for most of the yield-related traits, the P-BLUP model estimated a larger additive variance component, than the H-BLUP and G-BLUP models as illustrated in Figure 1 in Paper III. The BLUP values were highest under the P-BLUP method, followed by the H-BLUP and least by the G-BLUP for all traits (Supplementary Tables S3 and S4 in Paper III). The lower standard errors for variance components by the H-BLUP model illustrate the more accurate precision of this model in estimating genetic parameters of traits under study.

Table 5 in paper III shows the heritability estimates, reliabilities and prediction accuracies obtained from the P-BLUP, G-BLUP and H-BLUP models. The H-BLUP model had the highest range of H^2 (0.00 – 0.95), followed by the P-BLUP model (0.00 – 0.68) and the lowest (0.00 – 0.51) by the G-BLUP for all traits studied.

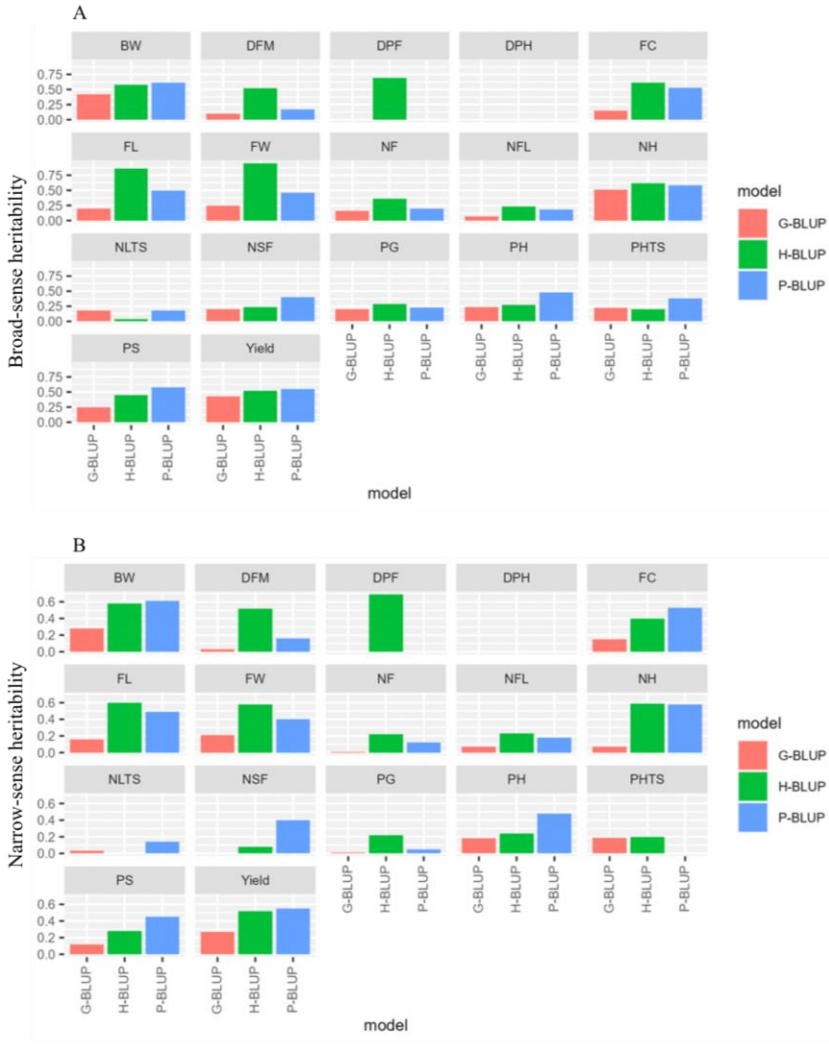


Figure 7. Broad-sense heritability (H^2) (A) and narrow-sense heritability (h^2) (B) by the P-BLUP, G-BLUP and H-BLUP models for all traits.

P-BLUP= pedigree-based best linear unbiased prediction, G-BLUP= genomic-based best linear unbiased prediction , H-BLUP= hybrid-based best linear unbiased prediction, BW = bunch weight at full maturity (kg), NH = number of hands, NF = number of fruits, FC = fruit circumference (mm), FL = fruit length (mm), FW = fruit finger weight (grams), Yield = yield/year/plant (kg), NSF = total number of suckers at flowering, NFL = number of standing functional leaves at flowering, PH = plant height at flowering (cm), PG = plant girth at 100 cm from soil surface (cm), PS = plant stature, PHTS= plant height of tallest sucker, NLTS= number of standing functional leaves of tallest sucker, DPF = days to flowering, DFM = days to fruit maturity and DPH = days from planting to harvest.

While the H-BLUP model estimated $h^2 > 0.50$ for most (five of the seven) yield traits, the P-BLUP model estimated $h^2 > 0.50$ for only three yield-related traits and the G-BLUP model had no $h^2 > 0.50$ for all yield traits. The trend of these parameters under the different models is illustrated in Figures 6A and B. The high (≥0.50) heritability estimates for most yield-related traits in this study, imply the possibility of their genetic improvement through direct selection (Ortiz, 2013; Tenkouano *et al*., 2011).

Linear mixed models that utilized genomic relationship matrices (G- and H-matrix) estimated values for additive variance and heritability with less bias compared to the model utilizing only the pedigree information (P-BLUP). The reduced bias could be because of a more accurate allocation of the genetic variance to dominance and epistatic effects, when these effects exist. This reduces the estimated additive variance and subsequent narrow-sense heritability. Furthermore, sometimes due to wrong pedigree records, the A matrix frequently makes approximations that can result in significant reconstruction errors of genetic relationships, which can affect genetic parameter estimations (Misztal *et al*., 2011; Muñoz *et al*., 2014a; Munoz *et al*., 2014b).

In addition, if the pedigree information of the parents used in the crosses is unknown as was the case in this study, the phenotype data of the parents is not included in the analysis when using the P-BLUP method. However, when using the genomic relationship matrices, the phenotype data of the parents is included since the pedigree relationship of the parents can be determined using the genetic markers. This increases the amount of data points and the establishment of genetic relationships between individuals, which improves the estimates for the variances and genetic parameters. This estimation is even further improved by the H-BLUP method that provides information for individuals that were not genotyped (Gezan *et al*., 2022). Better performance of genomic relationship methods compared to pedigree-based approaches has been reported in other studies (Festa & Whetten, 2021; Munoz *et al*., 2014b; Nazarian & Gezan, 2015; Norman *et al*., 2023). Creation of a hybrid H-matrix by combining the pedigree information from the A-matrix and the genetic information from the G-matrix, and using it in the LMM, significantly improved the accuracy of estimating the genetic parameters in our study.

As seen in Figure 7, the prediction accuracy, ρ, increased with an increase in heritability under all the methods, the H-BLUP model had higher

narrow-sense heritability estimates and prediction accuracy than estimates of the other two models for yield-related traits. According to Lee et al. (2010), the marker-based relationship matrix enhances the capacity to estimate narrow-sense heritability and breeding values of progeny may be predicted more accurately, which leads to a more accurate estimation of genetic gains (Festa & Whetten, 2021). The H-BLUP method was able to give reliabilities (ρ^2) and prediction accuracies (ρ) estimates for traits NH, NF, NSF, NFL, PG, PS, NLTS and all cycling traits that could not be estimated by using the G-BLUP method. The H-BLUP method had the highest prediction accuracies for yield-related traits and two cycling traits (DPF and DFM) while the P-BLUP model had the highest prediction accuracies for agronomic traits compared to the other two models.

Figure 8. Narrow-sense heritability (h^2) and prediction accuracy (ρ) estimates by the three models for the traits investigated.

P-BLUP= pedigree-based best linear unbiased prediction, G-BLUP= genomic-based best linear unbiased prediction , H-BLUP= hybrid-based best linear unbiased prediction, BW = bunch weight at full maturity (kg), NH = number of hands, NF = number of fruits, FC = fruit circumference (mm), FL = fruit length (mm), FW = fruit finger weight (grams), Yield = yield/year/plant (kg), NSF = total number of suckers at flowering, NFL = number of standing functional leaves at flowering, PH = plant height at flowering (cm), PG = plant girth at 100 cm from soil surface (cm), PS = plant stature, PHTS= plant height of tallest sucker, NLTS= number of standing functional leaves of tallest sucker, DPF = days to flowering, DFM = days to fruit maturity and DPH = days from planting to harvest.

6. Conclusions and future perspectives

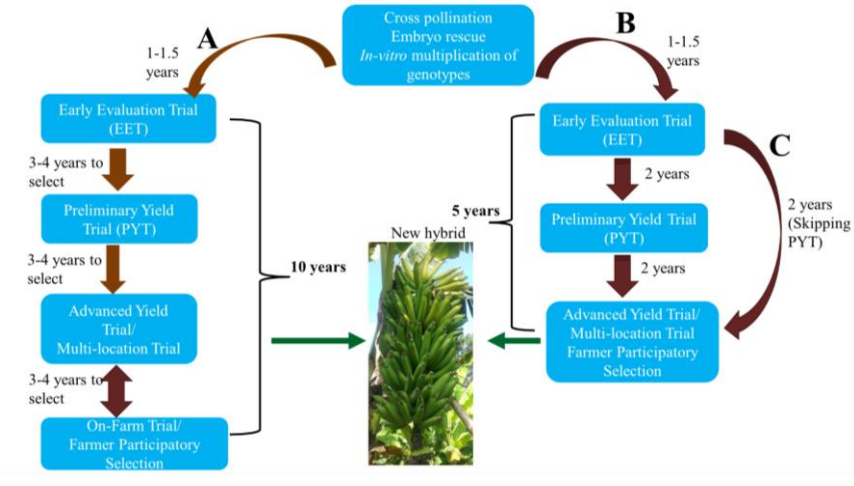


Figure 9. Approaches to hybrid selection in a Mchare banana breeding program. (A) the classical phenotypic evaluation and selection of hybrids, (B) phenotypic evaluation and selection with only cycle 1 data and (C) phenotypic evaluation and selection with only cycle 1 data and skipping the PYT stage.

The purpose of this thesis was to document the genetic diversity and population structure of *Musa* breeding germplasm available in five of the seven global banana breeding programs; and how to increase selection accuracy and improve genetic gains by the optimization of the breeding pipeline of the diploid Mchare banana breeding program of Tanzania. The significance and achievements of this thesis towards these goals can be highlighted using the breeder's equation estimating genetic gain over time (ΔG) (equation 1)

The **genetic variance** ($\boldsymbol{\sigma_A}$) of the germplasm used in the breeding programs even though found to be low, the use of unrelated parents from other sources as revealed and recommended in Paper I of this thesis, in crossing schemes would increase genetic variation for new cultivars, thereby increasing genetic gain in *Musa* breeding. In Papers II and III, the analyses revealed the existence of significant variability among the evaluated diploid Mchare hybrids and this variation is mainly additive; which demonstrates the possibility of obtaining superior genotypes and genetic gains through direct selection, especially for yield-related traits. In addition, the results of this thesis are the recommendation that the breeding strategy for improvement of Mchare diploid banana should focus on the development and selection of cultivars, by the accumulation of favourable alleles for maximum genetic gain.

The improved **prediction accuracy** (ρ) of > 0.5 for all yield-related traits was possible using genetic markers that allowed the construction of a hybrid genetic relationship matrix and hence the use of the H-BLUP method. This paved the way to model statistical analysis appropriately leading to improved prediction accuracy of breeding values of progeny, which contributes to a more accurate selection of hybrids and estimation of genetic gains that a breeder can anticipate.

The **time (*t*)** is the length of the selection cycle. The stages of a genetic improvement program for bananas are time-consuming due to the prolonged production cycles, which require time and resources. The results of this thesis allow the optimization of some of the steps. For example, the positive and significant genetic correlations among traits of interest identified in this study allow for simultaneous improvement and selection of multiple traits in the same breeding cycle. This saves time and increases genetic gain. A high correlation between production cycles 1 and 2 for yield traits documented in Paper II implies that banana breeders can already evaluate hybrids and make breeding decisions such as selections for advanced trials using data from the plant crop (approach B in Figure 8, saving two years between the early evaluation trials and preliminary yield trial). These two years could also be reduced from subsequent evaluation trials as improved statistical analysis exhibited in this thesis allows a move directly from EET to multi-locational trials simultaneously done with farmer participatory selection, with no need

for PYT (approach C in Figure 8), saving additional two years. The banana breeder at IITA-Tanzania (Allan Brown, personal communication) has also suggested this strategy. This possibly will reduce the diploid banana breeding cycle from ten years to approximately five years as illustrated in Figure 8. The number of hybrids we may lose by implementing these measures should be investigated through simulation *in silico* modelling.

This study should be done for triploid hybrid populations from the Matooke and Plantain breeding programs that are products of tetraploid x diploid crosses and triploid x diploid crosses, to estimate genetic parameters.

References

Adeniji, T., Tenkouano, A., Ezurike, J., Ariyo, C., & Vroh-Bi, I. (2010). Value-adding post harvest processing of cooking bananas (Musa spp. AAB and ABB genome groups). *African Journal of Biotechnology*, *9*(54), 9135–9141.

Adheka, J.G., Dhed'a Djailo, B., Blomme, G., Karamura, D., Swennen, R., and De Langhe, E. (2018a). Plantain diversity in the Democratic Republic of Congo and future prospects. *Acta Horticulturae*, (1225), 261–268. https://doi.org/10.17660/ActaHortic.2018.1225.36

Adheka, J.G., Dhed'a Djailo, B., Karamura, D., Blomme, G., Swennen, R., and De Langhe, E. (2018b). The morphological diversity of plantain in the Democratic Republic of Congo. *Scientia Horticulturae*, (234), 126–133. https://doi.org/10.1016/j.scienta.2018.02.034

Adheka, J.G., Komoy Losimba, J., Tamaru, C., Sivirihauma, C., Dhed'a Djailo, B., Karamura, D., De Langhe, E., Swennen, R., and Blomme, G. 2018. Banana diversity in the Oriental provinces, northeastern Democratic Republic of Congo *Acta Horticulturae*, (1196), 255–264. https://doi.org/10.17660/ActaHortic.2018.1196.31

Agre, P. A., Dassou, A. G., Loko, L. E., Idossou, R., Dadonougbo, E., Gbaguidi, A., Mondo, J. M., Muyideen, Y., *et al*. (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

Aguilar, M. (2013). Improvement of cavendish banana cultivars through conventional breeding. *Acta Horticulturae* (986), 205–208.

Akech, V. (2023). Standard Operating Procedures (SOPs) for Banana Breeding and Data Management. Faculty of Landscape Architecture, Horticulture and Crop Production Science, Swedish University of Agricultural Sciences. Available at https://publications.slu.se/?file=publ/show&id=120844

Akech, V., Bengtsson, T., Ortiz, R., Swennen, R., Uwimana, B., Ferreira, C. F., Amah, D., Amorim, E. P., *et al*. (2024). Genetic diversity and population structure in banana (*Musa* spp.) breeding germplasm. *The Plant Genome*. https://doi.org/10.1002/tpg2.20497

Amah, D., Alamu, E., Adesokan, M., van Biljon, A., Maziya-dixon, B., Swennen, R., and Labuschagne, M. (2019a). Variability of carotenoids in a *Musa* germplasm collection and implications for provitamin A biofortification.

Food Chemistry: X, 2, 100024. http://www.sciencedirect.com/science/article/pii/S2590157519300264

Amah, D., van Biljon, A., Brown, A., Perkins-Veazie, P., Swennen, R., and Labuschagne, M. (2019b). Recent advances in banana (*Musa* spp.) biofortification to alleviate vitamin A deficiency. *Critical Reviews in Food Science and Nutrition*, 59(21), 3498–3510. https://www.ncbi.nlm.nih.gov/pubmed/29999424

Amorim, E. P., dos Santos-Serejo, J. A., Amorim, V. B. O., Ferreira, C. F., & Silva, S. O. (2011a). *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.

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. 252–280. CRC Press

Amponsah Adjei, E., Esuma, W., Alicai, T., Bhattacharjee, R., Dramadri, I. O., Edema, R., Chamba, E. B., & Odong, T. L. (2023). Genetic diversity and population structure of Uganda's yam (*Dioscorea* spp.) genetic resource based on DArTseq. *PLoS ONE*, *18*(2). https://doi.org/10.1371/journal.pone.0277537

Andrews, C. (2010). The hardy-Weinberg principle. *Nature Education Knowledge*, *3*(10), 65. https://www.nature.com/scitable/knowledge/library/the-hardy-weinberg-principle-13235724/

Arinaitwe, I. K., Teo, C. H., Kayat, F., Tumuhimbise, R., Uwimana, B., Kubiriba, J., Swennen, R., Harikrishna, J. A., & Othman, R. Y. (2019). Evaluation of banana germplasm and genetic analysis of an F1 population for resistance to *Fusarium oxysporum* f. sp. *cubense* race 1. *Euphytica*, *215*, 1–11. https://doi.org/10.1007/s10681-019-2493-3

Aurore, G., Parfait, B., & Fahrasmane, L. (2009). Bananas, raw materials for making processed food products. *Trends in Food Science & Technology*, *20*(2), 78–91. https://doi.org/10.1016/j.tifs.2008.10.003

Ayesiga, S. B., Rubaihayo, P., Oloka, B. M., Dramadri, I. O., Edema, R., & Sserumaga, J. P. (2023). Genetic Variation among tropical maize inbred lines from NARS and CGIAR breeding programs. *Plant Molecular Biology Reporter*, *41*(2), 209–217. https://doi.org/10.1007/s11105-022-01358-2

Bakry, F., Carreel, F., Jenny, C., & Horry, J.-P. (2009). Genetic improvement of banana. In *Breeding plantation tree crops: tropical species*. 3–50. Springer. https://doi.org/10.1007/978-0-387-71201-7_1

Balloux, F., Lehmann, L., & de Meeûs, T. (2003). The population genetics of clonal and partially clonal diploids. *Genetics*, *164*(4), 1635–1644. https://doi.org/10.1093/genetics/164.4.1635

Baltunis, B. S., Wu, H. X., Dungey, H. S., Mullin, T. J. T., & Brawner, J. T. (2009). Comparisons of genetic parameters and clonal value predictions from clonal trials and seedling base population trials of radiata pine. *Tree Genetics & Genomes*, *5*(1), 269–278. https://doi.org/10.1007/s11295-008-0172-y

Balzarini, M. G. (2000). Biometrical models for predicting future performance in plant breeding. *PhD Thesis, Louisiana State University, US*. Available at https://doi.org/10.31390/gradschool_disstheses.7178

Barekye, A., Tongoona, P., Derera, J., Laing, M., & Tushemereirwe, W. (2009). Variation in diploid (AA) and tetraploid (AAAA) banana populations for black Sigatoka resistance and agronomic traits. 9th African Crop Science, Conference Proceedings, Cape Town, South Africa, 28 September-2 October 2009,

Batte, M., Nyine, M., Uwimana, B., Swennen, R., Akech, V., Brown, A., Hovmalm, H. P., Geleta, M., & Ortiz, R. (2020). Significant progressive heterobeltiosis in banana crossbreeding. *BMC Plant Biology*, *20*, 1–12. https://doi.org/10.1186/s12870-020-02667-y

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., *et al*. (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

Bayo, S. J., Massawe, V., Ndakidemi, P. A., Venkataramana, P., Mlaki, A., Mduma, H., Jomanga, K., Swennen, R., & Brown, A. (2024). Pollen amount and viability in Mchare and selected wild (AA) banana (*Musa acuminata*) genotypes: Prospects for breeding. *Hortscience*, *59*(5), 632–638. https://doi.org/10.21273/hortsci17608-23

Belser, C., Baurens, F.-C., Noel, B., Martin, G., Cruaud, C., Istace, B., Yahiaoui, N., Labadie, K., *et al*. (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

Biruma, M., Pillay, M., Tripathi, L., Blomme, G., Abele, S., Mwangi, M., Bandyopadhyay, R., Muchunguzi, *et al*. (2007). Banana *Xanthomonas wilt*:

a review of the disease, management strategies and future research directions. *African Journal of Biotechnology*, *6*(8).

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

BMGF (2014). *Multi crop value chain phase II Tanzania/Uganda cooking banana*. Bill & Melinda Gates Foundation (BMGF), Seattle, Washington. Available at: https://gatesopenresearch.org/documents/3-716 [2023-08-20]

Boloy, F. N., Nkosi, B. I., Losimba, J. K., Bungamuzi, C. L., Siwako, H. M., Balowe, F. W., Lohaka, J. W., Dheda Djailo, B., *et al*. (2014). Assessing incidence, development and distribution of banana bunchy top disease across the main plantain and banana growing regions of the Democratic Republic of Congo. *African Journal of Agricultural Research*. https://doi.org/10.5897/ajar2014.8751

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

Calberto, G., Staver, C., & Siles, P. (2015). An assessment of global banana production and suitability under climate change scenarios. *Climate change and food systems: global assessments and implications for food security and trade. Rome: Food Agriculture Organization of the United Nations (FAO)*, 265-291.

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*. Oxford University Press.

Christelová, P., De Langhe, E., Hřibová, E., Čížková, J., Sardos, J., Hušáková, M., Van den houwe, I., Sutanto, A., *et al*. (2017). Molecular and cytological characterization of the global Musa germplasm collection provides insights into the treasure of banana diversity [Article]. *Biodiversity and Conservation*, *26*(4), 801–824. https://doi.org/10.1007/s10531-016-1273-9

Čížková, J., Hřibová, E., Christelová, P., Van Den Houwe, I., Häkkinen, M., Roux, N., Swennen, R., & Doležel, J. (2015). Molecular and Cytogenetic Characterization of Wild *Musa* Species. *PLoS ONE*, *10*(8), e0134096. https://doi.org/10.1371/journal.pone.0134096

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., Hagen, T., Quinn, M., & Ng, E. H. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theoretical and Applied Genetics*, *132*, 627–645. https://doi.org/10.1007/s00122-019-03317-0

Cyrille, K. K. G., N'da Désiré P., Siaka T., Laetitia K. D., Olivier A. G. O, *et al*. (2022). Genetic Diversity of Plantain (*Musa* spp.) collection in Côte

d'Ivoire revealed by Single Nucleotide Polymorphism (SNP) markers. *Advances in Crop Science Technology*, *10*(495), 2.

Daniells, J., C. Jenny, D. Karamura and K. Tomekpe (2001). Musalogue: A catalogue of *Musa* germplasm. Diversity in the genus *Musa*. (E. Arnaud and S. Sharrock, Compil.), 1–213. INIBAP, Montpellier, France.

De Buck, S., & Swennen, R. (2016). Bananas, the green gold of the South. VIB, Fact series. Available at https://lirias.kuleuven.be/1718669&lang=en

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 [Article]. *BMC Plant Biology*, *13*(1), Article 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*, 361–370. Routledge.

De Langhe, E. (2009). Relevance of banana seeds in archaeology. *Ethnobotany Research and Applications*, *7*, 271–281. https://doi.org/10.17348/era.7.0.271-281

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

De Langhe, E., Karamura, D., & Mbwana, A. (2001). Tanzania Musa Expedition 2001.

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

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

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.

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.

Molecular ecology, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294x.2005.02553.x

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* 1–18. https://doi.org/https://doi.org/10.1002/9781119528265.ch1

Falconer, D. S., & Mackay, T. F. (1996). Introduction to Quantitative Genetics. Longman London, UK.

FAOSTAT. (2022). *Agriculture data*. Food and Agriculture Organization Statistics (FAOSTAT). Rome. Italy. Available at: http://faostat.fao.org Accessed 20-08-2023.

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

Fasahat P, Rajabi A., Rad J. M., Derera J. (2016). Principles and utilization of combining a.bility in plant breeding. *Biometrics & Biostatistics International Journal* 4:1–24. https://doi.org/10.15406/bbij.2016.04.00085

Festa, A. R., & Whetten, R. (2021). Simulation of Pedigree vs. fully-informative marker based relationships matrices in a Loblolly pine breeding population. Cold Spring Harbor Laboratory. https://dx.doi.org/10.1101/2021.11.30.468863

Foulley, J. L., & Ollivier, L. (2006). Estimating allelic richness and its diversity. *Livestock Science*, *101*(1-3), 150–158. https://doi.org/10.1016/j.livprodsci.2005.10.021

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/i1547-3465-07-333

Gaidashova, S.V. & Karamura, E.B. (2010). Agronomic performance of introduced banana varieties in lowlands of Rwanda. *African Crop Science Journal*, 16 (1), 9–16. https://doi.org/10.4314/acsj.v16i1.54321

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*. https://doi.org/10.1007/s00438-023-02018-0

Gezan, S. A., de Oliveira, A. A., Galli, G., & Murray, D. (2022). User's manual for ASRgenomics v. 1.1.4: An R package with complementary genomic functions. *VSN International*. https://doi.org/10.32614/cran.package.asrgenomics

Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics*, *13*(2), 135–145. https://doi.org/10.1038/nrg3118

Gilmour, A., Gogel, B., Cullis, B., & Thompson, R. (2015). ASReml user guide release 4.1 structural specification. VSN International Ltd, Hemel Hempstead, HP1 1ES, United Kingdom.

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, *136*(2), 245–257. https://doi.org/10.1007/s10709-008-9308-0

Gold, C., Kagezi, G., Night, G., & Ragama, P. (2004). The effects of banana weevil, Cosmopolites sordidus, damage on highland banana growth, yield and stand duration in Uganda. *Annals of Applied Biology*, *145*(3), 263–269. https://doi.org/10.1111/j.1744-7348.2004.tb00382.x

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

Häkkinen, M. (2013). Reappraisal of sectional taxonomy in *Musa (Musaceae)*. *Taxon*, 62(4), 809–813. https://doi.org/10.12705/624.3

Hardy, O. J., & Vekemans, X. (2001). Patterns of allozyme variation in diploid and tetraploid Centaurea jacea at different spatial scales. *Evolution*, *55*(5), 943–954. https://doi.org/10.1111/j.0014-3820.2001.tb00612.x

Henderson, C. R. (1984). Applications of linear models in animal breeding (Vol. 462). University of Guelph, Guelph.

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

Höglund, J. (2009). *Evolutionary conservation genetics*. Oxford University Press. Oxford, UK.

Hřibová, E., Čížková, J., Christelová, P., Taudien, S., de Langhe, E., & Doležel, J. (2011). The ITS1-5.8S-ITS2 Sequence Region in the *Musaceae*: Structure, Diversity and Use in Molecular Phylogeny. *PLoS ONE*, 6(3). https://doi.org/10.1371/journal.pone.0017863

IITA News and Events. (2023). https://www.iita.org/news-item/iita-and-partners-announce-first-fusarium-wilt-resistant-mchare-cooking-banana-in-tanzania/

International Food Policy Research Institute (IFPRI). (2022). Projections from IFPRI's IMPACT model: Climate change and food systems. In 2022 Global Food Policy Report: Climate Change and Food Systems. Chapter 14, 146–161. https://doi.org/10.2499/9780896294257_14

Isik, F., Li, B., & Frampton, J. (2003). Estimates of additive, dominance and epistatic genetic variances from a clonally replicated test of loblolly pine. *Forest science*, *49*(1), 77–88. https://doi.org/10.1093/forestscience/49.1.77

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

Jenny, C., Sachter-Smith, G., Breton. C., Rivallan, R., Jacquemoud-Collet, J. P., Dubois, C., Chabannes, M., Lý, N. S., *et al*. (2024). *Musa* species in mainland Southeast Asia: Frm wild to domesticate. *PLoS ONE* 19(10). https://doi.org/10.1371/journal

Kamira, M., Crichton, R. J., Kanyaruguru, J. P., Van Asten, P. J. A., Blomme, G., Lorenzen, J., Njukwe, E., *et al*. (2013). Agronomic evaluation of common and improved dessert banana cultivars at different altitudes across Burundi. In banana systems in the Humid Highlands of Sub-Saharan Africa: Enhancing resilience and productivity (eds Blomme, G., Van Asten P. & Vanlauwe, B). 37–47.
CAB International 2013. https://doi.org/10.1079/9781780642314.0037

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

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., *et al*. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms, 67–89. *Humana Press*. https://doi.org/10.1007/978-1-61779-870-2_5

Kilimo, T. (2012). Analysis of the Banana Value Chains in Tanzania and Uganda. Consumption, Productivity and Challenges (VCA report in improving the banana value chain in the EAC Region., Issue. K. Trust. https://kilimotrust.org/analysis-of-the-banana-value-chains-in-uganda-and-tanzania-2/

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

Lee, S. H., Goddard, M. E., Visscher, P. M., & Van Der Werf, J. H. (2010). Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. *Genetics Selection Evolution*, *42*, 1–14. https://doi.org/10.1186/1297-9686-42-22

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.

Lorenzen, J., A. Tenkouano, R. Bandyopadhyay, B. Vroh, D. Coyne, and L. Tripathi (2010). Overview of banana and plantain (*Musa* spp.) improvement in Africa: past and future. Proceedings of international conference on banana and plantain in Africa. *Acta Horticulturae* 879, 595–603. https://doi.org/10.17660/actahortic.2010.879.66

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

Luttikhuizen, P. C., Stift, M., Kuperus, P., & Van Tienderen, P. H. (2007). Genetic diversity in diploid vs. tetraploid *Rorippa amphibia* (*Brassicaceae*). *Molecular ecology*, *16*(17), 3544–3553. https://doi.org/10.1111/j.1365-294X.2007.03411.x

Madalla, N. A., Massawe, C., Shimwela, M., Mbongo, D., Kindimba, G., Kubiriba, J., Arinaitwe, I., *et al*. (2022a). TARIBAN1, TARIBAN2, TARIBAN3, and TARIBAN4 'Matooke' cooking banana cultivars for the Great Lakes Region of Africa. *Hortscience*, *57*(12), 1588–1592. https://doi.org/10.21273/hortsci16854-22

Madalla, N. A., Swennen, R., Brown, A., Carpentier, S., Van den Bergh, I., Crichton, R., Marimo, P., *et al*. (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

Madalla, N. A., Swennen, R., Brown, A. F., Massawe, C., Shimwela, M., Mbongo, D., Kindimba, G., *et al*. (2022b). Yield stability of East African highland cooking banana 'Matooke' hybrids. *Journal of the American Society for Horticultural Science*, *147*(6), 334–348 https://doi.org/10.21273/jashs05246-22

Mahy, G., Bruederle, L. P., Connors, B., Van Hofwegen, M., & Vorsa, N. (2000). Allozyme evidence for genetic autopolyploidy and high genetic diversity in tetraploid cranberry, *Vaccinium oxycoccos* (*Ericaceae*). *American Journal of Botany*, *87*(12), 1882–1889. https://doi.org/10.2307/2656840

Malikouski, R. G., Peixoto, M. A., Morais, A. L. d., Elizeu, A. M., Rocha, J. R. d. A. S. d. C., Zucoloto, M., & Bhering, L. L. (2021). Repeatability coefficient estimates and optimum number of harvests in graft/rootstock combinations for 'tahiti' acid lime. *Acta Scientiarum. Agronomy*, *43*. https://doi.org/10.4025/actasciagron.v43i1.51740

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*. https://doi.org/10.5772/59421

Marimo, P., Karamura, D., Tumuhimbise, R., Shimwela, M., Bergh, I. v. d., Batte, M., Massawe, C., *et al*. (2019). Post-harvest use of banana in Uganda and

Tanzania: Product characteristics and cultivar preferences of male and female farmers. *RTB Working Paper*.

Marin, D. H., Romero, R. A., Guzman, M., & Sutton, T. B. (2003). Black Sigatoka: an increasing threat to banana cultivation. *Plant Disease*, *87*(3), 208–222. https://doi.org/10.1094/pdis.2003.87.3.208

Martin, G., Cardi, C., Sarah, G., Ricci, S., Jenny, C., Fondi, E., Perrier, X., Glaszmann, J. C., *et al*. (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

Mbida Mindzie, C., Doutrelepont, H., Vrydaghs, L., Swennen, R. L., Swennen, R. J., Beeckman, H., De Langhe, E., & De Maret, P. (2001). First archaeological evidence of banana cultivation in central Africa during the third millennium before present. *Vegetation History and Archaeobotany*, *10*, 1-6. https://doi.org/10.1007/pl00013367

Mbo Nkoulou, L. F., Tchinda Ninla, L. A., Cros, D., Martin, G., Ndiang, Z., Houegban, J., Ngalle, H. B., *et al*. (2023). Analysis of genetic diversity and agronomic variation in banana sub-populations for genomic selection under drought stress in southern Benin. *Gene*, *859*, Article 147210. https://doi.org/10.1016/j.gene.2023.147210

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

Mertens, A., Bawin, Y., Vanden Abeele, S., Kallow, S., Toan Vu, D., Thi Le, L., Dang Vu, T., *et al*. (2021). Genetic diversity and structure of *Musa balbisiana* populations in Vietnam and its implications for the conservation of banana crop wild relatives. *PLoS ONE*, 16(6). https://doi.org/10.1371/journal.pone.0253255

Misztal, I., Tsuruta, S., Aguillar, I., Legarra, A., & Lawlor, T. (2011). Approximation of genomic accuracies in single-step genomic evaluation. *Interbull Bulletin* (44). Stavanger, Norway.

Moose, S. P., & Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant physiology*, *147*(3), 969-977. https://doi.org/10.1104/pp.108.118232s

Mrode, R. A. (2005). Linear models for the prediction of animal breeding values, CABI, Wallingford, United Kingdom. 311–316. https://doi.org/10.1079/9780851990002.0311

Muñoz, P. R., Resende Jr, M. F., Gezan, S. A., Resende, M. D. V., de Los Campos, G., Kirst, M., Huber, D., & Peter, G. F. (2014a). Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics*, 198(4), 1759–1768. https://doi.org/10.1534/genetics.114.171322

Muñoz, P. R., Resende, M. F. R., Huber, D. A., Quesada, T., Resende, M. D. V., Neale, D. B., Wegrzyn, J. L., Kirst, M., & Peter, G. F. (2014b). Genomic

Relationship Matrix for Correcting Pedigree Errors in Breeding Populations: Impact on Genetic Parameters and Genomic Selection Accuracy. *Crop Science*, 54(3), 1115–1123. https://doi.org/10.2135/cropsci2012.12.0673

Nakato, G. V., Christelová, P., Were, E., Nyine, M., Coutinho, T. A., Doležel, J., Uwimana, B., Swennen, R., & Mahuku, G. (2019). Sources of resistance in Musa to *Xanthomonas campestris* pv. *musacearum*, the causal agent of banana Xanthomonas wilt. *Plant Pathology*, *68*(1), 49–59. https://doi.org/10.1111/ppa.12945

Nayar, N. (2010). The Bananas: Botany, Origin, Dispersal. *Horticultural reviews*, *36*(3), 117–164. https://doi.org/10.1002/9780470527238.ch2

Nazarian, A., & Gezan, S. A. (2015). Integrating non-additive genomic relationship matrices into the study of genetic architecture of complex traits. *Journal of Heredity*, *107*(2), 153–162.

Ndabamenye, T., Van Asten, P., Vanhoudt, N., Blomme, G., Swennen, R., Annandale, J., & Barnard, R. (2012). Ecological characteristics influence farmer selection of on-farm plant density and bunch mass of low input East African highland banana (*Musa* spp.) cropping systems. *Field Crops Research*, *135*, 126–136. https://doi.org/10.1016/j.fcr.2012.06.018

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, *89*(3), 583–590. https://doi.org/10.1093/genetics/89.3.583

Nei, M., & Roychoudhury, A. K. (1974). Sampling variances of heterozygosity and genetic distance. *Genetics*, *76*(2), 379–390. https://doi.org/10.1093/genetics/76.2.379

Norman, P. E., Asfaw, A., Agre, P. A., Danquah, A., Tongoona, P. B., Danquah, E. Y., & Asiedu, R. (2023). Molecular and phenotypic profiling of white Guinea yam (*Dioscorea rotundata*) breeding lines. *Frontiers in Horticulture*, *2*, https://doi.org/10.3389/fhort.2023.1290521

Nowakunda, K., Khakasa, E., Ceballos, H., Kenneth, A., Tumuhimbise, R., Bugaud, C., Asasira, M., *et al*. (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

Nowakunda, K., Akankwasa, K., Khakasa, E., Asasira, M., Kisakye, S., Arnaud, E., Mukasa, Y., *et al*. (2022). *Gendered Matooke product profile in Uganda* https://doi.org/10.18167/DVN1/HQX13V

Nyine, M., Uwimana, B., Blavet, N., Hřibová, E., Vanrespaille, H., Batte, M., Akech, V., *et al*. (2018). Genomic prediction in a multiploid crop: genotype by environment interaction and allele dosage effects on predictive ability in banana. *The Plant Genome*, *11*(2), https://doi.org/10.3835/plantgenome2017.10.0090

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

Ocan, D., Mukasa, H., Rubaihayo, P., Tinzaara, W., & Blomme, G. (2008). Effects of banana weevil damage on plant growth and yield of East African Musa genotypes. *Journal of Applied Biosciences*, *9*(2), 407–415.

Ochola, D., Jogo, W., Ocimati, W., Rietveld, A., Tinzaara, W., Karamura, D., & Karamura, E. (2013). Farmers' awareness and perceived benefits of agro-ecological intensification practices in banana systems in Uganda. *African Journal of Biotechnology*, *12*(29). https://doi.org/10.5897/ajb12.2868

Ocimati, W., Tazuba, A. F., Tushemereirwe, W. K., Tugume, J., Omondi, B. A., Acema, D., Were, E., *et al*. (2021). First report of banana bunchy top disease caused by Banana bunchy top virus in Uganda. *New Disease Reports*, 44(2). https://doi.org/10.1002/ndr2.12052

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*, https://doi.org/10.17660/ActaHortic.2010.879.69

Onyango, M., Karamura, D., Keeley, S., Manshardt, R., & Haymer, D. (2009). Morphological characterisation of East African AAB and AA dessert bananas (*Musa* spp.). *Acta Horticulturae*, 897, 95–105. https://doi.org/10.17660/actahortic.2011.897.9

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. (2013). Conventional banana and plantain breeding. VII international symposium on banana: ISHS-ProMusa symposium on bananas and plantains: Towards Sustainable Global Production. *Acta Horticultura*e 986:177–194. https://doi.org/10.17660/actahortic.2013.986

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*, 110–146. Chapman & Hall.

Ortiz, R., & Swennen, R. (2014). From crossbreeding to biotechnology-facilitated improvement of banana and plantain. *Biotechnology Advances*, *32*(1), 158–169. https://doi.org/10.1016/j.biotechadv.2013.09.010

Ortiz, R., & Vuylsteke, D. (1994). Future strategy for Musa improvement. *International Network for the Improvement of Banana and Plantain (INIBAP)*, Montpellier, France.

Oselebe, H., & Tenkouano, A. (2008). Combining ability for important agronomic traits in Musa diploid populations. *Journal of Agriculture, Biotechnolgy and Ecology*, 1, 93–104

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*, https://doi.org/10.3389/fpls.2021.748750

Pappert, R. A., Hamrick, J., & Donovan, L. A. (2000). Genetic variation in *Pueraria lobata* (*Fabaceae*), an introduced, clonal, invasive plant of the southeastern United States. *American Journal of Botany*, *87*(9), 1240–1245. https://doi.org/10.2307/2656716

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*(3), 545–554.

Perrier, X., De Langhe, E., Donohue, M., Lentfer, C., Vrydaghs, L., Bakry, F., Carreel, F., *et al*. (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., *et al*. (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

Pillay, M., Ogundiwin, E., Nwakanma, D., Ude, G., & Tenkouano, A. (2001). Analysis of genetic diversity and relationships in East African banana germplasm. *Theoretical and Applied Genetics*, *102*, 965–970. https://doi.org/10.1007/s001220000500

Pillay, M., & Tripathi, L. (2008). Banana Breeding. In M. S. Kang & P. M. Priyadarshan (Eds.), *Breeding Major Food Staples.* 393–428. Blackwell Publishing.

Ploetz, R., & Evans, E. (2015). Banana diseases and the future of the industry. *Horticultural. Review*, *43*, 311–351. https://doi.org/10.1002/9781119107781.ch06

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Raboin, L. M., Carreel, F., Noyer, J. L., Baurens, F. C., Horry, J. P., Bakry, F., Montcel, H. T. D., *et al*. (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

Ray, P. K. (2002). *Breeding tropical and subtropical fruits*. Springer Science & Business Media.

Ribeiro, L. R., Silva, S. d. O., Oliveira, S. A. S. d., Amorim, E. P., Serejo, J. A. S., & Haddad, F. (2018). Sources of resistance to *Fusarium oxysporum* f. sp. *cubense* in banana germplasm. *Revista Brasileira de Fruticultura*, *40*. https://doi.org/10.1590/0100-29452018202

Sardos, J., Perrier, X., Doležel, J., Hřibová, E., Christelová, P., Van den houwe, I., Kilian, A., & Roux, N. (2016). DArT whole genome profiling provides insights on the evolution and taxonomy of edible Banana (*Musa* spp.). *Annals of Botany*, 118(7), 1269–1278. https://doi.org/10.1093/aob/mcw170

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*, *120*(14). https://doi.org/10.1073/pnas.2205780119

Sansaloni, C. P., Petroli, C. D., Carling, J., Hudson, C. J., Steane, D. A., Myburg, A. A., Grattapaglia, D., *et al*. (2010). A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus. *Plant Methods*, *6*(1), 1–11. https://doi.org/10.1186/1746-4811-6-16

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 R & D in Asia and the Pacific*, *10*, 103–106.

Sharrock, S. L., Horry, J. P., & Frison, E. A. (2000). The state of the use of *Musa* diversity. *Broadening the Genetic Base of Crop Production*, 223–243. https://doi.org/10.1079/9780851994116.0223

Shimwela, M. M., Mahuku, G., Mbanzibwa, D., Mkamilo, G., Mark, D., Mosha, H. I., Pallangyyo, B., *et al*. (2022). First report of banana bunchy top virus in banana and plantain (*Musa* spp.) in Tanzania. *Plant Disease*, *106*(4), 1312. https://doi.org/10.1094/pdis-07-21-1387-pdn

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). 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

Speijer, P., Kajumba, C., & Ssango, F. (1999). East African highland banana production as influenced by nematodes and crop management in Uganda. *International journal of pest management*, *45*(1), 41–49. https://doi.org/10.1080/096708799228030

Ssebuliba, R., Rubaihayo, P., Tenkouano, A., Makumbi, D., Talengera, D., & Magambo, M. (2005). Genetic diversity among East African Highland bananas for female fertility. *African Crop Science Journal*, *13*(1), 13–26.

Stevens, B., Diels, J., Vanuytrecht, E., Brown, A., Bayo, S., Rujweka, A., Richard, E., *et al*. (2020). Canopy cover evolution, diurnal patterns and leaf area index relationships in a Mchare and Cavendish banana cultivar under different soil moisture regimes. *Scientia Horticulturae*, 272, 109328. https://doi.org/10.1016/j.scienta.2020.109328

Stoeckel, S., Grange, J., Fernández-Manjarres, J. F., Bilger, I., Frascaria-Lacoste, N., & Mariette, S. (2006). Heterozygote excess in a self-incompatible and partially clonal forest tree species—*Prunus avium* L. *Molecular ecology*, *15*(8), 2109–2118. https://doi.org/10.1111/j.1365-294x.2006.02926.x

Stringer, J., Cox, M., Atkin, F., Wei, X., & Hogarth, D. (2011). Family selection improves the efficiency and effectiveness of selecting original seedlings and parents. *Sugar Tech*, *13*(1), 36–41. https://doi.org/10.1007/s12355-011-0073-5

Swennen, R. (1990). Limits of morpho taxonomy: 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; Montpellier, France.

Swennen, R., & De Langhe, E. (1985). Growth parameters of yield of plantain (*Musa* cv. AAB). *Annals of botany*, *56*(2), 197-204.

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

Swennen, R. & Vuylsteke, D. (2001) Banana (*Musa* L.). In: Raemaekers R.H. (ed), *Crop production in tropical Africa*. pp. 530–552.

Tenkouano, A., Crouch, J., Crouch, H., & Ortiz, R. (1998). Genetic diversity, hybrid performance, and combining ability for yield in *Musa* germplasm. *Euphytica*, *102*, 281–288.

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

Tenkouano, A., Ortiz, R., & Baiyeri, K. (2002). Phenotypic and genetic correlations in *Musa* populations in Nigeria. *African Crop Science Journal*, 10(2) https://doi.org/10.4314/acsj.v10i2.27544

Tenkouano, A., Pillay, M., & Ortiz, R. (2011). Breeding techniques. In M. Pillay & A. Tenkouano (Eds.), *Banana Breeding: Progress and Challenges* (pp. 181–202). Taylor & Francis Group.

Tenkouano, A., & Swennen, R. (2004). Plantains and banana: progress in breeding and delivering improved plantain and banana to African farmers. *Chronica Horticulturae*, 44 (1), 9–15.

Tinzaara, W., Karamura, E., Kubiriba, J., Ochola, D., Ocimati, W., Blomme, G., & Ssekiwoko, F. (2016). The banana Xanthomonas wilt epidemic in east and central Africa: current research and development efforts. *Acta Horticulturae,*, *1114*, 267–274. https://doi.org/10.17660/actahortic.2016.1114.36

Tixier, P., Malézieux, E., & Dorel, M. (2004). SIMBA-POP: a cohort population model for long-term simulation of banana crop harvest. *Ecological Modelling*, *180*(2-3), 407–417. https://doi.org/10.1016/j.ecolmodel.2004.04.028

Tushemereirwe, W., Batte, M., Nyine, M., Tumuhimbise, R., Barekye, A., Ssali, T., Talengera, D., Kubiriba, *et al*. (2015). Performance of NARITA banana hybrids in the preliminary yield trial for three cycles in Uganda.

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*(1), 4. https://doi.org/10.3390/agronomy11010004

van Asten, P. J., Fermont, A., & Taulya, G. (2011). Drought is a major yield loss factor for rainfed East African highland banana. *Agricultural water management*, *98*(4), 541–552. https://doi.org/10.1016/j.agwat.2010.10.005

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Vantyghem, M., Merckx, R., Hood-Nowotny, R., Swennen, R., Heiling, M., Resch, C., Dercon, G., & Heilig, M. (2023). Evaluating phloem sap δ13C as a short-term indicator of drought stress in banana. *Acta Horticulturae* (1367), 133–140. https://doi.org/10.17660/actahortic.2023.1367.15

Varma, V., & Bebber, D. P. (2019). Climate change impacts on banana yields around the world. *Nature Climate Change*, *9*(10), 752–757. https://doi.org/10.1038/s41558-019-0559-9

Viana, J. M. S., & Matta, F. d. P. (2003). Analysis of general and specific combining abilities of popcorn populations, including selfed parents. *Genetics and Molecular Biology*, *26*, 465–471. https://doi.org/10.1590/S1415-47572003000400010

Viljoen, A., Mahuku, G., Massawe, C., Ssali, R.T., Kimunye, J., Mostert, G., Ndayihanzamaso, P. and Coyne, D.L. (2017). *Banana Diseases and Pests: Field Guide for Diagnostics and Data Collection*. International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria.

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics*, *9*(4), 255–266. https://doi.org/10.1038/nrg2322

Vuylsteke, D. (2000). Breeding bananas and plantains: from intractability to feasibility. *Acta Horticulturae*, 540, 149–156. https://doi.org/10.17660/actahortic.2000.540.16

Vuylsteke, D. (2001). Strategies for utilization of genetic variation in plantain improvement. *Published PhD Thesis, Katholieke Universiteit Leuven*.

Vuylsteke, D. R., Swennen, R. L., & Ortiz, R. (1993a). Development and performance of black sigatoka-resistant tetraploid hybrids of plantain (Musa spp., AAB group). *Euphytica*, *65*, 33–42.

Vuylsteke, D., Swennen, R., & Ortiz, R. (1993b). Registration of 14 improved tropical Musa plantain hybrids with black sigatoka resistance. *Hortscience*, *28*(9), 957–959.

Wairegi, L. W., Van Asten, P. J., Tenywa, M. M., & Bekunda, M. A. (2010). Abiotic constraints override biotic constraints in East African highland banana systems. *Field Crops Research*, *117*(1), 146–153. https://doi.org/10.1016/j.fcr.2010.02.010

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B., & Nussey, D. H. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology*, *79*(1), 13–26. https://doi.org/10.1111/j.1365-2656.2009.01639.x

Wong, C., R. Kiew, G. Argent, O. Set, S.K. Lee and Y.Y. Gan (2002). Assessment of the Validity of the Sections in *Musa (Musaceae)* using AFLP. *Annals of Botany* 90(2): 231–238. https://doi.org/10.1093/aob/mcf170

Wright, S. (1968). Evolution and the Genetics of Populations, 4 vols. In: University of Chicago Press, Chicago.

Zhou, M., & Mokwele, A. (2016). Family versus individual plant selection for stem borer (*Eldana saccharina*) resistance in early stages of sugarcane breeding in South Africa. *South African Journal of Plant and Soil*, *33*(2), 89–96. https://doi.org/10.1080/02571862.2015.1084546

Popular science summary

As the global population continues to grow so does the demand for food. To satisfy this world demand, 90% must be met by increasing crop harvests per hectare. One way to achieve this is through plant breeding to improve the crops' performance. Bananas and plantains, sometimes collectively referred to as bananas, grown in tropical and sub-tropical regions, are a critical income security and staple food crop for millions of people in Africa. For example in Uganda, the East African highland banana called 'Matooke' is synonymous with food. Despite the importance and increasing demand for bananas, their current yield is still below the potential yield due to mainly climate change, disease and pest pressure.

Breeding for improved varieties is challenging especially through cross-breeding because bananas have reduced fertility and are slow growing, requiring up to 20 years to get an improved hybrid released. This makes the banana breeding and evaluation process long, expensive and tedious, reducing the genetic gain that can be expected in banana breeding programs. Genetic gain is the progress achieved by the genetic improvement of a desirable trait in a crop from one generation to the next. Improving the efficiency of plant breeding will increase genetic gain if genetic variation and selection intensity are maximized and the time it takes from one generation to the next generation is minimized. Genetic variation is the genetic resources available and accessible in the breeding population that offer a pool of genes that can be used to improve crops. This thesis focused on documenting the genetic variation available and identifying new sources of variation that could be used in banana improvement schemes, and optimization of the breeding pipeline of the diploid Mchare breeding program of Tanzania as a blueprint for the triploid and tetraploid improvement programs.

The banana genetic resources studied include wild populations, cultivars, and improved breeding clones used in current breeding programs. A detailed study focused on diploids and tetraploids as these have the possibility of being utilized as parents across breeding programs. The genetic diversity was analysed using DNA markers. The results indicate low genetic diversity among the diploid and tetraploid clones used as parents. However, seven diploids were identified that could be used as potential sources of pests and pathogen resistance genes. The inclusion of these diploids in the breeding programs will broaden the genetic base and contribute to the increase in genetic gain in bananas.

Hybrids from crosses using Mchare cultivars as females, and wild and improved diploids as males, were used to determine the level of genetic variation, partition the genetic variation into additive and non-additive components, estimate genetic parameters and investigate if the use of information from genetic markers improves the estimates of genetic characteristics. The aim was to improve the selection efficiency for important traits in banana breeding. The results showed that there is significant additive variation and it is the strongest genetic variance for all yield-related traits. This implies that improving parents and selecting the best-performing parents to use in the breeding designs will lead to genetic gains in Mchare banana populations. This is further supported by the high values for narrow-sense heritability for these yield-related traits that were reported in this thesis. The correlations among the yield-related traits were high and positive and can therefore be enhanced simultaneously. The correlation between the plant crop and the first ratoon crop for all yield-related traits was significant, high and positive. This enhances genetic gain by reducing the breeding cycle since only data from the plant crop is sufficient to make selection decisions. The use of genomic relationships established through genetic markers significantly improves the estimates of genetic characteristics in diploid banana breeding.

The significance and achievements of this thesis are highlighted by its contribution towards documenting and identification of available genetic resources that can be exploited in banana breeding. Additionally, designing breeding schemes, evaluation and selection efficiency can be enhanced by the approaches employed and recommended in this thesis to increase genetic gains that can meet the current and future demands of bananas.

Populärvetenskaplig sammanfattning

När världens befolkning fortsätter att växa, ökar även efterfrågan på mat. För att tillgodose den globala efterfrågan måste 90% mötas genom att öka skördarna per hektar. Ett sätt att uppnå detta är att genom växtförädling förbättra grödornas prestation. Bananer och plantain (mjölbanan), ibland kollektivt kallade bananer, som odlas i tropiska och subtropiska regioner, är en kritisk inkomstkälla och basföda för miljontals människor i Afrika. Till exempel, i Uganda, är den östafrikanska höglandsbananen kallad "Matooke" synonym med mat. Trots bananernas betydelse och den ökande efterfrågan är den nuvarande avkastningen fortfarande under den potentiella avkastningen, främst på grund av klimatförändringar, sjukdomar och skadegörare.

Förädling för förbättrade sorter är utmanande, särskilt genom korsningsförädling, eftersom bananer har nedsatt fertilitet och växer långsamt, vilket kräver upp till 20 år för att få fram en förbättrad hybrid. Detta gör bananförädlingen och utvärderingsprocessen lång, dyr och mödosam, vilket minskar den genetiska vinsten som kan förväntas i bananförädlingsprogram. Genetisk vinst är det framsteg som uppnås genom genetisk förbättring av en önskvärd egenskap i en gröda från en generation till nästa. Att förbättra effektiviteten i växtförädlingen kommer att öka den genetiska vinsten om genetisk variation och urvalets intensitet maximeras och tiden mellan generationerna minimeras. Genetisk variation är de genetiska resurser som finns och är tillgängliga i förädlingspopulationen och som erbjuder en genpool som kan användas för att förbättra grödor. Denna avhandling fokuserade på att dokumentera den genetiska variation som finns tillgänglig och identifiera nya källor till variation som kan användas i bananförädlingsprogram, samt optimering av förädlingsprocessen i förädlingsprogrammet för diploid Mchare i Tanzania som en modell för förädlingsprogram för triploider och tetraploider.

De genetiska resurserna som studerades inkluderar vilda populationer, sorter och förbättrade förädlingskloner som används i nuvarande förädlingsprogram. En

detaljerad studie fokuserade på diploider och tetraploider eftersom dessa har potential att användas som föräldrar i olika förädlingsprogram. Den genetiska mångfalden analyserades med hjälp av DNA-markörer. Resultaten indikerar låg genetisk mångfald bland de diploida och tetraploida kloner som används som föräldrar. Däremot identifierades sju diploider som kan användas som potentiella källor till resistensgener mot skadegörare och patogener. Inkluderingen av dessa diploider i förädlingsprogrammen kommer att bredda den genetiska basen och bidra till att öka den genetiska vinsten i bananer.

Hybrider från korsningar med Mchare-sorter som honor och vilda och förbättrade diploider som hanar användes för att bestämma nivån av genetisk variation, dela upp den genetiska variationen i additiva och icke-additiva komponenter, uppskatta genetiska parametrar och undersöka om användningen av information från genetiska markörer förbättrar uppskattningarna av genetiska egenskaper. Målet var att förbättra selektionseffektiviteten för viktiga egenskaper i bananförädling. Resultaten visade att det finns signifikant additiv variation och att denna är den starkaste genetiska variansen för alla avkastningsrelaterade egenskaper. Detta innebär att förbättrade föräldrar och val av de bäst presterande föräldrarna att använda i förädlingsdesignen kommer att leda till genetiska vinster i Mchare-bananpopulationer. Detta stöds ytterligare av de höga värdena för arvbarhet (h^2) för dessa avkastningsrelaterade egenskaper som rapporterades i denna avhandling. Korrelationerna mellan de avkastningsrelaterade egenskaperna var höga och positiva, och kan därför förbättras samtidigt. Korrelationen mellan den första skörden och den första rotskott-skörden för alla avkastningsrelaterade egenskaper var signifikant, hög och positiv. Detta ökar den genetiska vinsten genom att förkorta förädlingscykeln eftersom det är tillräckligt med data från den första skörden för att kunna göra ett urval. Användningen av genomiska relationer fastställda genom genetiska markörer förbättrar signifikant uppskattningarna av genetiska egenskaper i diploid bananförädling.

Betydelsen och framstegen av denna avhandling betonas genom dess bidrag till dokumentation och identifiering av tillgängliga genetiska resurser som kan utnyttjas i bananförädling. Dessutom kan utformningen av förädlingsscheman, utvärdering och selektionseffektivitet förbättras genom de tillvägagångssätt som används och rekommenderas i denna avhandling för att öka de genetiska vinsterna som kan möta de nuvarande och framtida behoven av bananer.

Acknowledgements

I would like to extend my sincere gratitude to my supervisors, Prof. Rodomiro Ortiz, Assoc. Prof. Therese Bengtsson, Prof. Emeritus. Rony Swennen and Dr. Allan Brown for expertise and guidance on this journey. I feel honoured to have had the opportunity to work under the guidance of such senior scientists, yet still capable of listening and offering quick responses to my opinions, ideas and work. You all contributed significant pieces to the puzzle that now shapes my career path. Thank you for sharing your databases of knowledge with me so I could decipher complex genetics and statistics concepts. I would like to thank you all for extending grace, patience and kindness to me when I lost my mother in the middle of this PhD journey. Therese, thank you for stepping out of the academic zone to remind me to take care of my wellness and live a well-balanced life so I can stay sane.

Even if she is no longer here with us, I would like to thank my mother, Jane Nantongo, who once told me 'I was denied an education but I hope you get it all', and would have been proud of this achievement. May your beautiful soul continue to rest in peace Mama.

To my dad, the retired Col. Dan Opito, who on the very first day starting my BSc. Agriculture at Makerere University, told me "Akech, you must do well and then get a PhD". Here dad, I got the PhD.

I want to thank Dr. Marnik Vuylsteke for helping me navigate the data analysis for the second part of this thesis. Thank you for your interest in my work.

My sincere gratitude to Dr. Brigitte Uwimana as a mentor and friend. More than assisting with my research, which has been instrumental, she encourages me, challenges me, provides learning opportunities, promotes my work, and has fostered enormous growth in me as an early career academic. Thank you for teaching me how to chase my dreams and always reminding whenever I doubted myself, that I am still a good mum. It is immeasurable

how much I have learnt from Brigitte and I am extremely grateful for having worked with her during these formative years.

It was a great opportunity to be affiliated to both the International Institute of Tropical Agriculture (IITA) and the Swedish University of Agricultural Sciences (SLU). I want to thank my colleagues at work, Rosemary Simba, Miriam Awio, Noah Nyanzi for sample collection and preparation for genotyping, and Stanley Bayo, Jackie Ulotu and Hassan Mduma for phenotyping.

Many thanks to my colleagues at SLU, past and present, for your friendship that made my stay in Sweden easier. Thank you for listening to my frustrations during lunch breaks and 'fikas'. Special thanks go to Anders Ferm and Barbro Nedstam who opened up their home to me and my children, and the warm delicious dinners and conversations about Uganda that made me feel home away from home. I would like to thank Camilla Stjärnäng for your help with the administrative issues and your kindness.

To the friends before and those I made along the way, in Uganda, Sweden and beyond, I am tremendously grateful for your friendship and interest in my work. Thank you for cheering me on and making me feel like I was doing an important thing with this PhD. If I had enough space, I would mention you all by name but thank you for giving me many, many things to enjoy outside of the PhD. Oriane, Eliora, and Noella Havugimana, thank you for your sincere prayers for me, for everything that only prayers could help.

I would like to thank my brother Abas Kabuye, for being the captain of the cheerleading group cheering me on and supporting me. Having my back always, taking care of everything home when I was away and for simply being the best brother I could ever ask for. My extended family in Uganda and Kenya, the late Mr. Obita, Mathe/Shosho-Ramdas, my aunty (Nalongo) Madina, Uncle Andrew, Mama Brenda and Rev. Margaret Nakalembe thank you for mothering me when mum died, your prayers, support and encouragement. Mathe, Aunty Madina, Julie, JJ, Sharon, Phiona, Brenda, Tendo, Hamida and the 'village', thank you for taking care of the twins when I was away at school.

Last but not least, I give thanks to my family for their love, unwavering support and inspiration: Brielle and Adriel, if I am proud of this work and my achievements, that doesn't compare to how proud I am to be your mummy. I give particular thanks to Allan for being an amazing support and aid through all the tough times. Thank you for holding the fort together, for

playing the double role of daddy and mummy but yet always reminding me that I am still a good mum even when I had to do it from miles away. Thank you for being my calm when the storms hit. I only wish I could support you to pursue your dreams as you have supported me to achieve mine, so far.

I thank you all for allowing me to be everything I am capable of.

I

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.

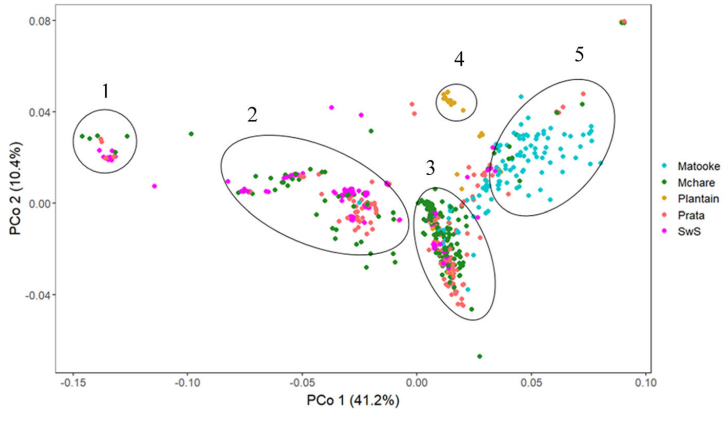


FIGURE 1 Genetic distance revealed by a principal coordinate analysis (PCoA) based on Nei's unbiased genetic distance of single nucleotide polymorphisms (SNP) among the accessions. A strong population structure was revealed by the PCoA based on diploids from each breeding program with the first and second 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

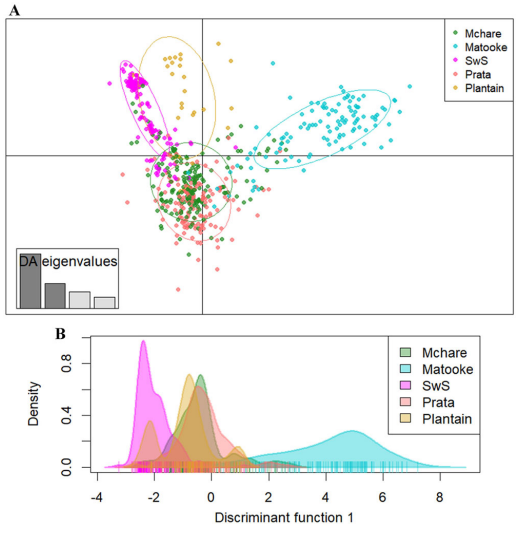


FIGURE 2 Discriminant analysis of principal components (DAPC). (A) Scatter plot of the 641 *Musa* accessions from the five breeding programs using two discriminant functions with an inset plot of discriminant analysis (DA) eigen values; (B) plot of the densities of individuals on the first discriminant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

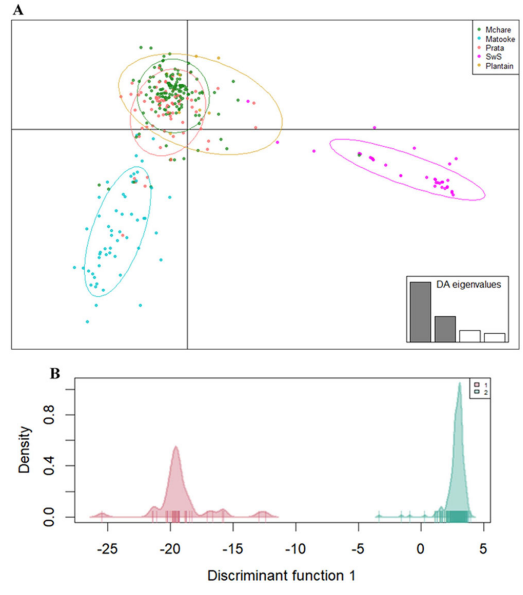


FIGURE 3 Discriminant analysis of principal components (DAPC). (A) Scatter plot of the 327 diploid *Musa* accessions from the five breeding programs using two discriminant functions with an inset plot of discrimant analysis (DA) eigen values and (B) plot of the densities of individuals on the first discriminant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

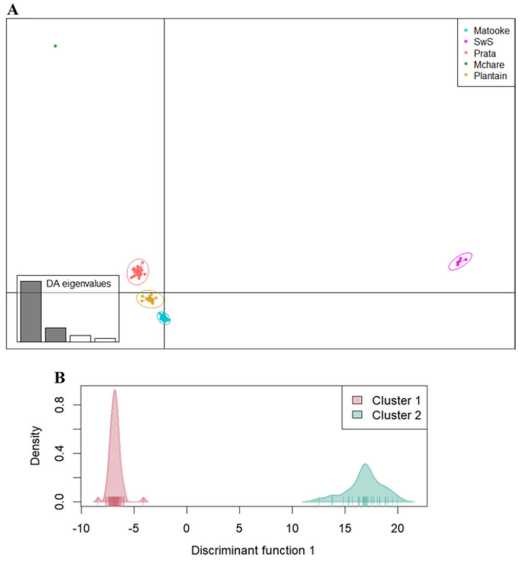


FIGURE 4 Discriminant analysis of principal components (DAPC). (A) Scatter plot of the 108 tetraploid *Musa* accessions from the five breeding programs using two discriminant functions with an inset plot of discriminant analysis (DA) eigen values and (B) plot of the densities of individuals on the first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

Acta Universitatis agriculturae Sueciae
Doctoral Thesis No. 2024:94

Banana improvement through crossbreeding is a slow and tedious process. Hence, its breeding efficiency should be improved. This thesis provides new information on available genetic diversity to broaden the narrow genetic base, and how to enhance the accuracy of estimating genetic parameters using marker-based models, paving the way for further genomic exploration in banana breeding programs for increased genetic gain.

Violet Akech received her doctoral education at the Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Alnarp, Sweden. She received her MSc in Plant Breeding and Seed Systems from Makerere University, Uganda.

Acta Universitatis agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-421-5
ISBN (electronic version) 978-91-8046-429-1