Genome-wide association for agro-morphological traits in a triploid banana population with large chromosome rearrangements

S. Rio[1,2,α], L. Toniutti[2,3,α], F. Salmon[2,3], C. Hervouet[1,2], C. Cardi[1,2], P. Mournet[1,2], C. Guiougou[2,3], F. Marius[2,3], C. Mina[2,3], J.M. Delos[2,3] , F. Lambert[2,3], C. Madec[2,3] , J.C. Efile[2,3], C. Cruaud[5], J.M. Aury[5], A. D'Hont[1,2], J.Y. Hoarau[2,4,β], G. Martin[1,2,β]

[1]CIRAD, UMR AGAP Institut, F-34398 Montpellier, France

[2]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

[3]CIRAD, UMR AGAP Institut, F-97130 Capesterre-Belle-Eau, Guadeloupe, France

[4]CIRAD, UMR AGAP Institut, F-97494 Sainte-Clotilde, La Réunion, France

[5]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France

[α]These authors contributed equally to the work

[β]These authors contributed equally to the work

Email addresses:

- Simon Rio: simon.rio@cirad.fr
- Lucile Toniutti: lucile.toniutti@cirad.fr
- Frédéric Salmon: frederic.salmon@cirad.fr
- Catherine Hervouet: catherine.hervouet@cirad.fr
- Céline Cardi: celine.cardi@cirad.fr
- Pierre Mournet: pierre.mournet@cirad.fr
- Chantal Guiougou: chantal.guiougou@cirad.fr
- Franck Marius: franck.marius@cirad.fr
- Claude Mina: claude.mina@cirad.fr
- Jean-Marie Delos: jean-marie-eric.delos@cirad.fr
- Frédéric Lambert: frederic.lambert@cirad.fr
- Camille Madec: camille.madec@palmelit.com
- Jean-Claude Efile: jean-claude.efile@cirad.fr
- Corinne Cruaud: cruaud@genoscope.cns.fr
- Jean-Marc Aury: jmaury@genoscope.cns.fr
- Angélique D'Hont: angelique.dhont@cirad.fr
- Jean-Yves Hoarau: jean-yves.hoarau@cirad.fr
- Guillaume Martin: guillaume.martin@cirad.fr

Abstract (250 words)

Banana breeding is hampered by the very low fertility of domesticated bananas and the lack of knowledge about the genetic determinism of agronomic traits. We analysed a breeding population of 2 723 triploid hybrids resulting from crosses between diploid and tetraploid *M. acuminata* parents, which was evaluated over three successive crop-cycles for 24 traits relating to yield components and plant, bunch and fruit architectures. A subset of 1 129 individuals was genotyped-by-sequencing revealing 205 612 single nucleotide polymorphisms. Most parents were heterozygous for one or several large reciprocal chromosomal translocations, which are known to impact recombination and chromosomal segregation. We applied two linear mixed models to detect associations between markers and traits: (i) a standard model with a kinship calculated using all SNPs and (ii) a model with chromosome-specific kinships that aims at recovering statistical power at alleles carried by long non-recombined haplotypic segments. For 23 of the 24 traits, we identified one to five significant quantitative trait loci (QTLs) for which the origin of favourable alleles could often be determined among the main ancestral contributors to banana cultivars. Several QTLs, located in the rearranged regions, were only detected using the second model. The resulting QTL landscape represents an important resource to support breeding programs. The proposed strategy for recovering power at SNPs carried by long non-recombined rearranged haplotypic segments is an important methodological advance for future association studies in banana and other species affected by chromosomal rearrangements.

Introduction

Dessert and cooking bananas (*Musa* spp.) are staple foods and an important source of income in many tropical and subtropical producing countries. There are about a thousand different banana cultivars, but the world banana production is based on a very limited number of natural hybrid cultivars and their somaclonal variants (Bakry et al., 2021). The 'Cavendish' bananas alone, which represent a few natural phenotypic somaclonal variants, account for about 57% of world banana production (Lescot et al., 2023). Such a narrow genetic base makes world's banana cultivation very vulnerable to the outbreak of diseases and pests, and variations caused by climate change or human practices. In this context, breeding for more diverse disease-resistant varieties that meet yield and quality commercial production criteria is essential for achieving a sustainable banana production.

Cultivated bananas are natural hybrids between species and subspecies of the genus *Musa* initially selected in Southeast Asia (Simmonds, 1962; Perrier et al., 2011; Sardos et al., 2022; Martin et al 2023). One of the main selected traits of cultivated bananas has been their ability to produce edible seedless fleshy fruits, due to sterility and parthenocarpy (Dodds and Simmonds, 1948, Simmonds 1953). For banana, one way to achieve complete or almost complete sterility is through the production of triploid individuals (3x), a ploidy level that provides more vigorous plants with larger bunches than diploids (Bakry et al., 2021). A common breeding strategy for obtaining progenies of triploid individuals involves crossings a diploid parent (2x) with a tetraploid parent (4x) (Tomekpe et al., 2004, Noumbissie et al., 2016, Nyine et al., 2018, Bakry et al., 2021, Salmon et al., 2023). The tetraploid parents are doubled diploid accessions obtained from a colchicine treatment or selected from crosses between triploid and diploid parents. Alongside this cross-breeding step leading to the selection of commercial triploid hybrids, a recurrent breeding step involving genetic improvement of diploid parents can be carried out. This strategy of improving parents through cycles of recombination and selection is likely to facilitate the simultaneous improvement of a larger number of agronomic traits of interest. However, banana breeding remains difficult as the most interesting banana progenitors have very low levels of fertility and germination rates, requiring embryo rescue. In addition, selecting triploid hybrids in the field requires a lot of space and time, given the large plant biomass and the relatively long cultivation cycles. In this context, knowledge about the genetic architecture of the main target agronomic traits could greatly help choosing the best resources and crossing schemes to accelerate the production of new cultivars.

Genome-wide association studies (GWAS) have been successfully applied in numerous crop species to identify quantitative trait loci (QTLs) controlling a wide range of agronomic and biochemical traits [see Gupta et al. (2019) for a review]. These studies exploit linkage disequilibrium (LD) between SNPs and causal variants at QTLs. Regarding banana, very few QTLs have been detected using GWAS approaches and for a limited number of traits: seedless phenotype (Sardos et al., 2016), bunch weight and its morphological components (Nyine et al., 2019). QTLs for organoleptic fruit quality during banana ripening (Biabiany et al. 2022) and resistance to subtropical race 4 of *Fusarium oxysporum* f. sp. *cubense* (Chen et al., 2023) have also been identified using QTL mapping.

A common issue in GWAS is controlling the detection of spurious associations caused by population structure, which generates LD between loci not necessarily physically linked. The most common way of limiting these false positive associations is to take into account genetic structure or kinship among individuals in the model (Yu et al.,2006). A drawback of this standard approach is that it limits statistical power (i.e. the probability of detecting true signals) in genomic regions with a large extent of LD. This is due to the fact that markers are used both for testing associations and estimating kinship. Rincent et al. (2014) proposed a method for efficiently recovering statistical power in regions

with large extent of LD, in which SNPs present on the same chromosome as the tested SNP are discarded to estimate kinship.

Banana cultivar genomes are a mosaic of ancestral contributions (Baurens et al., 2019, Cenci et al., 2020, Martin et al., 2020a, Martin et al., 2023). Some of the contributing species and subspecies differ by a few large chromosomal rearrangements, mainly large reciprocal translocations, sometimes associated with inversions (Shepherd, 1999, Martin et al., 2017, Dupouy et al 2019, Wang et al., 2019, Martin et al., 2020b, Liu et al. 2023), resulting in structural heterozygosities in hybrid cultivars. So far, zero to four large chromosome rearrangements have been observed in the genome of cultivars (https://banana-genome-hub.southgreen.fr/translocation). These structural heterozygosities generated segregation distortions and the inversions prevented recombination (Martin et al., 2017, Baurens et al., 2019, Dupouy et al 2019, Martin et al., 2020b). In the QTL mapping study reported by Biabiany et al. (2022), the presence of a large structural heterozygosity in one parent - resulting from a recriprocal translocation between chromosome 1 and 7 associated with an inversion - blocked recombination along chromosome 1 and generated co-segregation between chromosome 1 and 7. This co-segregation prevented the precise location of the fruit quality QTLs. The consequences of these structural heterozygosities have not yet been assessed in QTL banana studies based on GWAS approaches.

In this work, we analysed a large breeding population of 2 723 triploid hybrids from CIRAD's banana varietal improvement program (Salmon et al., 2023; Tonuitti et al., 2023). This triploid population was bred from representative *M. acuminata* accessions containing some large chromosomal rearrangements. They were phenotyped for 24 agro-morphological traits of breeding interest relating to yield components as well as plant, bunch and fruit architectures. The objectives of the study were to i) evaluate the impact of large chromosome rearrangements on QTL detection, ii) propose a new GWAS model to limit their negative impact on the ability to detect QTL and iii) obtain an extensive overview of the QTL landscape for the traits in the *M. acuminata* resources studied.

Results

In this study, we analysed a large breeding population of 2 723 triploid banana hybrids evaluated for 24 traits relating to yield components as well as plant, bunch and fruit architectures (Table 1). A subset of 1 129 hybrids were genotyped for 205 612 polymorphic bi-allelic SNPs and used for GWAS.

Table 1: Description of traits.

Category	Description	Unit
Plant architecture	Pseudostem height (PH) - Measured at flowering	cm
Plant architecture	Pseudostem girth (PG) - Measured at 1m above soil level (at flowering?)	cm
Plant architecture	Robustness index (PH/PG) - Robustness of the pseudostem	
Plant architecture	Number of leaves at flowering - Counted on standing leaves	
Plant architecture	Number of leaves at harvesting - Counted on standing leaves	
Plant architecture	Leaf blade length (LL) - Measured on rank 3 leaf	cm
Plant architecture	Leaf blade width (LW) - Measured on rank 3 leaf	cm
Plant architecture	Leaf index (LL/LW)	
Bunch architecture	Bunch angle - Angle between the bunch and the pseudostem	°
Bunch architecture	Peduncle length (PL)	cm
Bunch architecture	Peduncle diameter (PD)	cm
Bunch architecture	Peduncle index (PL/PD)	
Bunch architecture	Bunch length at maturity (BL) - Measured at maturity	cm
Bunch architecture	Bunch compactness index (BL/NH)	cm
Fruit architecture	Fruit pedicel length	mm
Fruit architecture	Fruit pedicel diameter	mm
Fruit architecture	Fruit length	mm
Fruit architecture	Fruit grade	mm
Yield component	Number of hands on a bunch (NH) - Counted on a bunch	
Yield component	Number of fruits on a bunch (NF) - Counted on a bunch	
Yield component	Number of fruits per hand (NF/NH)	
Yield component	Bunch weight - Measured at maturity	kg
Yield component	Fruit weight - Measured at maturity	g
Yield component	Days to fruit maturity - Interval between flowering and harvesting	days

Impact of large chromosome rearrangements on population structure

Hybrids resulted from crosses between diploid and tetraploid *M. acuminata* parents, most of which were heterozygous for one to three large reciprocal translocations (Table 2). These translocations involved four couples of chromosomes: 1/4, 1/7, 2/8 and 3/8. Parents heterozygous for translocations 1/4, 1/7 and 3/8 display absence or reduction of recombination involving large chromosome segments, while for translocations 2/8 a reduction of recombination is observed only at the breakpoints (Martin et al 2020b). Moreover, some chromosomes are involved in distinct translocations. For example, chromosome 1 is involved in three distinct chromosome structures: the reference chromosome structure, the 1/4 reciprocal translocation (1T4, 4T1 haplotypes) and the 1/7 reciprocal translocation (1T7, 7T1 haplotypes) (Figure 1A).

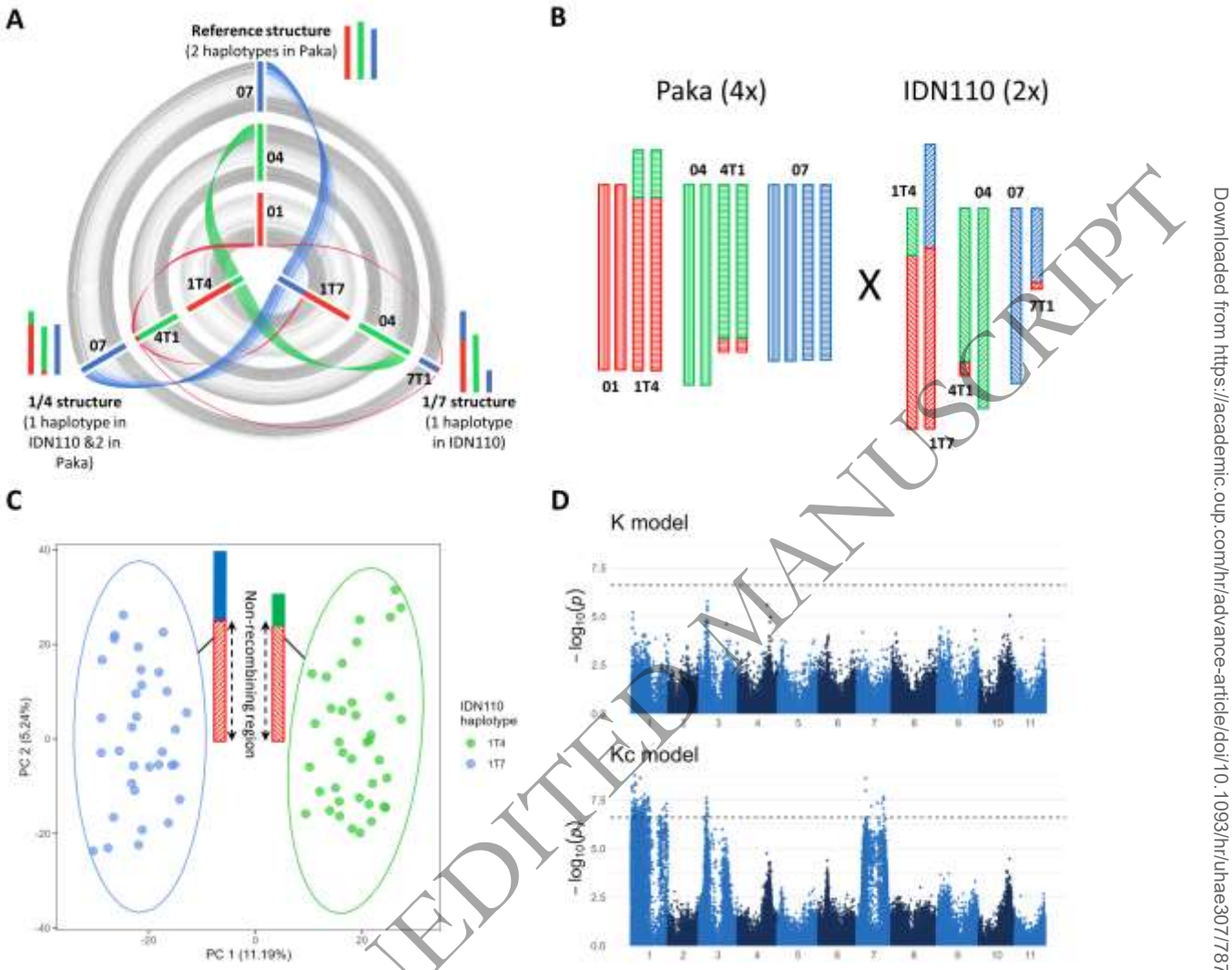


Figure 1: Impact of large chromosome rearrangements on GWAS. **A**) Comparison of two reciprocal translocations (1/4 and 1/7) involving chromosomes 1, 4 and 7 with the reference chromosome structure. **B**) Chromosome structural heterozygosities in the tetraploid Paka (4x) and in the diploid IDN110 (2x) accessions. **C**) SNP-based principal component analysis performed on the 71 genotypes of the Paka (4x) x IDN110 (2x) population. The large region of chromosome 1 that does not recombine due to the structural heterozygosity is indicated in red. **D**) Manhattan plots obtained for bunch angle from a standard model (K model) compared to the model proposed to recover signals (Kc model)

To evaluate the impact of distinct chromosome structures on the estimation of population structure, we exploited a progeny from the cross Paka (4x) x IDN110 (2x). Paka has two copies of the reference chromosome 1 and two copies of the 1/4 translocated chromosomes (1T4 and 4T1), while IDN110 has one copy of the 1/4 translocated chromosomes and one copy of the 1/7 translocated chromosomes (1T7 and 7T1) (Figure 1B). A principal component analysis performed with the SNP data (Figure 1C) clustered the hybrids of the Paka (4x) x IDN110 (2x) progeny into two groups. These groups corresponded to the presence of the 1T4 haplotype or the 1T7 haplotype inherited from IDN110, presenting an absence of recombination on the chromosome 1 portion of these haplotypes. This illustrates the impact that large chromosomal rearrangements can have on population structure.

Table 2: Diploid and tetraploid parents of the 1 129 hybrids used in the GWAS. Presence of large reciprocal translocation in heterozygous or homozygous states or absence (-) involving 4 couples of chromosomes: 1/4, 1/7, 2/8 and 3/8.

	Number of crosses		Large reciprocal translocations			
Parents	Diploid	Tetraploid	1/4	1/7	2/8	3/8
Akondro Mainty[a]	0	5	heterozygous	-	-	heterozygous
Chicame[a]	0	8	heterozygous	-	-	heterozygous
Gu Nin Chiao[c]	1	0	heterozygous	heterozygous	-	-
Khi Maeo[b]	3	0	heterozygous	heterozygous	-	heterozygous
IDN 077	1	1	heterozygous	heterozygous	-	-
IDN 110[c]	3	5	heterozygous	heterozygous	-	-
IRFA 903[c]	0	2	heterozygous	heterozygous	-	-
Malaccensis nain	3	0	homozygous	-	-	-
Manang	3	1	-	heterozygous	heterozygous	-
Microcarpa	1	0	-	-	-	heterozygous
Nzumoheli II[a]	0	1	heterozygous	-	-	heterozygous
Pa (Patthalong)[b]	2	0	heterozygous	heterozygous	-	heterozygous
Pahang	2	0	heterozygous	-	-	-
Paka	2	6	heterozygous	-	-	-
Pisang Jaran	2	0	-	heterozygous	-	heterozygous
Pisang Lilin	1	4	heterozygous	-	-	-
Pisang Madu	7	2	-	heterozygous	-	-
Pisang Pipit	0	3	-	heterozygous	-	heterozygous
Sinwobogi	2	0	-	-	-	-
THA 052[b]	3	0	heterozygous	heterozygous	-	heterozygous
Thong Det	2	0	-	heterozygous	-	-

[a], [b] and [c] are groups of somaclones

Accounting for population structure generated by large chromosomal rearrangements in GWAS

In standard GWAS models, a polygenic background effect is generally included whose covariance is proportional to a kinship matrix calculated with SNPs. The aim is to control for false positives by limiting statistical power at SNPs whose polymorphism is correlated with population structure. As a consequence, the statistical power at SNPs tagging the segregation of the 1T4 and 1T7 haplotypes from IDN110 was also limited when applying the standard K model. We circumvented this problem by proposing an alternative GWAS model, the Kc model, in which a kinship is calculated specifically for each chromosome carrying the SNPs to be tested and by excluding SNPs from that chromosome and from other chromosomes involved in structural variation with it. The added value provided by this new GWAS methodology can be exemplified by the analysis of the bunch angle. For this trait, the standard K model did not reveal any significant associations for a 5% Bonferroni threshold, while the new Kc model helped to recover signals on chromosomes 1, 3 and 7 (Figure 1D). The associations on chromosomes 1 and 7 likely identified a QTL with an allele located on a 1T7 haplotype that did not recombine in the population, such as that carried by IND110. The Kc model also helped recovering

signals on chromosome 3 that may have been hidden in the K model GWAS due to limited recombination of rearranged 3/8 haplotypes in other crosses, the 3/8 reciprocal translocation being absent in Paka and IDN110.

QTL detection for agro-morphological traits

The moderate to high heritability estimated for the 24 traits over the experimental design confirmed the relevance of this dataset to perform GWAS (Table S1). Both K and Kc GWAS models were applied to all traits and SNPs (see summary statistics in Dataverse repository, QQ-plots and Manhattan plots in Figures S1 and S2 for the K and Kc models, respectively) with two significance thresholds, i.e. a 5% Bonferroni threshold and a 5% false discovery rate (FDR). Significant associations were clustered into 62 consensus QTLs over the two models and LD intervals were calculated between the most significant SNP of the QTL and neighbouring SNPs (Figure 2). Co-segregation between QTLs located on different chromosomes were identified and could always be related to the presence of the 1/7 and 1/4 translocations (Table S2). Regarding the comparison of GWAS models, a same number of 43 QTLs were detected for each model considering a 5% FDR, but a higher number of QTLs were detected using the Kc model (26) than using the K model (19) considering a 5% Bonferroni threshold. A comparison of QTL LD intervals according to the model is presented in Figure S3 and information on each QTL is presented in Table S3.

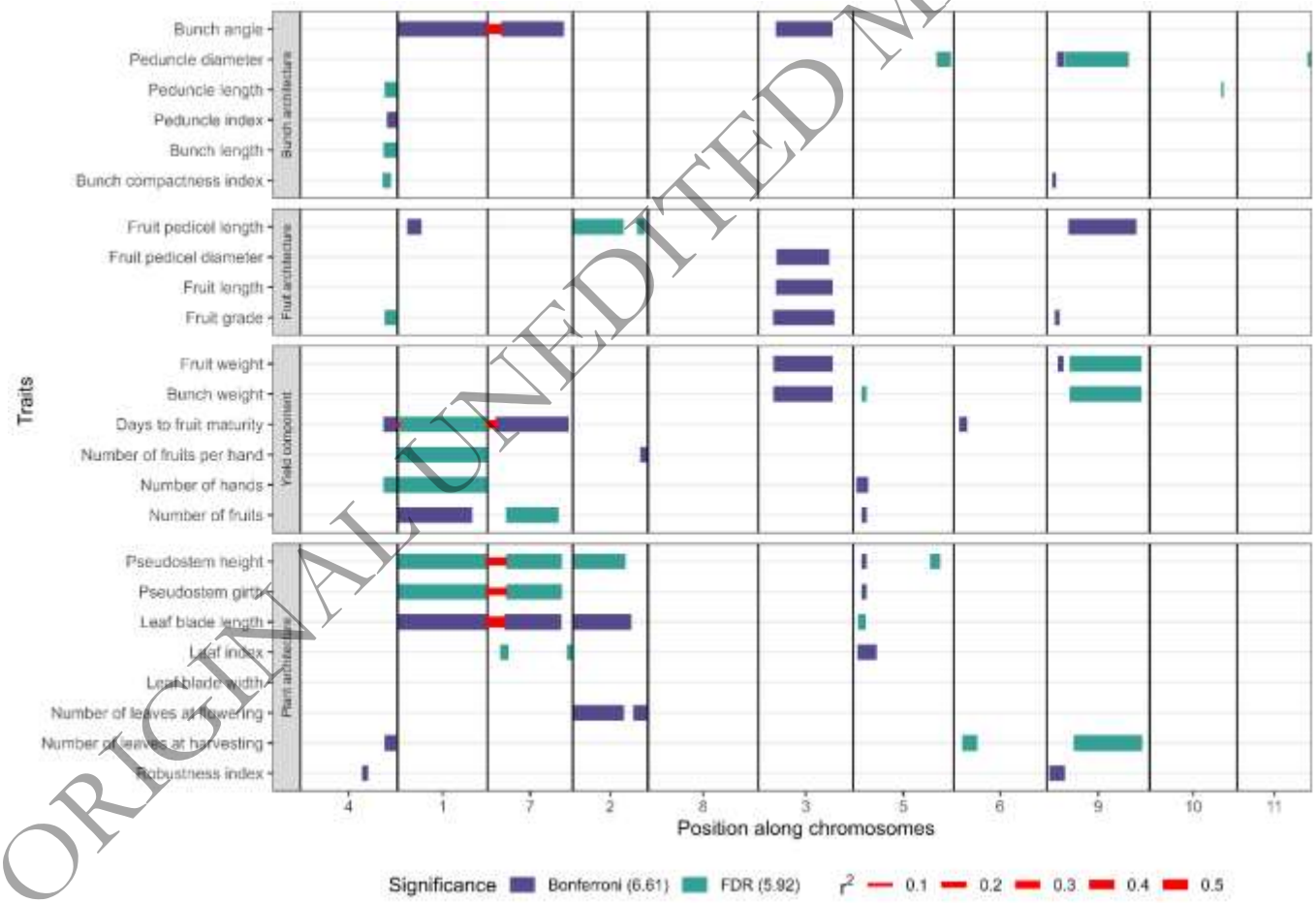


Figure 2: Localization of consensus QTL LD intervals along chromosomes according to two significance thresholds: Bonferroni ($-\log_{10}(p) = 6.61$) and FDR ($-\log_{10}(p) = 5.92$). Co-segregations between the most significant SNPs of each interval are indicated by a red segment with a width proportional to their LD (r^2) whose values are shown in Table S2. The continuous r^2 size scale is

represented by discrete values from 0.1 to 0.5. The order of the chromosomes on the x-axis was chosen so as to position the chromosomes involved in a reciprocal translocation close to each other.

QTL allele ancestries

The determination of the ancestral origin of some of the SNP alleles made it possible to characterize the origin of the favourable and unfavourable alleles of part of the QTLs detected in this study. A focus was done on four QTLs involved in the genetic determinism of days to fruit maturity, fruit grade, bunch angle and number of fruits (Figure 3). Figures representing the characterization of the allele ancestry of each QTL are available in Figure S4 and S5 for the K and Kc model, respectively.

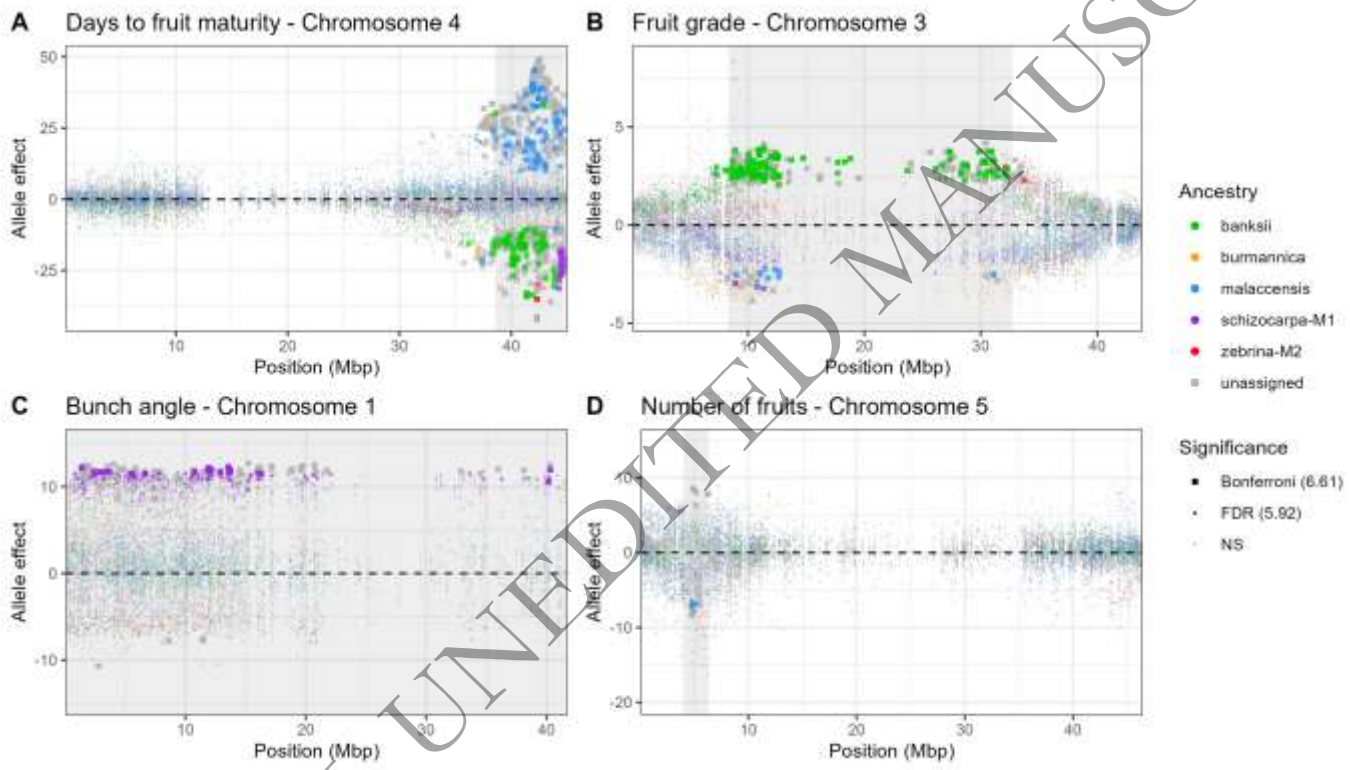


Figure 3: Estimated allele effects for (**A**) days to fruit maturity QTL on chromosome 4, (**B**) fruit grade QTL on chromosome 3, (**C**) bunch angle QTL on chromosome 1, and (**D**) number of fruits QTL on chromosome 5. The plotted effects were obtained from the Kc model, except for number of fruits for which the effects were obtained from the K model. Dots were coloured according to allele ancestry and shaped according to the level of significance of the test. When no ancestry could be assigned, the effect represented was that of the alternative allele. The QTL interval is indicated by a grey area.

For days to fruit maturity (Figure 3A), a QTL has been detected with both models at the end of chromosome 4, with a LD interval ranging from 38.55Mbp to 44.72Mbp. As cycle length is one of the components of banana yield due to the asynchronism of growth cycle between different plants, shortening the interval between flowering and harvesting is of interest from a breeding perspective. As a consequence, the effects associated with the presence of favourable alleles have a negative sign. Most favourable alleles showed banksi ancestry, except for two zebrina-M2 alleles as well as a

large number of schizocarpa-M1 alleles at the end of the LD interval. In contrast, most unfavourable alleles showed malaccensis ancestry, with the exception of two banski alleles.

For fruit grade (Figure 3B), a QTL has been detected with both models around the centromere of chromosome 3, with a large LD interval ranging from 8.33Mbp to 32.70Mbp. Provided that a breeder seeks to increase fruit diameter, favourable alleles have a positive sign. Again, most favourable alleles showed banksii ancestry, with the exception of a zebrina-M2 allele at the end of the interval. In contrast, unfavourable alleles showed malaccensis or schizocarpa-M1 ancestry.

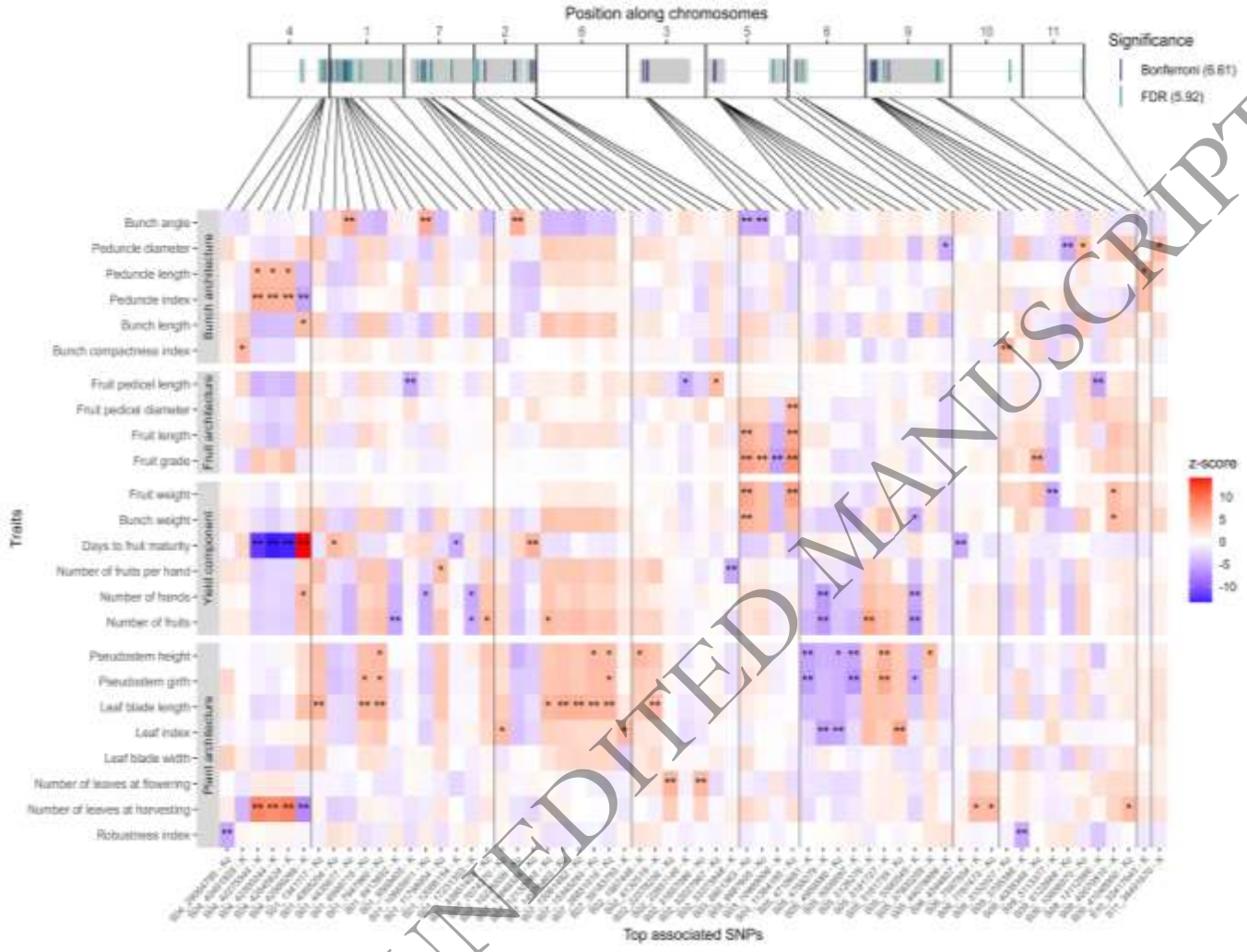
For bunch angle (Figure 3C), a QTL has been detected with the Kc model only and spans the entire chromosome 1. From a breeding perspective, the smallest angle between the bunch and the pseudostem is generally desirable in order to harvest fruits of uniform dimension from the bunch. Most alleles with positive signs had schizocarpa-M1 ancestry with few exceptions including burmannica, malaccensis and zebrina-M2 ancestry. No allele whose presence is associated with an effect of negative sign (favourable alleles) could be assigned to an ancestry. Note that this QTL co-segregated with a QTL on chromosome 7 as a result of the segregation of rearranged 1/7 haplotypes, and this co-segregating QTL showed a similar ancestry pattern with several schizocarpa-M1 favourable alleles (Figure S5).

Finally, for number of fruits (Figure 3D), a QTL has been detected with both models at the beginning of chromosome 5, with a LD interval ranging from 3.98Mbp to 6.34Mbp. Few alleles with negative signs (unfavourable alleles for bunch weight) suggested a malaccensis ancestry. No allele whose presence is associated with an effect of positive sign could be assigned.

Meta-analysis

Based on Figure 2, some QTLs detected for different traits colocalized to the same genomic regions. They may result from a single causal locus with pleiotropic effects on traits. To investigate the existence of such effects, we performed a meta-analysis using the most associated SNPs of all QTLs by transforming the effect of alleles alternative to those of the reference genome into z-scores (Figure 4). Note that colocalizing QTLs did not necessarily have the same most significant SNP. Three sets of colocalized QTLs with a high level of significance for several traits are described hereafter.

A first set of colocalized QTLs involved peduncle length, peduncle index, bunch length, bunch compactness index, days to fruit maturity, number of hands, and number of leaves at harvesting, for



which at least one of the following SNPs located at the end of chromosome 4 was significantly detected: S04_40491859, S04_42275544, S04_42300244, S04_42545624 and S04_42589288. For instance, the presence of the alternative allele of SNP S04_42545624 was significantly associated with a decrease in days to fruit maturity but an increase in number of leaves at harvesting, peduncle length and peduncle index. In contrast, the alternative allele of SNP S04_42589288 showed opposite effect signs compared to SNP S04_42545624 for these same traits. This opposition of signs was not simply due to the arbitrary coding of the alleles of the two SNPs (based on the reference genome), as both SNPs showed similar genotypic classes frequencies (see Table S3). It rather resulted from the existence of two parental haplotypes with contrasted effects segregating in some crosses, each haplotype being tagged by different SNPs.

Figure 4: QTL meta-analysis. For each QTL, the effect of the most associated SNP in the interval was transformed into a z-score using the estimate obtained from the model for which it was most significant. The effect considered was that of the allele alternative to the reference genome. The model was reported after each SNP name on the x-axis and the level of significance is indicated using "*" and "**" if the SNP was detected using FDR or both FDR and Bonferroni, respectively. The order of the chromosomes on the x-axis was chosen so as to position the chromosomes involved in a reciprocal translocation close to each other.

A second set of colocalizing QTLs involved bunch angle, fruit pedicel diameter, fruit length, fruit grade, fruit weight and bunch weight, for which at least one of the following SNPs located on chromosome 3 was significantly detected: S03_8901363, S03_8987605, S03_10969006 and S03_11364185. For instance, the presence of the alternative allele of S03_8901363 was significantly associated with an increase in fruit length, fruit grade, fruit weight and bunch weight but a decrease in bunch angle.

A last set of colocalizing QTLs involved number of hands, number of fruits, pseudostem height, pseudostem girth and leaf index, for which at least one of the following SNPs located on chromosome 5 was significantly detected: S05_4718681, S05_4758279, S05_4959580, S05_5009597, S05_5126278, S05_5181727, S05_5181739, S05_5390549. While none of these SNPs were significant for all traits, the sign of the effect associated with the presence of the alternative allele was consistent across all traits.

Discussion

Accounting for large chromosome rearrangements in GWAS

Large chromosome rearrangements at the heterozygous state in accessions disrupt recombination and segregation during the meiosis (Tadmor et al., 1987; McKim et al 1988; Jáuregui et al., 2001; Stevison et al., 2011). In banana, Martin et al. 2020b reported the presence of several reciprocal translocations, sometimes associated with inversions, in the various genetic groups involved in cultivars. They also showed that some of these translocations at heterozygous state in parents led to haplotype segments showing an absence or a reduction of recombination and/or co-segregations with other haplotype segments. At the scale of our multi-parental banana population involving such parents, limited recombination of haplotype segments generated large blocks of markers in LD. The transmission of non-recombined rearranged haplotypes was associated with population structure in the progeny. This genetic structuring of the progeny could theoretically translate into phenotypic structuring provided that the non-recombined haplotypes carry one or more QTLs with strong effects.

Because the markers involved in such haplotypes were used both for testing associations with traits and estimating kinship, they were found to correlate with the population structure, which limited their statistical detection power in the GWAS mixed model. To overcome this limitation, we followed the method of Rincent et al. (2014) by proposing the Kc model in which a kinship is calculated with all markers, excluding those located on the same chromosome as the marker tested. We adapted the original method by also excluding the markers that are located on chromosomes involved in a network of reciprocal translocations (e.g. chromosomes 1, 4, and 7 for the 1/4 and 1/7 reciprocal translocations). In our population, this strategy was supported by the existence of LD between markers located on different chromosomes according to the reference genome structure, but located on a same chromosome in rearranged chromosomes.

The Kc model helped recover several QTLs when compared to the standard K model that considered a kinship estimated using all SNPs, especially on chromosomes 1 and 7. The 1/7 alternative chromosome structures segregated in a large proportion of crosses, which is known to be associated with suppressed recombination on a large portion of the 1T7 haplotype (Martin et al., 2020b). Because of the limitation of statistical power mentioned hereabove, any QTL allele specifically present on the 1T7 haplotype would be particularly difficult to detect using a standard K model.

Conversely, some QTLs were only detected with the K model. This could be explained by the procedure of correcting the Wald statistics using the inflation factor λ, which penalized more the Kc model than the K model. This stronger penalty for the Kc model probably resulted from the exclusion of all markers from certain chromosomes, which may have limited the accuracy of kinship estimation. We recommend applying both K and Kc models jointly to the data, and aggregate results to obtain the most complete QTL landscape for the studied traits.

Because chromosome rearrangements in the heterozygous state are pervasive in the progenitors of banana breeding programs, it seems important to apply our methodology to future GWAS in banana, so as not to miss out on detecting part of the QTL landscape.

GWAS design for triploid banana populations

This triploid population, resulting from crosses between tetraploid (doubled-diploid) and diploid parents, was suboptimal for QTL detection with respect to the statistical properties of segregating markers. For a tetraploid parent with genotype 0:0:1:1, the segregation of its markers assuming polysomy at meiosis gives the following gametes: 0:0 (1/6), 0:1 (4/6) and 1:1 (1/6). When crossed to a diploid parent giving a single allele, e.g. 0, it generates three genotypic classes with the same frequencies as the gametes mentioned above. With the diploid SNP coding used in this study, this is equivalent to observing two genotypic classes (i.e. homozygous 0:0 and heterozygous 0:1), one of which is rare. It has been demonstrated that the existence of rare genotypic classes is associated with poor statistical power in GWAS (Sham and Purcell, 2014). Even supposing a triploid SNP coding, two-thirds of the progeny would be grouped into a single genotypic class (i.e. 0:0:1) thus recovering only a modest amount of statistical power. Note that assuming preferential pairing between doubled/identical chromosomes at meiosis in the tetraploid parent, the statistical power would decrease further with increasing frequency of 0:1 gametes. Alternatively, the segregation of markers in diploid parents allow for balanced genotypic classes in the progeny, which should allow for optimal power for detecting QTLs. In the future, we may consider generating crosses between diploid parents for the detection of QTLs and use triploid populations for validation only. However, it should be noted that the studied triploid population offered the advantage of detecting and evaluating directly the effect of QTLs in a triploid genetic background, which is that of most cultivars. In addition, it allowed to exploit the phenotyping effort that had already been carried out as part of CIRAD's banana breeding program, providing a large population phenotyped for several traits.

The choice of diploid coding for a triploid population for this analysis can be questioned. This choice was motivated by two reasons: (i) the GBS approach does not allow the two heterozygous classes (0:0:1 and 0:1:1) to be easily distinguished; (ii) the segregation of tetraploid parents produces a non-negligible proportion of aneuploid individuals, whose proportion is only increased by the presence of structural variations (Baurens et al., 2019). These two phenomena together mean that attempting to predict dosage in a triploid will produce a non-negligible number of errors. In this context, it seemed safer to genotype individuals only for their homozygous/heterozygous state.

The GWAS implemented in this study relied on hybrid genotypic values estimated over three growth cycles, which amounts to detecting QTLs with a relatively stable effect over all cycles. These QTL are of priority interest for breeding, as bananas are generally grown over many cycles. Any detection of QTL effects specific to particular cycles would require a multi-cycle GWAS including an interaction effect between the tested marker and the cycle. However, unlike in our study, replicates of each hybrid would be required to correctly estimate the cycle-specific genotypic values to be used as response variables in the multi-cycle GWAS. In addition, it would be interesting to evaluate each hybrid in several environments to distinguish the QTLs with a stable effect in all environments from those that interact with the environment.

Genetic architecture of agro-morphological traits

In this study, a set of 62 consensus QTLs were detected for 23 of 24 traits which were located on 10 of the 11 chromosomes of the banana genome. These results consist of the second GWAS results for yield components and fruit size (bunch weight, number of hands and fruits, fruit length and diameter) after Nyine et al. (2019), and the first GWAS results for all other traits related to plant architecture and bunch architecture. Compared with this first study, the size of the population we studied was much larger (almost four times) and the density of SNP markers was also much higher (more than seven times). This greater experimental input has made it possible to increase the power of QTL detection for the traits shared between the two studies relating to yield components and fruit size. For these traits, Nyine et al. (2019) detected 25 genomic loci mostly localized on chromosome 3. In our study, for these same traits, we aslo identified a large genomic region around the centromere of this chromosome 3 with significant QTL signals. In addition, the QTL landscape obtained for all 24 traits shows QTLs spread across all but one chromosome in the genome (chromosome 8). QTLs were detected for most traits, with one to five significant QTL detected for each trait, with the exception of leaf blade width for which no QTL were detected.

This relatively modest number of significant QTL per trait and their relatively modest effects suggested that variation of most traits in the studied population is controlled by many other genetic factors not detectable (due to small effect and/or unbalanced genotypic classes, causal loci in low LD with our SNPs, non-additive genetic effects). Most traits appeared essentially quantitative in nature. In general, QTLs displayed a pleiotropic effect on different traits, which is known to cause genetic correlations between traits (Falconer and Mackay, 1996). Such correlations between banana agro-morphological traits have already been reported by Nyine et al. (2017) using a population from an East African highland banana breeding program.

The QTL associated with bunch angle, bunch weight and four of its components relative to fruit dimension (fruit weight, fruit length, fruit grade and fruit pedicel diameter) on chromosome 3 is a first example of connected genetic architecture between traits. Bunch weight and its fruit components showed a QTL effect of the same sign, while the QTL effect of bunch angle was of the opposite sign. This QTL probably corresponded to the QTL detected by Nyine et al. (2019) for bunch weight component traits. Based on the allele ancestry assignments of Martin et al. (2023), we highlighted alleles of banksii origin associated with increased values for bunch weight and its fruit components and with decreasing values for bunch angle (Figures 3B, S4 and S5). This pleiotropic effects of opposite sign between bunch weight and its fruit dimensions, on the one hand, and bunch angle with the pseudo-stem, on the other, is congruent with the fact that heaviest-bunch cultivars tend to have pendulous orientation while the smallest-bunch cultivars tend to have sub-horizontal inflorescences (Karamura and Karamura 1995). It is conceivable that this QTL has a direct positive effect on bunch weight, which leads to greater bending of the peduncle, resulting in a smaller angle between the bunch and the pseudostem.

A second example consists of the QTL at the end of chromosome 4, which is associated with several traits and whose effect is particularly large for days to fruit maturity and number of leaves at harvesting, but with opposite sign. The negative relationship between these two traits can be explained by the fact that leaf emission stops after flowering and leaves are more likely to disappear when the interval between flowering and harvest increases due to senescence, wind damage or diseases. From a breeding perspective, a short interval between flowering and harvesting is of interest, as it increases the number of cycles in a given period. Again, we showed that the favourable allele for this QTL had essentially a banksii origin.

A last example is the QTL at the beginning of chromosome 5, which is notably associated with the number of hands and fruits per bunch and pseudostem height and girth, with the same sign of effect. While a greater number of fruits is desirable to reach higher yields, taller plants are undesirable because of their vulnerability to lodging. The allele statistically associated with smaller plants and a smaller number of fruits had a malaccensis origin.

Because of the multiparental nature of the population, QTL intervals were too wide to identify candidate genes. They covered more than 10 Mbp when they were located close to centromeres or when they tagged haplotypes whose recombination was limited by chromosomal rearrangements. As a result, the number of genes annotated from the *M. acuminata* reference sequence V4 (Belser et al., 2021) in each QTL interval was very large, ranging from 169 to 2743. Further work is needed to reduce the size of these intervals to be able to identify candidate genes.

New perspectives for banana breeding

Based on these QTLs, it is possible to suggest ways of improving banana breeding schemes, which have so far made little use of the genomic information.

Firstly, parents can be characterized at QTLs so that crosses between parents carrying favourable alleles can be prioritized. These cross choices could be done among the set of *M. acuminata* parents currently available in the breeding program, as well as future improved parents obtained from recurrent parental selection.

Secondly, an early selection of progeny could be made prior to field evaluation based on their genotype at QTLs. This could enable a larger set of promising hybrids to be evaluated during the first phase of field evaluation. However, for quantitative traits, marker-assisted selection (MAS) has often proved disappointing (Moreau et al., 2004), which can be explained by the insufficient proportion of the genetic variance explained by detected QTLs. Genomic prediction has often proved to be a more promising strategy than MAS (Wang et al., 2014 ; Arruda et al., 2016 ; Zhang et al., 2016), as it is not limited by statistical power associated with QTL detection. In banana, genomic prediction has already been evaluated for agro-morphological traits with moderate to high prediction accuracies (Nyine et al. 2018), confirming the interest of this approach.

The characterization of allele origin has enabled us to identify the ancestral groups that have contributed numerous QTL favourable alleles. The most striking example is the banskii group that has contributed the favourable allele for the bunch and fruit weight QTL as well as for the QTL allele associated with shorter cycle length. These results confirm the major role of the banksii group in the formation of dessert banana cultivars, and suggest that particular attention should be paid to germplasm carrying alleles of banksii origin.

The disruption of recombination generated by large chromosome rearrangements in heterozygous state in parents needs careful consideration in breeding. Both favourable and unfavourable alleles may co-segregate due to their localization on non-recombined haplotypic segments. This situation of genetic load makes it difficult to take advantage of the potential genetic variability associated with crossbreeding of the diploid and tetraploid parents so far available for triploid breeding. To solve this issue, pre-breeding programs at the diploid level could be set up to generate new improved parents that are homozygous for chromosome rearrangements. In the homozygous state, the alternative chromosome structures could recombine normally and unfavourable alleles would be purged more easily by selection. Pre-breeding programs at the diploid level would also enable an improvement in the cross-breeding value of the parents, prior to final crosses leading to triploid hybrids. The extensive QTL information produced in this study could be useful to guide such pre-breeding programs.

Materials and methods

Breeding population

The breeding population consisted of 2 723 triploid hybrids resulting from biparental crosses implying 38 *M. acuminata* accessions including wild accessions and cultivars (Toniutti et al., 2023). Triploid hybrids were obtained by crossing a diploid parent with a tetraploid parent, the latter resulting from chromosome doubling of a diploid accession using a colchicine treatment. The population represented 116 full-sib families, giving a relatively modest average number of progenies per family (23.47) due to the generally low fertility levels of most parental combinations. Hybrids were all produced and evaluated at CIRAD Neufchateau station, Capesterre Belle-Eau, Guadeloupe, French West Indies (16°05'N, 61°35'W, elevation 250 m, average rainfall 3500 mm, average temperature of 25°C and soil classified as andosol). A subset of 1 463 hybrids were genotyped using a genotyping-by-sequencing (GBS) approach, which led after SNP quality filtering (see below) to a total of 1 129 hybrids available for the GWAS analysis. This final subset of hybrids was derived from 21 *M. acuminata* accessions involved in 38 biparental combinations as diploid and/or tetraploid parents (Table 2). These accessions comprised three groups of somaclonal mutants (three mutant triplets), that are genetically indistinguishable but phenotypically distinct. The genome of these 21 accessions taken as a whole encompassed four major reciprocal translocations between four pairs of chromosomes, as compared with the ancestral *M. acuminata* structure (Martin et al., 2020) that is the structure of the *M. acuminata* reference sequence V4 (Belser et al., 2021). All accessions were structurally heterozygous for one to three of these large reciprocal translocations, with the exception of one accession (Sinwogobi).

Experimental design and phenotyping

All 2 723 hybrids planted in field experiments were evaluated for the 24 agronomic traits listed in Table 1 that were related to yield components as well as plant, bunch and fruit architectures.

The experimental design consisted of 12 trials comprising two to nine experimental unit blocks (48 in total) successively planted from 2011 to 2016. Each block contained 64 plants comprising 56 unreplicated hybrids and eight checks (five Cavendish, one Pisang Ceylan, one Pisang Madu and one Calcutta 4) (Toniutti et al., 2023). Phenotypic data were collected over three successive growth cycles (from 2012 to 2017). For each trait, best linear unbiased prediction (BLUP) of hybrid performances were calculated over the experimental design using the linear mixed model described in Toniutti (2023) that accounts for diploid and tetraploid parental effects. Inference of model parameters was performed using ASReml-R (v3, Gilmour et al., 2009). Their estimates and trait heritability are presented in Table S1.

Genotyping

A subset of 1 463 individuals from the phenotyped triploid population was genotyped-by-sequencing (GBS). Leaf samples were collected on the third leaf after the cigar leaf on adult individuals and DNAs were extracted from 3 g of leaf according to a modified MATAB method (Risterucci et al. 2000). Libraries were made at the GPTR platform (https://umr-agap.cirad.fr/en/plateformes/plateformes-regionales/genotyping) using PstI and MseI restriction enzymes and single-end sequencing performed on the GeTPlaGe platform (https://get.genotoul.fr) or Genoscope (http://www.genoscope.cns.fr) using an Illumina HiSeq sequencer (Illumina, San Diego, CA, USA). Raw sequence reads were demultiplexed using GBSX, version 1.2 (Herten et al., 2015). Adapters were removed and reads were quality trimmed using the CUTADAPT program (Martin, 2011).

A triploid variant calling was performed on individuals using the *M. acuminata* reference sequence V4 (Belser et al., 2021) with vcfhunter toolbox (https://github.com/SouthGreenPlatform/VcfHunter)

(Garsmeur et al., 2018) as described in Baurens et al. (2019). Only bi-allelic sites with no indels were selected for the analyses. The genotype call was then diploidized in the sense that the two difficult to distinguish triploid heterozygous classes (0:0:1 and 0:1:1) were combined into a single heterozygous class (0:1), 0 and 1 being the reference and alternative allele, respectively. Genotypic data points were set as missing values if their read depth was below 10 or above 10 000 and, for heterozygous data points, if an allele was supported by less than three reads or with an allele depth ratio (i.e., allele depth to total read depth) below 0.05. A prefiltered vcf was obtained by first eliminating SNPs with more than 50% missing values and then eliminating 334 individuals with more than 50% missing values. Polymorphic sites were additionally filtered following several criteria: (i) Removal of sites with more than 20% missing data on the 1 129 remaining individuals using vcfFilter.1.0.py of vcfhunter toolbox. (ii) Selection of sites for which a proportion of heterozygous individuals is comprised within [0.1; 0.9] in at least one biparental population using the vcf2PopStat.py script added to vcfhunter toolbox. (iii) Selection of sites with minor allele frequency (MAF) greater than 0.01 and a global heterozygosity comprised within [0.01; 0.99].

The final vcf file included 205 612 SNPs for 1 129 individuals representing 38 families ranging in size from 2 to 141 individuals with a median number of individuals of 28.

GWAS

The standard GWAS model of Yu et al. (2006) was applied at each of the M SNPs and is referred to as the "K model" further in the text:

$$Y_{ik} = \mu + \alpha_k + \beta^m x_{ik}^m + G_{ik} + E_{ik} \quad (1)$$

where Y_{ik} is the reference phenotypic value of hybrid i from family k (i.e. the BLUP calculated over the experimental design), μ is the intercept, α_k is effect of family k, β^m is the effect of SNP m, $x_{ik}^m \in \{0,0.5,1\}$ is the genotypic score of hybrid i from family k at SNP m, G_{ik} is the polygenic background effect with $\boldsymbol{g} \sim N(0, \sigma_G^2 \boldsymbol{K})$, $\boldsymbol{g}$ is the vector of all G_{ik}, σ_G^2 is the genetic variance, $\boldsymbol{K}$ is the genomic relationship matrix, E_{ik} is the error with $\boldsymbol{e} \sim N(0, \sigma_E^2 \boldsymbol{I})$, $\boldsymbol{e}$ is the vector of all E_{ik}, σ_E^2 is the error variance, $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{g}$ and $\boldsymbol{e}$ being independent.

The genomic relationship K_{ij} between two hybrids i and j is calculated following VanRaden (2008):

$$K_{ij} = \frac{\sum_{m=1}^{M} w_{im}\, w_{jm}}{\sum_{m=1}^{M} f_m(1-f_m)} \quad (2)$$

where $w_{im} = x_{im} - f_m$ is the centered genotypic score of hybrid i at SNP m and f_m is the frequency of the alternative allele at SNP m. For the calculation of K_{ij} only, missing x_{im} values were imputed as f_m.

A second GWAS model adapted from Rincent et al. (2014) was applied and is referred to as the "Kc model" further in the text. It aimed at preventing the limitation of statistical power for large haplotypic segments showing reduced recombination and/or co-segregations with another haplotype segments due to reciprocal translocations at heterozygous state in parents. At each tested SNP m from chromosome c, the GWAS model of Eq. (1) was adapted by computing the following genomic relationship K_{ij}^c specific to chromosome c:

$$K_{ij}^c = \frac{\sum_{m \in S_c} w_{im}\, w_{jm}}{\sum_{m \in S_c} f_m(1-f_m)} \quad (3)$$

where S_c is the set of markers to be included in the calculation of K_{ij}^c, all other terms being identical to those presented in Eq. (2). Each S_c excludes all markers from its own chromosome c. When

additional chromosomes are involved in a network of reciprocal translocations with chromosome c in some parents, they were also excluded from S_c: (i) chromosomes 1, 4, and 7 were all excluded from their respective S_c because of the existence of the 1/7 and 1/4 reciprocal translocations, and (ii) chromosomes 2, 3, and 8 were all excluded from their respective S_c because of the existence of the 2/8 and 3/8 reciprocal translocations.

Model parameters were estimated using restricted maximum likelihood and the effect of each marker β^m was tested using a Wald test, both implemented in the R-package "MM4LMM" (Laporte et al., 2022) available from the CRAN. As the second GWAS model tended not to control sufficiently for false positive, an inflation factor λ was calculated as the median value of the Wald statistic over the M SNPs divided by the expected median. Following Delvin and Roeder (1999), the Wald statistic of each test was adjusted by dividing it by λ. The family-wise error rate was controlled using either (i) a Bonferroni correction by dividing the type I error ($\alpha = 5\%$) by the number of SNPs M or (ii) by applying the false discovery rate procedure of Benjamini and Yekutieli (2001) jointly to the set of p-values of all traits and GWAS methods.

For all GWAS, quantile-quantile (Q-Q) and Manhattan plots were generated. Significant SNPs were aggregated into QTL LD-based intervals using the following procedure: (i) adjacent significant SNPs were first grouped into clusters when they were less than 2Mbp apart, (ii) LD interval was calculated using the position of the first and last SNPs (significant or not) in LD of at least 0.25 with the most significant SNP of the cluster, (iii) when overlapping LD intervals were observed for a given trait, they were merged into a single interval, and (iv) LD intervals shorter than 1kbp were discarded as they likely resulted from one or few markers incorrectly positioned on the reference genome. The LD between two SNPs m and m' from chromosome c was adapted from Mangin et al. (2012) to correct for bias due to relatedness between hybrids:

$$r^2_{m,m'} = \frac{(\boldsymbol{w}_m^T \boldsymbol{K}_c^{-1} \boldsymbol{w}_{m'})^2}{(\boldsymbol{w}_m^T \boldsymbol{K}_c^{-1} \boldsymbol{w}_m)(\boldsymbol{w}_{m'}^T \boldsymbol{K}_c^{-1} \boldsymbol{w}_{m'})} \quad (4)$$

where $\boldsymbol{w}_m^T = (w_{1m}, \dots, w_{im}, \dots, w_{Nm})$ and $\boldsymbol{K}_c$ is the genomic relationship matrix of Eq. (3). Co-segregation between QTL was highlighted by computing the $r^2_{m,m'}$ between the most significant SNPs of QTL LD intervals. A representation of the consensus QTL intervals for each trait was made by merging overlapping LD intervals obtained for models K and Kc.

Ancestral origin of alleles

Among all SNPs, 40 340 presented an allele for which an ancestral origin was determined in Martin et al. (2023). Ancestral origins correspond to the following *Musa* species and genetic groups: banksii, burmannica, malaccensis, shizocarpa, zebrina and two unknown ancestral groups M1 and M2. The two unknown ancestral groups M1 and M2 were associated with the groups to which they are most closely related, i.e. schizocarpa for M1 and zebrina for M2 (Martin et al., 2023). This strategy was motivated by the fact that a large number of alleles of M1 and M2 origin were probably incorrectly attributed to schizocarpa and zebrina, respectively. This is due to the small number of M1 and M2 representatives (always admixed) that allowed these alleles to be assigned. For each detected QTL, estimated allele effects were plotted along the chromosome with colouring according to ancestral origin.

Meta-analysis

A meta-analysis of all traits and methods was performed to assess possible pleiotropic effects of identified QTLs. Using the most significant markers of each LD interval, a z-score Z_{mt} of marker m for trait t was calculated as following:

$$Z_{mt} = -\Phi^{-1}(0.5p_{mt}) \times sign(\beta_{mt}) \quad (5)$$

where $\Phi(.)$ stands for the standard Gaussian cumulative distribution function and $sign(\beta_{mt})$ is the the sign of the estimated effect of marker marker m for trait t. Note that the z-scores were calculated using the p-value and sign of the effect corresponding to those of the method (i.e. K or Kc model) by which the marker was detected. When a same marker was detected using both methods, a single z-score was calculated using the method with the most significant p-value. The effect considered was that of the allele alternative to the reference genome.

Acknowledgements

This research was supported by the Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) and Genoscope (from French Alternative Energies and Atomic Energy Commission (CEA)), the European Agricultural Fund for Rural Development (FEADER) and Région Guadeloupe through 'Plan Banane Durable 1' and 'Plan Banane Durable 2' programs, the France Génomique (ANR-10-INBS-09-08) project DYNAMO, the CGIAR Research Programme on Roots, Tubers and Bananas and the Agropolis Fondation (ID 1504-006) 'GenomeHarvest' project through the French Investissements d'Avenir program (Labex Agro: ANR- 10-LABX-0001-01). This work has been realized with the support of MESO@LR-Platform at the University of Montpellier and the technical support of the bioinformatics group of the UMR AGAP Institute, member of the French Institute of Bioinformatics (IFB) - South Green Bioinformatics Platform. We thank Sébastien Ricci for his thoughts on the experimental design. We thank Christian Vingadassalon, Frédéric Vingadassalon, Raymond Crispin, Alexin Clotaire, Ginot Karramkan, Ginot Karramkan, Gérard Numitor and Nathanaelle Leclerc for their contributions to the maintenance of the experimental set-up and phenotyping. We thank Franc-Christophe Baurens for its biomolecular support. We thank the GPTR (Great regional technical platform) of Montpellier core facility for its technical support.

Contributions

JYH, GM, FS and AD conceived the study and contributed to funding acquisition. FS, CG, FM, ClM, JMD, FL, CaM, JCE, generated breeding material, implemented the experimental design and acquired phenotypic data. CH, CC and GENOSCOPE acquired the genotypic data. SR, LT, JYH and GM performed GWAS analyses, interpreted results and wrote the first draft, which was reviewed and edited by all authors. All authors read and approved the final manuscript.

Data availability statement

All phenotypic and genotypic data underlying this study are available from the following CIRAD Dataverse repository (temporary private link to be replaced by public link upon acceptance): https://dataverse.cirad.fr/privateurl.xhtml?token=43d9d35e-49cf-4e3c-8811-d8b2c6110f68, including the vcf file "Genotyping_1129hybrids_200Ksnps.vcf.gz", the pedigree information and BLUPs for all traits in "Phenotyping_2727hybrids_24traits.tsv", and GWAS summary statistics fro the K and Kc in archives "GWAS_summary_stats_K_model.tar.gz" and "GWAS_summary_stats_Kc_model.tar.gz", respectively. The vcf file is also available on the exploration and visualization tool Gigwa: https://gigwa.cgiar.org/gigwa/?module=GWAS_agromorphotraits. Raw GBS data is available on NCBI sequence read archive: upon acceptance. The illumina data of the progenies are available in the SRA database under project PRJNA1106767 (temporary private link to be replaced by public link upon

acceptance:
https://dataview.ncbi.nlm.nih.gov/object/PRJNA1106767?reviewer=bb3v0k6dii50rqhacur0mjr2co).

Conflict of interests

The authors declare no conflict of interest.

Supplementary information

Supplementary data is available at Horticulture Research online.

Table S1: Summary table of fixed effects (cycle and block), variance components (2x, 4x, interaction 2x-4x and within-cross) and heritabilities estimated for all traits

Table S2: Co-segregations between the most significant SNPs of each QTL interval calculated using the LD formula from Eq. (4) for consensus QTLs and QTLs detected from each GWAS model. Only values above 0.05 are reported.

Table S3: Summary tables of QTL detected for all traits according to the GWAS model

Figure S1: Manhattan plots and QQ plots (before and after correction with inflation factor λ) for all traits using the K model

Figure S2: Manhattan plots and QQ plots (before and after correction with inflation factor λ) for all traits using the Kc model

Figure S3: Localization of QTL LD intervals along chromosomes for each trait and GWAS model, according to two significance thresholds: Bonferroni (-log10(p) = 6.61) and FDR (-log10(p) = 5.92). Co-segregations between the most significant SNPs of each interval are indicated by a red segment with a width proportional to the level of LD (r^2) whose values are shown in Table S2. The continuous r^2 size scale is represented by discrete values from 0.1 to 0.5. The order of the chromosomes on the x-axis was chosen so as to position the chromosomes involved in a reciprocal translocation close to each other.

Figure S4: Estimated allele effects for all traits using the K model. Dots were coloured according to allele ancestry and shaped according to the level of significance of the test. When no ancestry could be assigned, the effect represented was that of the alternative allele. The QTL interval is indicated by a grey area.

Figure S5: Estimated allele effects for all traits using the Kc model. Dots were coloured according to allele ancestry and shaped according to the level of significance of the test. When no ancestry could be assigned, the effect represented was that of the alternative allele. The QTL interval is indicated by a grey area.

References

Arruda, M. P., Lipka, A. E., Brown, P. J., Krill, A. M., Thurber, C., et al. 2016. Comparing genomic selection and marker-assisted selection for *Fusarium* head blight resistance in wheat (*Triticum aestivum* L.). Molecular Breeding, 36(7): 84.

Bakry, F., Horry, J.-P., & Jenny, C. 2021. Making banana breeding more effective. Achieving sustainable cultivation of bananas. Volume 2 - Germplasm and genetic improvement (Kema Gert H.J. (ed.), Drenth André (ed.)): 217–256. Cambridge: Burleigh Dodds Science Publishing.

Baurens, F.-C., Martin, G., Hervouet, C., Salmon, F., Yohomé, D., et al. 2019. Recombination and Large Structural Variations Shape Interspecific Edible Bananas Genomes. Molecular Biology and Evolution, 36(1): 97–111.

Belser, C., Baurens, F.-C., Noel, B., Martin, G., Cruaud, C., et al. 2021. Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. Communications Biology, 4(1): 1–12.

Benjamini, Y., & Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics, 29(4): 1165–1188.

Biabiany, S., Araou, E., Cormier, F., Martin, G., Carreel, F., et al. 2022. Detection of dynamic QTLs for traits related to organoleptic quality during banana ripening. Scientia Horticulturae, 293: 110690.

Cenci, A., Sardos, J., Hueber, Y., Martin, G., Breton, C., et al. 2021. Unravelling the complex story of intergenomic recombination in ABB allotriploid bananas. Annals of Botany, 127(1): 7–20.

Chen, A., Sun, J., Martin, G., Gray, L.-A., Hřibová, E., et al. 2023. Identification of a Major QTL-Controlling Resistance to the Subtropical Race 4 of *Fusarium oxysporum* f. sp. cubense in *Musa acuminata* ssp. malaccensis. Pathogens, 12(2): 289.

Devlin, B., & Roeder, K. 1999. Genomic control for association studies. Biometrics, 55(4): 997–1004.

Dodds, K. S., & Simmonds, N. W. 1948. Sterility and parthenocarpy in diploid hybrids of Musa. Heredity, 2(1): 101–117.

Dupouy, M., Baurens, F.-C., Derouault, P., Hervouet, C., Cardi, C., et al. 2019. Two large reciprocal translocations characterized in the disease resistance-rich *burmannica* genetic group of *Musa acuminata*. Annals of Botany, 124(2): 319-329.

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., et al. 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nature Communications, 9(1): 2638.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., & Thompson, R. 2009. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. www.vsni.co.uk.

Gupta, P. K., Kulwal, P. L., & Jaiswal, V. 2019. Chapter Two - Association mapping in plants in the post-GWAS genomics era. In D. Kumar (Ed.), Advances in Genetics, vol. 104: 75–154. Academic Press.

Herten, K., Hestand, M. S., Vermeesch, J. R., & Van Houdt, J. K. 2015. GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. BMC Bioinformatics, 16(1): 73.

Jáuregui, B., de Vicente, M., Messeguer, R. et al. 2001. A reciprocal translocation between 'Garfi' almond and 'Nemared' peach. Theoretical and Applied Genetics, 102: 1169–1176.

Karamura, E. B., & Karamura, D. A. 1995. Banana morphology - part II: the aerial shoot. In: Gowen, S. (eds) Bananas and Plantains. World Crop Series. Springer, Dordrecht

Laporte, F., Charcosset, A., & Mary-Huard, T. 2022. Efficient ReML inference in variance component mixed models using a Min-Max algorithm. PLOS Computational Biology, 18(1): e1009659.

Lescot, T. 2023. World banana production in its genetic diversity and uses. FruiTrop, 287: 100-104

Liu, X., Arshad, R., Wang, X., et al. 2023. The phased telomere-to-telomere reference genome of *Musa acuminata,* a main contributor to banana cultivars. Sci Data, 10(1): 631.

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., et al. 2012. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity, 108(3): 285–291.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal, 17(1): 10–12.

Martin, G., Carreel, F., Coriton, O., Hervouet, C., Cardi, C., et al. 2017. Evolution of the Banana Genome (*Musa acuminata*) Is Impacted by Large Chromosomal Translocations. Molecular Biology and Evolution, 34(9): 2140–2152.

Martin, G., Cardi, C., Sarah, G., Ricci, S., Jenny, C., et al. 2020a. Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. The Plant Journal, 102(5): 1008–1025.

Martin, G., Baurens, F.-C., Hervouet, C., Salmon, F., Delos, J.-M., et al. 2020b. Chromosome reciprocal translocations have accompanied subspecies evolution in bananas. The Plant Journal, 104(6): 1698–1711.

Martin, G., Cottin, A., Baurens, F.-C., Labadie, K., Hervouet, C., et al. 2023. Interspecific introgression patterns reveal the origins of worldwide cultivated bananas in New Guinea. The Plant Journal, 113(4): 802–818.

McKim, K. S., Howell, A. M., Rose, A. M. 1988. The effects of translocations on recombination frequency in *Caenorhabditis elegans*. Genetics, 120 (4): 987–1001.

Moreau, L., Charcosset, A., & Gallais, A. 2004. Experimental evaluation of several cycles of marker-assisted selection in maize. Euphytica, 137(1): 111–118.

Noumbissié, G. B., Chabannes, M., Bakry, F., Ricci, S., Cardi, C., et al. 2016. Chromosome segregation in an allotetraploid banana hybrid (AAAB) suggests a translocation between the A and B genomes and results in eBSV-free offsprings. Molecular Breeding, 36(4): 38.

Nyine, M., Uwimana, B., Swennen, R., Batte, M., Brown, A., et al. 2017. Trait variation and genetic diversity in a banana genomic selection training population. PLOS ONE, 12(6): e0178734.

Nyine, M., Uwimana, B., Blavet, N., Hřibová, E., Vanrespaille, H., 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): 170090.

Nyine, M., Uwimana, B., Akech, V., Brown, A., Ortiz, R., et al. 2019. Association genetics of bunch weight and its component traits in East African highland banana (*Musa* spp. AAA group). Theoretical and Applied Genetics, 132(12): 3295–3308.

Rincent, R., Moreau, L., Monod, H., Kuhn, E., Melchinger, A. E., et al. 2014. Recovering Power in Association Mapping Panels with Variable Levels of Linkage Disequilibrium. Genetics, 197(1): 375–387.

Risterucci, A.M., Grivet, L., N'Goran, J.A.K., Pieretti, I., Flament, M.H. and Lanaud, C. 2000. A high-density linkage map of Theobroma cacao L. Theor. Appl. Genet. 101, 948–955.

Salmon, F., Bakry, F., Efile, J. C., Ricci, S., Toniutti, L., Horry, J. P. 2023. Banana breeding at CIRAD: creating resistant new cultivars to avoid the use of pesticides. Acta Horticulturae, (1367): 201-208

752 Sardos, J., Rouard, M., Hueber, Y., Cenci, A., Hyma, K. E., et al. 2016. A Genome-Wide Association
753 Study on the Seedless Phenotype in Banana (*Musa* spp.) Reveals the Potential of a Selected Panel to
754 Detect Candidate Genes in a Vegetatively Propagated Crop. PLOS ONE, 11(5): e0154448.

755 Sham, P. C., & Purcell, S. M. 2014. Statistical power and significance testing in large-scale genetic
756 studies. Nature Reviews Genetics, 15(5): 335–346.

757 Sheperd, K., & Plantain, I. N. for the I. of B. and. 1999. Cytogenetics of the genus *Musa*.
758 https://cgspace.cgiar.org/handle/10568/104258.

759 Simmonds, N. W. 1953. Segregations in some diploid bananas. Journal of Genetics, 51(3): 458–469.

760 Stevison, L. S., Hoehn, K. B., Noor, M. A. F. 2011. Effects of Inversions on Within- and Between-
761 Species Recombination and Divergence. Genome Biology and Evolution, 3: 830–841.

762 Tadmor, Y., Zamir, D. & Ladizinsky, G. 1987. Genetic mapping of an ancient translocation in the
763 genus *Lens*. Theoretical and Applied Genetics, 73: 883–892.

764 Tomekpe, K., Jenny, C., & Escalant, J. V. 2004. A review of conventional improvement strategies for
765 Musa. Infomusa, 13(2): 2–6.

766 Toniutti, L., Rio, S., Martin, G., Hoarau, J. Y., & Salmon, F. 2023. Variation of morphological and yield
767 traits in a banana (*Musa acuminata*) breeding program of triploid populations: lessons for selection
768 procedures and criteria. Acta Horticulturae, (1362): 539–546.

769 Wang, Y., Mette, M. F., Miedaner, T., Gottwald, M., Wilde, P., et al. 2014. The accuracy of prediction
770 of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted
771 selection and is equally augmented by multiple field evaluation locations and test years. BMC
772 Genomics, 15(1): 556.

773 Wang, Z., Miao, H., Liu, J., Xu, B., Yao, X., et al. 2019. *Musa balbisiana* genome reveals subgenome
774 evolution and functional divergence. Nature Plants, 5(8): 810–821.

775 Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., et al. 2006. A unified mixed-model method
776 for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38(2): 203–
777 208.

778 Zhang, J., Song, Q., Cregan, P. B., & Jiang, G.-L. 2016. Genome-wide association study, genomic
779 prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). Theoretical and
780 Applied Genetics, 129(1): 117–130.