

Shared pedigree relationships and transmission of unreduced gametes in cultivated banana

Guillaume Martin^{1,2,*}, Franc-Christophe Baurens^{1,2}, Karine Labadie³, Catherine Hervouet^{1,2}, Frédéric Salmon^{2,4}, Franck Marius^{2,4}, Nilda Paulo-de-la-Reberdiere^{2,5}, Ines Van den Houwe⁶, Jean-Marc Aury⁷, Angélique D'Hont^{1,2} and Nabila Yahiaoui^{1,2,*}

¹CIRAD, UMR AGAP Institut, Montpellier F-34398, France, ²UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France, ³Genoscope, Institut François Jacob, CEA, Université Paris-Saclay, Evry, France, ⁴CIRAD, UMR AGAP Institut, F-97130 Capesterre-Belle-Eau, Guadeloupe, France, ⁵CIRAD, UMR AGAP Institut, CRB-PT, F-97170 Roujol Petit-Bourg, Guadeloupe, France, ⁶Bioversity International, Willem De Croylaan 42, B-3001, Leuven, Belgium and ⁷Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France.

* For correspondence. E-mail guillaume.martin@cirad.fr, nabila.yahiaoui@cirad.fr

- **Background and aims:** Cultivated bananas resulted from inter-(sub-)specific hybridizations involving *Musa* species and subspecies (*M. acuminata* subspecies, *M. schizocarpa*, *M. balbisiana*) and the subsequent selection, centuries ago, of hybrids with parthenocarpic, seedless fruits. Cultivars have low fertility and are vegetatively propagated, forming groups of somaclones. Relatively few of them, mainly triploids, are grown on a large scale and characterization of their parental relationships may be useful for breeding strategies. Here, we investigate parental relationships and gamete-type contributions among diploid and polyploid banana cultivars.
- **Methods:** We used SNP genotyping data from whole genome sequencing of 178 banana individuals including 111 cultivars, 55 wild bananas and 12 synthetic F1 hybrids. We analysed the proportion of SNP sites in accordance with direct parentage with a global statistic and along chromosomes for selected individuals.
- **Key Results:** We characterised parentage relationships for seven diploid, eleven triploid and one tetraploid cultivars. Results showed that both diploid and triploid cultivars could have contributed gametes to other banana cultivars. Diploids may have contributed 1x or 2x gametes and triploids, 1x to 3x gametes. The Mchare diploid cultivars group nowadays only found in East Africa, was found as parent of 2 diploid and 8 triploid cultivars. In five of its identified triploid offspring, corresponding to main export or locally popular dessert bananas, Mchare contributed a 2x gamete with full genome restitution without recombination. Analyses of remaining haplotypes in these Mchare offspring suggested ancestral pedigree relationships between different interspecific banana cultivars.

- **Conclusions:** Current cultivated banana resulted from different pathways of formation, with implication of recombined or un-recombined unreduced gametes produced by diploid or triploid cultivars. Identification of dessert banana's parents and the type of their contributed gametes should support the design of breeding strategies.

Key words: *Musa acuminata*, *Musa balbisiana*, *Musa schizocarpa*, banana cultivars, Mchare, parentage, unreduced gamete, polyploid, whole genome approach.

Accepted Manuscript

INTRODUCTION

Cultivated bananas (*Musa sp.*) resulted from a complex process of natural inter-(sub-)specific hybridization (Martin *et al.*, 2020b, Simmonds, 1962, Perrier *et al.*, 2011, Martin *et al.*, 2023). They are vegetatively propagated and only a very limited number of these cultivars are grown on a large scale. For example, the commercial dessert Cavendish banana represents more than 50% of the world production (Lescot, 2020). Banana culture faces major biotic stresses, including the fungal disease caused by a *Fusarium sp.* lineage named TR4 that currently threatens banana cultures (Viljoen *et al.*, 2020). There is thus a crucial need for breeding disease resistant bananas while maintaining agro-morphological and fruit quality traits. However, breeding strategies are hampered by the sterility or very low fertility of cultivars which limits crossing abilities and progeny number. In this context, knowledge of cultivar parentage would represent a very useful resource for breeding.

Main domestication traits of bananas relate to fruit edibility, with the selection by humans of plants with parthenocarpic seedless fruits, and therefore with low fertility. Three main species are involved in their formation process: *Musa acuminata* (A genome, $2n=2x=22$), *Musa balbisiana* (B genome, $2n=2x=22$) and *Musa schizocarpa* (S genome, $2n=2x=22$). So far, the A genome was present in all analysed banana cultivars, except the small group of Fe'i cultivars that derived only from *Musa* species of the Australimusa series (T genome, $2n=2x=20$) (Jarret *et al.*, 1992, Simmonds and Shepherd, 1955, Martin *et al.*, 2023).

Hybridizations involving *M. schizocarpa* and *M. acuminata* subspecies (ssp. *banksii* and possibly ssp. *zebrina*) in the New Guinea Island are proposed as the starting point of banana domestication (Martin *et al.*, 2023). Along with diffusion of early cultivars throughout Southeast Asia, additional hybridizations occurred with other *Musa* species and subspecies such as *M. acuminata* ssp. *zebrina*, *malaccensis*, *burmannica*, *halabanensis*, *Musa*

balbisiana and at least one unknown contributor that might be an uncharacterized *M. acuminata* subspecies (Martin *et al.*, 2020b, Sardos *et al.*, 2022, Martin *et al.*, 2023). Large chromosomal rearrangements (*i.e.* translocation, inversion) present in some wild *Musa* were transmitted to many cultivars (Martin *et al.*, 2020a). This context of inter-(sub-)specific hybridization may have favoured production of 2x gametes from diploid hybrids, leading to the formation of triploid cultivars (Simmonds, 1962, Perrier *et al.*, 2011). The current diversity of cultivars includes diploids and triploids with different global genomic combinations (*e.g.* AA, AB, AAA, AAB, ABB, AAT) modulated by interspecific recombination and with an A genome introgressed by *M. schizocarpa* (Simmonds and Shepherd, 1955, Baurens *et al.*, 2019, Cenci *et al.*, 2021, Nemeckova *et al.*, 2018, Martin *et al.*, 2023). Within these genomic groups, cultivars are classified in distinct subgroups, that are thought to derive each from one seed and centuries or millennia of vegetative propagation and thus represent subgroups of phenotypically different somaclonal mutants (Perrier *et al.*, 2011, Simmonds, 1954). The most grown cultivars subgroups are triploid, including AAA dessert bananas such as Cavendish, AAA cooking East African Highland bananas or AAB Plantain bananas.

The complex hybridization processes at the origin of cultivars would be difficult to rapidly reproduce in breeding programs. The identification and use of direct parents of successful cultivars could thus be useful to facilitate banana breeding (Raboin *et al.*, 2005). Several parentages have been proposed based on analyses with low density markers (RFLP and SSR) involving in particular the diploid Mchare cultivar subgroup (previously named Mlali) (Raboin *et al.*, 2005, Hippolyte *et al.*, 2012, Perrier *et al.*, 2009). Higher density markers should allow confirming parentage relationships at whole genome scale, and should provide additional information on cultivar pedigree and on gamete transmission processes.

The aim of this study was to use high density single nucleotide polymorphism (SNP) genotyping information from 178 banana individuals to identify potential parents of cultivars from our sample. Special emphasis was placed on the contribution of the Mchare cultivar subgroup and the parental relationships between Mchare-derived cultivars. This work also addresses questions on gamete transmission type leading to major cultivars and the potential role of triploids as parents to other triploids.

MATERIALS AND METHODS

Materials

A set of 178 banana individuals was first considered for this analysis (**Supplementary data Table S1**). These individuals correspond to wild *M. acuminata* (A genome, $2n=2x=22$), *M. schizocarpa* (S genome, $2n=2x=22$), *M. balbisiana* (B genome, $2n=2x=22$) and *Australimusa* spp. accessions (T genome, $2n=2x=20$), to cultivars derived from these species, and to 12 synthetic F1 hybrids generated at the CIRAD breeding platform by crosses between parents also represented in the dataset (**Supplementary data Table S1**). SNP-based genotyping data (not phased) of these individuals were extracted from a larger vcf file obtained from high coverage Illumina sequencing data used to perform chromosome ancestry painting, described in (Martin *et al.*, 2023). SNP positions are located along the DH-Pahang V4 reference genome (Belser *et al.*, 2021). All used scripts are available in the vcfhunter toolbox (<https://github.com/SouthGreenPlatform/VcfHunter>). Compared to the original file, a new genotype calling was performed (using TotalRecal.1.0.py), for accession DB_Pisang_Awak that was tetraploid instead of triploid according to the (Martin *et al.*, 2023) analysis. Among the 178 individuals, 26 were similar to other accessions (*i.e.* somaclones or duplicates of other accessions). For these similar accessions (**Supplementary data Table S1**), only one was reported in results tables. Representations of genome ancestry mosaic painting were

previously published (Martin *et al.*, 2023; <https://banana-genome-hub.southgreen.fr/node/50/1598445>), except for newly generated results of deduced haplotypes (see below).

Method

Two types of SNP data analysis were performed (**Supplementary data** Fig. S1). The first one consisted in the identification of triplets of individuals that match a direct parents-child trio relationship (*i.e.* the genotype from an individual (2x, 3x or 4x) defined as child is compatible with a combination of two individuals (2x, 3x or 4x) defined as parents when looking at their genotypes). The second one consisted in the identification of pairs of individuals that match a direct parent-child duo relationship involving one individual from the diploid Mchare clonal subgroup as a parent (*i.e.* the genotype from an individual (2x, 3x or 4x) defined as child is compatible with Mchare as parent when looking at the genotype). Both analyses were performed in two steps: i) calculation of the global proportion of SNP sites in accordance with each tested trio or duo; ii) validation of identified trios or duos through local analysis along chromosomes of the proportion of SNP sites in accordance with such trios or duos.

Identification of potential parents-child trios through global SNP analysis

This analysis was performed using the ValPar.py tool, added to the vcfhunter toolbox. The tool works as follows: (i) For each potential parent, all possible gametes were generated. These gametes were all combinations of k elements, with repetition and without order, sampled among n non-redundant alleles found at the studied position in the potential parent (k being the ploidy of the potential gamete). For example, for a position where a parent is ATT, the different possible diploid gametes are: AA, AT and TT. If the potential child was triploid, two values of n were tested for each parent, *i.e.* n=1 and n=2. (ii) For each potential

parent pair, all possible zygote allele combinations were generated by combining possible gametes from both parents (only gamete combinations giving the correct zygote ploidy were generated). (iii) The allele combination found in the potential child was searched for, among the possible zygote allele combinations from the parents. (iv) The proportion of sites in which the potential child allele combination matched one of the possible zygote allele combinations from the parent pair was then calculated for each potential parent pair. Ploidy was considered: for example, if the potential child was triploid, two final values were reported; one with parent 1 being the 1x gamete donor and parent 2 being the 2x gamete donor and a second one with parent 1 being the 2x gamete donor and parent 2 being the 1x gamete donor.

To save computation time, accessions without *Australimusa* ancestry were analysed using a filtered vcf including only polymorphic sites in the 167 accessions from *Musa acuminata*, *Musa schizocarpa* and/or *Musa balbisiana* origin. The filtered vcf contained 6 867 490 polymorphic sites. Accessions with an *Australimusa* ancestry were analysed using a filtered vcf including only polymorphic sites in the 178 individuals (the 167 A, S and/or B accessions + the 11 *Australimusa* or *Australimusa* hybrid accessions) (**Supplementary data Table S2**). The filtered vcf contained 8 285 170 polymorphic sites.

The statistic was calculated for each trio with a cultivated accession as a child (**Supplementary data Table S3**). The following criteria were used to consider parental relationships as valid: (i) The proportion of sites in accordance with the tested parentage should be equal to or higher than 0.999 which is the minimal value observed in the 12 synthetic parents-child trios (**Supplementary data Table S4**). (ii) Among couples of parents that validated this threshold only those with the minimal cumulative ploidy were retained. The reason behind this is that we considered that diploids were more likely to be parents compared to triploids.

Identification of potential Mchare contribution through global SNP analysis

The contribution of Mchare was specifically analysed by looking for a potential 1x and 2x Mchare gamete restitution on all individuals of the dataset.

According to Martin *et al.*, (2023), Mchare ancestral contributors were only from *Musa acuminata* ssp. and *Musa schizocarpa*. Thus, only polymorphic sites between and within these genetic groups were kept for the analysis. For this, sites containing private alleles from other species (*i.e.* *Musa balbisiana* or *Australimusa*, **Supplementary data** Table S2) were identified and removed using `IdentPrivateAllele.py`, `allele_ratio_group.py` and `vcfSelect.py` as previously described (Martin *et al.*, 2023). If not removed, such sites would generate an important proportion of sites in accordance with the Mchare contribution only because they are homozygous in all individuals from *Musa acuminata* and/or *Musa schizocarpa* origin. The resulting vcf file contained 5 412 666 polymorphous SNP sites.

The analysis was then performed in several steps (**Supplementary data** Fig. S1, purple steps): (i) Selecting, for each duo (Mchare, potential child), only sites that are polymorphic within and/or between individuals of the duo using `vcfFilter.1.0.py`. (ii) Calculating, the proportion of sites validating a potential 1x contribution of Mchare to the potential child using the `ACRO.py` tool added to `vcfhunter` toolbox (**Supplementary data** Table S5). (iii) In case of a polyploid potential child, calculating the proportion of sites validating a potential 2x (recombined or un-recombined) contribution of Mchare to the potential child using the `ACRO.py` tool (**Supplementary data** Table S5). (iv) In case of polyploid potential child, looking for the proportion of sites validating a complete Mchare genome (*i.e.* un-recombined 2x gamete named hereafter 2x^c) in the potential child using `vcfRemove.1.0.py` (**Supplementary data** Table S5).

The summary statistics obtained from steps (ii, iii and iv) were used to identify potential duos with a Mchare contribution. For this, duos that have a proportion of SNP sites in accordance with a 1x or 2x contribution greater than 99% were selected (**Supplementary data** Table S5). This value was selected based on an observed shift in the summary statistics values (**Supplementary data** Table S5).

Validation of parents-child trios through local SNP analysis along chromosomes

To validate selected parents-child trios, the distribution of the proportion of shared SNP was inspected along reference chromosomes. This was performed in several steps (**Supplementary data** Fig. S1, green steps): (i) Selecting, in the vcf file, the SNP sites that are polymorphic within and/or between individuals of the trio. (ii) Calculating from the obtained vcf file, the proportion of sites in accordance with the trio on windows of 201 SNP size. Sites in accordance were calculated as described in the trio identification section above. (iii) Filtering the vcf in order to keep, for each parent-child duo, polymorphic SNPs within and between individuals of the duo and calculating, from the resulting vcf file, the proportion of sites in accordance with the duo on window sizes of 201 SNPs. This calculation considers the gamete ploidy of the tested parent. (iv) In parallel, the number of alleles shared between the parent and the child is calculated to test for complete genome restitution. (v) In cases where the complete genome of the parent is found in the child, the alleles from this parent were removed from the child's genotype in the vcf obtained at (i). This allowed access to the gamete given by the other parent. (vi) the remaining genotype(s) were compared to the genotype of the second potential parent, to analyse the proportion of sites in accordance with such parentage. This analysis was performed using the APAR.py tool that was added to the vcfhunter toolbox.

For ease of representation, the proportions of sites in discordance with tested parentages were displayed along each chromosome and all results were visualised using Circos (Krzywinski *et al.*, 2009). Configuration and files required for Circos visualisation were automatically generated by the APAR.py tool.

Validation of parent-child duos through local SNP analysis along chromosomes

Validation of parent child duos (**Supplementary data** Fig. S1, blue steps) was performed as follows: (i) For each duo, selection in the vcf of sites that are polymorphic within and between individuals of the duos using vcfFilter.1.0.py. (ii) Looking along chromosomes, for the proportion of SNP sites validating a 1x contribution of the potential parent to the potential child using the ACRO.py tool. (iii) In case of polyploid potential child, looking along chromosomes, for the proportion of SNP sites validating a potential 2x contribution of the potential parent to the potential child using the ACRO.py tool. (iv) In case of polyploid potential child, looking along chromosomes for the proportion of sites validating a complete parental genome in the potential child using vcfIdent.1.0.py tool of vcfhunter toolbox. These proportions were calculated along chromosomes on windows of 201 SNPs size.

The proportions of sites in discordance with tested parentages were represented along each chromosome and results were visualised using Circos (Krzywinski *et al.*, 2009) and the DrawCircos.py tool added to the vcfhunter toolbox.

In silico chromosome ancestry painting of remaining haplotype(s) in polyploid progeny where complete genome restitution of a parent was identified

The genotype of the parent was removed from the polyploid child in the vcf using the vcfRemove.1.0.py tool. Remaining alleles were considered as a haplotype, representative of

the gamete from the second parent. The *in silico* chromosome ancestry painting of the remaining haplotype(s) corresponding to the gamete from the second parent was performed using the ancestry specific alleles and the process described in (Martin *et al.*, 2023). The process was automated with the SPRH.py tool added to the vcfhunter toolbox (**Supplementary data** Fig. S1).

Obtained mosaics for Mchare-derived AAB cultivars were compared to published banana mosaics (Martin *et al.*, 2023) by looking for similar ancestry mosaic patterns or compatible mosaics that could derive from recombination between parental haplotypes.

Management of aneuploidy

The triploid Lady Finger (Nadan) accession had a supernumerary chromosome 08. To manage this aneuploidy, the genotype calling of Lady Finger along chromosome 08 was recalculated with a ploidy of 4 using the TotalRecal.1.0.py script of vcfhunter toolbox. For this accession, the analyses described previously (with exception of the global one) were performed separately on triploid chromosomes and on chromosome 08 and results were pooled. For tetraploid Calypso, aneuploid regions were identified by analysing read coverage of SNP sites along chromosomes, using the vcf2cov.py program added to the vcfhunter toolbox.

RESULTS

Global parentage analysis identifies one or the two parents for several cultivars

Parentage relationships were searched among 178 banana individuals (**Supplementary data** Table S1) using SNP genotyping data obtained by (Martin *et al.*, 2023) from whole genome Illumina sequencing. The dataset included 55 wild accessions representing *M. acuminata* ssp. (A genome), *M. schizocarpa* (S genome), *M. balbisiana* (B genome) and *Australimusa* spp.

(T genome), 111 cultivars derived from these species and 12 F1 diploid hybrids for which both parents were known. Banana cultivars belonging to a subgroup correspond to somaclonal variants that are currently indistinguishable at the genomic level. In order to facilitate the reading of the manuscript, when relevant, we will refer to the name of the subgroup (**Supplementary data** Table S1) rather than to the name of the accession.

Ten parents-child trios were identified, as they showed a global proportion of SNP sites in accordance with the trio, that was equal to or higher than the proportion (99,9%) observed for the 12 F1 synthetic parents-child trios (Table 1, **Supplementary data** Table S3). Three additional parents-child trios that involved *Australimusa* species were identified, however, two of them may correspond to clonal relationship between cultivars and the third to similarity between two wild accessions (**Supplementary data** Table S3).

The targeted search for parental contributions of the Mchare cultivar subgroup (**Supplementary data** Table S5) yielded a total of 13 Mchare-child duos (Table 1). Three of these 13 duos (involving Cavendish, Mnalouki, Colatina Ouro) were already identified among the parents-child trios. The duos involved two diploid, ten triploid and one tetraploid accessions. Among them, six accessions (five triploids and one tetraploid) had more than 99.6% of sites in accordance with Mchare being the 2x gamete donor and more than 99.4% of sites suggested that both haplotypes of Mchare are found in each of these 6 accessions (**Supplementary data** Table S5).

SNP analysis along chromosomes validates most candidate parents from the parents-child trios and parent-child duos analysis

The proposed parentages for cultivars (10 trios and 13 duos) were further investigated by local analysis of shared SNPs along their chromosomes. Two additional trios proposed by the literature (Perrier et al., 2009, Hippolyte et al., 2012) with Cavendish cultivars as a child,

Mchare as 2x gamete donor and diploid cultivars Pisang Madu or Pisang Pipit as 1x gamete donor were also tested. The proportion of SNP in discordance along the chromosomes was calculated and represented together with the *in silico* chromosomal ancestry painting (Fig. 1). Fig. 1A represents the tested trio Gros Michel (child), Mchare (2x parent) and Khai Nai On (1x parent). The proportion of SNP in discordance was close to zero along all chromosomes, validating this parents-child trio. In contrast, the two trios proposed by literature with Cavendish as child showed large regions, on nearly all chromosomes, in which the proportion of SNP in discordance was relatively high (Fig. 1B, **Supplementary data** Fig. S2A) indicating that the direct trio parentages were not valid.

All ten trios suggested by the global analysis were validated by this approach (Fig. 1A, 1C, **Supplementary data** Fig. S2 B to I). For two of them, a peak of discordance was observed on a small region on chromosome 08 (Fig. 1C, **Supplementary data** Fig. S2D), probably due to small differences (gene conversion, aneuploidy) resulting from divergence after vegetative propagation over centuries or problems of read mapping and variant calling. Among the ten validated trios, three distinct trios were proposed with Galeo as child (in italics in Table 1). This impossible situation is further investigated in the next section of results.

The seven remaining validated trios, revealed different types of crosses involving diploids and triploids as parents of diploid or triploid cultivars (Table 1, Fig. 1, **Supplementary data** Fig. S2). Fig. 1C illustrates a case where a Plantain (AAB triploid genome) is proposed as a 2x gamete donor for another triploid cultivar. All identified parents are cultivars except in one case where a wild *Musa balbisiana* accession (PKW) is a parent of the triploid cultivar Saba (Table 1, **Supplementary data** Fig. S2).

Similar analyses were performed with the 13 predicted Mchare-duos (including three already present in trios) and validated 11 of them (Table 1). A 2x contribution of the Mchare was validated for five triploid cultivars: Nadan (Fig. 2A), Gros Michel (**Supplementary data** Fig. S3A), Cavendish (**Supplementary data** Fig. S3B), Nendra Padaththi (**Supplementary data** Fig. S3C) and Pome (**Supplementary data** Fig. S3D). A 2x contribution of the Mchare was also found in tetraploid cultivar Calypso (**Supplementary data** Fig. S3E). A 1x contribution of the Mchare was validated for two diploid cultivars (Paka (**Supplementary data** Fig. S3F) and Colatina Ouro (**Supplementary data** Fig. S3G)) and three triploid cultivars (Koja (**Supplementary data** Fig. S3H), Pisang Ambon (**Supplementary data** Fig. S3I) and Mnalouki (**Supplementary data** Fig. S3J)). In Pisang Ambon (**Supplementary data** Fig. S3I), a small region at the beginning of chromosome 01 was discordant for a 1x Mchare contribution but corresponded to a region where Pisang Ambon is aneuploid with a ploidy of 2x instead of 3x (Martin *et al.*, 2023). The segment of Mchare origin was likely lost from this region during years of vegetative propagation of Pisang Ambon.

Finally, for two triploids, Hom Thong Mokho (Fig. 2B) and Wh-O-Gu (**Supplementary data** Fig. S3K), regions of discordance, even with a 1x contribution of the Mchare were observed suggesting that these accessions were not directly derived from Mchare accessions but are related to them.

Complete 2x and 3x gamete transmission among cultivated banana

The proportions of shared SNPs along chromosomes were analysed in the seven predicted cases of 2x gamete transmission by a diploid accession (Table 1) to precise the nature of the transmitted 2x gametes. One case involved the diploid Tjau Lagada that transmitted a recombined 2x gamete to Pisang Papan (Fig. 2C). In six cases, a complete diploid Mchare genome restitution was detected (Fig. 2A, **Supplementary data** Fig. S3 A to

E). These cases included the tetraploid accession Calypso which is an improved cultivar derived from a cross between Highgate, a dwarf clone of the triploid Gros Michel cultivar and a diploid individual (Borges *et al.*, 2014). The Mchare contribution to Calypso identified in this duo is due to the presence of the Mchare genome in Gros Michel. The test of the complete genome restitution of Gros Michel along chromosomes of Calypso showed that only one region at one extremity of chromosome 10 was not in accordance with such complete genome restitution (Fig. 2D). Calypso is aneuploid for a few chromosomal regions, including a missing chromosomal region at one extremity of chromosome 10 (Fig. 2D, **Supplementary data** Fig. S4). The results were thus in accordance with a complete 3x gamete restitution of Gros Michel to Calypso.

A particular situation was found for the Galeo diploid cultivar that was proposed as child in three distinct trios involving the Palang triploid accession, as one parent, and either of the diploid accessions Khai Nai On, Sinwobogi or SF265 as the second parent (Table 1). These three diploid potential parents are not somaclones as they do not have the same ancestral mosaic genome (Martin *et al.*, 2023), which raised questions about the process generating these trios. Further examination of the trios involving Galeo revealed that the complete genome of Galeo was present in triploid Palang (**Supplementary data** Fig. S5A). This led us to propose that Galeo may not be the child but rather the parent of these accessions and transmitted a 2x gamete to Palang and a 1x gamete to Khai Nai On, Sinwobogi and SF265 (Table 1). The first proposed trios are a consequence of the presence of the Galeo genome in Palang which allows for recombinant haplotypes complementing those present in Khai Nai On, Sinwobogi and SF265 to generate Galeo (**Supplementary data** Fig. S5B). Access to real haplotypes (phased data from genetic studies or long read sequencing technologies) of these individuals should allow validating or rejecting this hypothesis.

In silico extraction and analysis of the remaining 1x gamete from Mchare derived accessions

For accessions derived from a complete gamete restitution (2x or 3x), it is possible to deduce the complementary 1x gamete. In most Mchare-derived polyploid accessions resulting from such gametes, the parental origins of the second parent were not identified (Table 1). To learn more about the second parent, alleles from the complete genome donor (Mchare or Gros Michel) were removed from the child genotypes to deduce the gamete from the second parent. These gametes were predicted to be 1x except for Nadan where the deduced gamete had 2 copies of chromosome 8, suggesting a 1x + 1 gamete. These deduced gametes were then processed as described in (Martin *et al.*, 2023) to obtain their genome ancestry mosaic (**Supplementary data** Fig. S6). This revealed a *M. balbisiana* introgressed status of the second parents of AAB cultivars Nadan, Nendra Padaththi and Pome. The comparison of the genome ancestry mosaic of deduced gametes of these three AAB cultivars with the ancestral mosaics of A/B hybrid cultivars obtained in (Martin *et al.*, 2023) suggested potential similarity between their 1x gamete donors and 2 ABB triploid cultivars (Monthan and Ney Mannan, Fig. 3, **Supplementary data** Fig. S6, Fig. 4). Nadan showed potential additional similarity with Saba and Pome showed potential additional similarity with Saba, Pelipita and Pisang Kelat (**Supplementary data** Fig. S6). This led us to investigate potential shared pedigree between these cultivars. As illustrated for chromosome 8 (Fig. 3A, B, C), for all three AAB cultivars (Pome, Nendra Padaththi and Nadan), large regions of agreement but also large regions of discordance were observed when parentage of their deduced gamete with Ney Mannan and Monthan was tested. Such patterns were visible on several chromosomes (**Supplementary data** Fig. S6) suggesting close pedigree relationships. In contrast, comparison with Saba, Pelipita and Pisang Kelat did not show such large regions of

agreement (Fig. 3B and C, **Supplementary data** Fig. S6), excluding very close pedigree relationships.

The Fig. 3 also illustrates some similar agreement/discordance patterns when comparing the deduced gametes from the AAB cultivars to Ney Mannan and Monthan. This suggests that all tested AAB cultivars and the ABB Ney Mannan and Monthan are related to each other.

The mosaics of the other deduced gametes also revealed a *M. acuminata* ssp. *malaccensis* and ssp. *banksii* hybrid as second parent of Calypso and a more complex hybrid status of the second parents of Cavendish and Gros Michel (Fig. 4). As expected from validated trios, the remaining haplotype of Gros Michel matched a potential 1x gamete from Khai Nai On (**Supplementary data** Fig. S6K). The remaining haplotype of Cavendish, showed large accordance/discordance segments with Pisang Pipit or Pisang Madu as 1x gamete donors suggesting that the 1x gamete donor of Cavendish is a close relative of both accessions (**Supplementary data** Fig. S6 L and M).

DISCUSSION

We analysed the direct parentage between 178 banana individuals including 111 cultivars and 55 wild accessions and characterised parentage relationships for seven diploid cultivars, 11 triploid cultivars and one tetraploid, some of them representing major subgroups of dessert banana. These results highlighted that diploids but also triploids cultivars could be parents of other diploid, triploid or tetraploid cultivars by contributing 1x, 2x or 3x gametes. A targeted analysis of the Mchare subgroup parentage relationships showed an important contribution of this subgroup to banana cultivars (Fig. 4).

The Mchare contributions to banana cultivars

The Mchare bananas form a phenotypically diverse subgroup of AA diploids. They are nowadays only found in some East African regions and Islands where they are particularly appreciated and culturally important (Perrier *et al.*, 2019). We showed that Mchare contributed 1x gametes to 2 diploid and 3 triploid cultivars, and also 2x gametes to 5 triploid cultivars of our sample. It is also present in tetraploid Calypso through the Gros Michel contribution (Fig. 4). The Gros Michel dessert banana was very popular in the first half of the previous century, but being susceptible to the Panama disease (*Fusarium oxysporum* fsp. *cubense* Race 1), it was replaced by Cavendish. We confirmed that both cultivars resulted from the transmission of a 2x gamete from Mchare (Raboin *et al.*, 2005, Hippolyte *et al.*, 2012, Perrier *et al.*, 2009, Martin *et al.*, 2020b) and demonstrated that in both cases the complete Mchare genome was transmitted. In addition, we validated the implication of Mchare as 2x gamete donor in the AAB cultivars Nadan, Pome/Prata and Nendra Padaththi (Hippolyte *et al.*, 2012, Perrier *et al.*, 2009) with additional information on complete genome transmission but we excluded the 2x contribution of Mchare to Hom Thong Mokho (Perrier *et al.*, 2009). Differences between our results and previous studies are likely due to the higher genotyping density and to the methodology that we used to validate parentage along chromosomes. However, we cannot completely exclude differences between studied individuals from distinct collections or resulting from miss-labelling. In addition, we also showed that Mchare contributed a 1x gamete to the triploid cultivars Mnalouki (AAB from Comoros), Koja (AAA from Comoros) and Pisang Ambon (AAA) and to diploid cultivars Paka (AA from East Africa) and Colatina Ouro (AA) (Fig. 4). Mchare clones have contributed their genomes to cultivars of diverse origins, including cultivars found in Africa but also triploid dessert bananas that are predicted to originate from Southeast Asia. Thus,

Mchare most probably originated in Southeast Asia and were transported by humans to Africa (Perrier *et al.*, 2019).

In all cases in which the Mchare was proposed as a 2x gamete donor, a full genome of the Mchare was found. This is the case for the two main dessert AAA banana cultivar subgroups Cavendish and Gros Michel and also for AAB cultivars from Pome/Prata, Nandan and Nendra Padaththi subgroups. One can hypothesise that a complete genome of Mchare brings highly favourable agro-morphological trait combinations which could explain selection of individuals resulting from complete Mchare gamete restitution. An alternate hypothesis is that this genome is prone to perform a full gametic restitution thereby increasing the probability to find such contribution to cultivated bananas. A combination of both these hypotheses could also be considered.

Other pedigree relationships

Second parents in Mchare-derived cultivars were various (Fig. 4). We confirmed that the second gamete present in Gros Michel was a 1x gamete from Khai Nai On (Hippolyte *et al.*, 2012, Raboin *et al.*, 2005). Moreover, we showed that the Galeo accession may be a grand-parent of Gros Michel through Khai Nai On. We excluded Pisang Pipit or Pisang Madu as 1x gamete donors to Cavendish (Hippolyte *et al.*, 2012, Perrier *et al.*, 2009), although the presence of large regions consistent with the proposed relationship suggested some shared pedigree.

Other second parents to Mchare-derived cultivars included cultivars as diverse as diploid Sucrier and triploid Plantains and also unknown A/B hybrids. For AAB Nandan, Nendra Padaththi and Pome, we showed that their unknown A/B hybrids second parents were related to ABB Ney Mannan and Monthan and that they transmitted 1x gametes (or in the case of Nandan an aneuploid 1x+1 gamete) (Fig. 4). Note that ABB cultivars from the

subgroup named Bluggoe were found similar at the genome mosaic level to cultivars from subgroup Monthan (**Supplementary data** Table S1; (Cenci *et al.*, 2021, Martin *et al.*, 2023)). In this study, we also suggested that Saba cultivars derived from Monthan. Thus, the AAB cultivars Nadan, Nendra Padaththi, Pome and ABB cultivars Monthan, Ney Mannan, Bluggoe and Saba are related. This is coherent with their shared A/B recombination breakpoints on chromosome 9 (Martin *et al.*, 2023).

In this study, all identified parental relationships only involved cultivars as parents to one exception. This can be explained by the fact that cultivars have a fixed genotype conserved through vegetative propagation, which allows the identification of natural hybridization events that occurred centuries ago. Conversely, most wild parents involved in hybridizations some centuries ago may not exist anymore, due to sexual reproduction. The only exception we found is a wild individual, Pisang Klutuk Wulung (PKW) that was proposed as a parent of triploid cultivar Saba. The PKW accession as well as some accessions of *M. balbisiana* are popular in Indonesia for various uses and are vegetatively propagated (Ahmad *et al.*, 2014). This may explain that we could still identify parentage involving Pisang Klutuk Wulung.

Different types of gametes and triploid individuals are involved in the formation of many banana cultivars

Several cases of transmission of un-recombined or recombined 2x-gametes from diploid to triploid cultivars were identified. They involved mainly Mchare cultivars but also Tjau Lagada and likely Galeo. More surprisingly, in few cases, a triploid was proposed as parent of another triploid or diploid cultivar, either as a 2x or as a 1x gamete donor. Banana cultivars either diploid or triploid are poorly fertile and sometimes totally sterile. Crosses have been undertaken since the 1930s in order to exploit cultivar's residual fertility in

breeding programs. This often required a huge number of pollinations. Cytogenetic analysis of obtained progenies showed that diploid bananas generally produce 1x gametes while triploids may produce from 1x up to 6x gametes (Larter, 1948, Wilson, 1946c, Wilson, 1946a, Wilson, 1946b, Shepherd, 1999). These features are exploited in breeding programs to generate tetraploids by crossing triploid cultivars with diploids. Triploids could thus also potentially be involved through natural hybridization in the formation of other cultivars. However, this should be rare given their poor levels of fertility. Based on nuclear and cytoplasmic genome information, a possible implication of triploid cultivars in the formation of other cultivars has been proposed (Carreel *et al.*, 2002, De Langhe *et al.*, 2010). Here, we identified such cases. For two triploid cultivars (ABB Saba and AAB Mnalouki), we identified a triploid parent (ABB or AAB, respectively) that could have contributed a 2x gamete. In addition, we observed one trio and one duo in which a 1x and 3x gamete respectively could have been provided by AAA triploids. This illustrates that such a process may occur in different types of triploid banana. However, we cannot exclude that a yet not identified diploid accession could be the parent of both the child and the triploid parent proposed in our parents-child trio analysis.

Previous studies of progeny obtained by crossing triploid Gros Michel with different diploids showed that the main class of obtained offspring was tetraploid (Wilson, 1946c, Larter, 1948). Based on breeding behaviour and cytological observations, it was proposed that triploid gametes were the main class of viable gametes produced by Gros Michel (Larter, 1948, Wilson, 1946c). Our analysis along chromosomes showed that the tetraploid cultivar Calypso resulted from the transmission of an un-recombined 3x gamete from the triploid Gros Michel. One region of discrepancy on chromosome 10 may be explained by aneuploidy in the triploid parent (a Gros Michel clone named Highgate, (Borges *et al.*, 2014)) and/or in Calypso compared to the accession Gros Michel we have used. A triploid Plantain was also

previously shown to have transmitted a recombined 3x gamete (Supplementary Material in Baurens *et al.*, 2019).

Unreduced gametes are generally more frequent in interspecific hybrids and in allopolyploids, and result in most cases, from disturbed meiotic processes, although pre-meiotic or post-meiotic doubling have also been mentioned (Mason and Pires, 2015, De Storme and Geelen, 2013). If the first division step of meiosis does not occur, meiosis is similar to mitosis and produces un-recombined unreduced gametes, retaining full parental heterozygosity (De Storme and Geelen, 2013). Other disturbances during meiosis such as cytological alterations during the second division step (*e.g.* parallel spindles, (d'Erfurth *et al.*, 2008)), or absence of second division, can result in unreduced gametes with recombined chromosomes (De Storme and Geelen, 2013). The inter(sub)specific origins of cultivated bananas often resulted in the presence of large chromosomal rearrangements at an heterozygous state in their genome (Martin *et al.*, 2020a). This may disturb meiotic processes and/or lead to the production of nonviable, unbalanced 1x gametes that could favour the transmission of unreduced gametes. For example, in the case of Mchare, the presence of two independent heterozygous large reciprocal translocations (Martin *et al.*, 2020a) will statistically generate 75% of unbalanced gametes during meiosis, which should be non-viable. This phenomenon mathematically increases the proportion of 2x gametes in the pool of viable gametes.

Prospects for banana breeding

This study illustrated that different processes are involved in the formation of banana cultivars involving various types of gametes. They include recombined or un-recombined 2x gametes from diploids and various types of gametes including 3x un-recombined gametes from triploids. These peculiar gametes must be rare and thus are potentially selected to

conserve particularly favourable hybrid genomic combinations that required several generations to be obtained.

In this context, pre-breeding strategies that aim at maintaining such favourable genomic combinations, while introgressing desirable traits such as resistance, could be favoured. For example, the important contribution of the Mchare subgroup to other cultivars, often through un-recombined 2x gamete, suggested that its genomic combination should be preserved as much as possible. High throughput genotyping technologies may be used to select individuals with the targeted genomic combination. These strategies, as others, will be complicated by the poor fertility of bananas but conversely may be particularly useful in this context.

SUPPLEMENTARY DATA

Supplementary data are available online at <https://academic.oup.com/aob> and consist of the following. Fig. S1: Flowchart and scripts used to perform parentage analyses. Fig. S2: Validation along chromosomes of predicted parent-child trios. Fig. S3: Validation along chromosomes of predicted parent child duos and transmitted gamete types. Fig. S4: Calypso Illumina reads coverage along chromosomes of DH-Pahang V4. Fig. S5: Analysis of Galeo contribution to cultivars. Fig. S6: Local proportions of SNP sites in discordance with AAB/ABB tested parentage relationships. Table S1: Accession information. Table S2: List of accessions used for global trio analysis and of accessions used for identification of *M. balbisiana* and *Australimusa* private alleles sites for the Mchare analysis. Table S3: Global parentage trio analysis: proportions of SNP sites in accordance with trios. Table S4: Global parentage trio results for synthetic F1 trios. Table S5: Global analysis of Mchare contribution.

ACKNOWLEDGMENTS

The authors thank the CRB Plantes Tropicales Antilles CIRAD-INRA Guadeloupe France and the Bioversity International Transit Centre (<https://www.bioversityinternational.org/banana-genebank/>) for providing the original plant materials, Xavier Perrier for contributing DNA; Caroline Belser for her contribution to the DYNAMO sequencing project; Chantal Guiougou, Jean-Marie Delos and Jean-Claude Efile for their contribution to banana crosses and embryo rescue, and Françoise Carreel and Simon Rio for critical reading of the manuscript. This work has been realized with the support of MESO@LR-Platform at the University of Montpellier, the financial support of CIRAD, 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. The filtered

vcf file is available from <https://banana-genome-hub.southgreen.fr/> in the Download section.

Scripts were added to the vcfhunter toolbox (<https://github.com/SouthGreenPlatform/VcfHunter>).

FUNDING

This work was supported by the France Génomique (ANR-10-INBS-09-08) project DYNAMO, the Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), the Agropolis Fondation (ID 1504-006) ‘GenomeHarvest’ project through the French Investissements d’Avenir program (Labex Agro: ANR- 10-LABX-0001-01).

Accepted Manuscript

LITERATURE CITED

- Ahmad F, Megia R, Poerba YS. 2014.** Genetic diversity of *Musa balbisiana* Colla in Indonesia based on AFLP marker. *HAYATI Journal of Biosciences*, **21**: 39-47.
- Baurens FC, Martin G, Hervouet C, Salmon F, Yohome D, Ricci S, Rouard M, Habas R, Lemainque A, Yahiaoui N, D'Hont A. 2019.** Recombination and large structural variations shape interspecific edible bananas genomes. *Molecular Biology and Evolution*, **36**: 97-111.
- Belser C, Baurens FC, Noel B, Martin G, Cruaud C, Istace B, Yahiaoui N, Labadie K, Hribova E, Dolezel J, Lemainque A, Wincker P, D'Hont A, Aury JM. 2021.** Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. *Communications Biology*, **4**: 1047.
- Borges CV, Amorim VB, Ramlov F, Ledo CA, Donato M, Maraschin M, Amorim EP. 2014.** Characterisation of metabolic profile of banana genotypes, aiming at biofortified *Musa* spp. cultivars. *Food Chemistry*, **145**: 496-504.
- Carreel F, Gonzalez de Leon D, Lagoda P, Lanaud C, Jenny C, Horry JP, Tezenas du Montcel H. 2002.** Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. *Genome*, **45**: 679-92.
- Cenci A, Sardos J, Hueber Y, Martin G, Breton C, Roux N, Swennen R, Carpentier SC, Rouard M. 2021.** Unravelling the complex story of intergenomic recombination in ABB allotriploid bananas. *Annals of Botany*, **127**: 7-20.
- d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Simon M, Jenczewski E, Mercier R. 2008.** Mutations in *AtPSI* (*Arabidopsis thaliana* parallel spindle 1) lead to the production of diploid pollen grains. *PLoS Genetics*, **4**: e1000274.

- De Langhe E, Hribova E, Carpentier S, Dolezel J, Swennen R. 2010.** Did backcrossing contribute to the origin of hybrid edible bananas? *Annals of Botany*, **106**: 849-57.
- De Storme N, Geelen D. 2013.** Sexual polyploidization in plants--cytological mechanisms and molecular regulation. *The New phytologist*, **198**: 670-684.
- Hippolyte I, Jenny C, Gardes L, Bakry F, Rivallan R, Pomies V, Cubry P, Tomekpe K, Risterucci AM, Roux N, Rouard M, Arnaud E, Kolesnikova-Allen M, Perrier X. 2012.** Foundation characteristics of edible *Musa* triploids revealed from allelic distribution of SSR markers. *Annals of Botany*, **109**: 937-51.
- Jarret RL, Gawel N, Whitemore A, Sharrock S. 1992.** Rflp-Based Phylogeny of *Musa* Species in Papua-New-Guinea. *Theoretical and Applied Genetics*, **84**: 579-584.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009.** Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**: 1639-45.
- Larter LNH. 1948.** Hybridism in *Musa*. *Journal of Genetics*, **49**: 69-85.
- Lescot T. 2020.** Diversité génétique de la banane. *Fruitrop*, **269**: 98-102.
- Martin G, Baurens F-C, Hervouet C, Salmon F, Delos J-M, Labadie K, Perdereau A, Mournet P, Blois L, Dupouy M, Carreel F, Ricci S, Lemainque A, Yahiaoui N, D'Hont A. 2020a.** Chromosome reciprocal translocations have accompanied subspecies evolution in bananas. *The Plant Journal*, **104**: 1698-1711.
- Martin G, Cardi C, Sarah G, Ricci S, Jenny C, Fondi E, Perrier X, Glaszmann JC, D'Hont A, Yahiaoui N. 2020b.** Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *The Plant Journal*, **102**: 1008-1025.
- Martin G, Cottin A, Baurens FC, Labadie K, Hervouet C, Salmon F, Paulo-de-la-Reberdiere N, Van den Houwe I, Sardos J, Aury JM, D'Hont A, Yahiaoui N.**

- 2023.** Interspecific introgression patterns reveal the origins of worldwide cultivated bananas in New Guinea. *The Plant Journal*, **113**: 802-818.
- Mason AS, Pires JC. 2015.** Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends in Genetics*, **31**: 5-10.
- Nemeckova A, Christelova P, Cizkova J, Nyine M, Van den Houwe I, Svacina R, Uwimana B, Swennen R, Dolezel J, Hribova E. 2018.** Molecular and cytogenetic study of East African Highland banana. *Frontiers in Plant Science*, **9**: 1371.
- Perrier X, Bakry F, Carreel F, Jenny C, Horry JP, Lebot V, Hippolyte I. 2009.** Combining biological approaches to shed light on the evolution of edible bananas *Ethnobotany Research and Applications*, **7**: 199-216.
- Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry JP, Jenny C, Lebot V, Risterucci AM, Tomekpe K, Doutrelepon H, Ball T, Manwaring J, de Maret P, Denham T. 2011.** Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proceedings of the National Academy of Sciences of the U S A*, **108**: 11311-8.
- 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**: 19-36.
- Raboin LM, Carreel F, Noyer JL, Baurens FC, Horry JP, Bakry F, Montcel HTD, Ganry J, Lanaud C, Lagoda PJJ. 2005.** Diploid ancestors of triploid export banana cultivars: Molecular identification of 2n restitution gamete donors and n gamete donors. *MOLECULAR BREEDING*, **16**: 333-341.
- 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**.

- Shepherd K. 1999.** *Cytogenetics of the genus Musa*. Montpellier, France: INIBAP
- Simmonds NW. 1954.** Varietal identification in the Cavendish group of bananas. *Journal of Horticultural Science*, **29**: 81-88.
- Simmonds NW. 1962.** *The evolution of the bananas*. London Great Britain: Longmans.
- Simmonds NW, Shepherd K. 1955.** The taxonomy and origins of the cultivated bananas. *Journal of the Linnean Society of London. Botany*, **55**: 302-312.
- Viljoen A, Ma L-J, Molina A. 2020.** Fusarium Wilt (Panama disease) and monoculture in banana production: resurgence of a century-old disease. In: J. B. Ristaino and A. Records e, ed. *Emerging Plant Diseases and Global Food Security*: The American Phytopathological Society.
- Wilson GB. 1946a.** Cytological studies in the Musae. 11. Meiosis in some diploid clones. *GENETICS*, **31**: 475-482.
- Wilson GB. 1946b.** Cytological studies in the Musae. 111. Meiosis in some seedling clones. *GENETICS*, **31**: 483-493.
- Wilson GB. 1946c.** Cytological studies in the Musae. I. Meiosis in some triploid clones. *GENETICS*, **31**: 241-258.

Figure Legends

Fig. 1: Validation along chromosomes of predicted parents-child trios. Accession chromosome ancestry mosaics obtained from (Martin *et al.*, 2023) and the local proportion of sites in discordance with tested trios are represented. Child pseudo-haplotypes are represented on outer circles and are separated from tested parents' pseudo-haplotypes (on inner circles) by the local proportion of alleles in discordance with the tested parentage (value between 0 and 1). Ploidy of the parental tested gamete is indicated between square brackets and the ploidy of the child is indicated between brackets. Colour codes “banksii”, “burmannica”, “zebrina”, “malaccensis”, “truncata”, “sumatrana” and “halabanensis” stand for an origin from *M. a. ssp banksii/microcarpa/errans*, *burmannica*, *zebrina*, *malaccensis*, *truncata*, *sumatrana* and *halabanensis* respectively. “schizocarpa” and “balbisiana” stand for *Musa schizocarpa* and *M. balbisiana* respectively. “M_2” is an uncharacterized contributor to banana and “Unknown” corresponds to regions in which no origin could be attributed. **A, B, C**, correspond to three tested trios whose names are indicated in each figure centre.

Fig. 2: Validation along chromosomes of predicted parent-child duos and transmitted gamete types. Accession chromosome ancestry mosaics obtained from (Martin *et al.*, 2023) and the local proportion of sites in discordance with tested duos and gamete types are represented. Child pseudo-haplotypes are represented on outer circles and are separated from tested parent pseudo-haplotypes (on inner circles) by the local proportion of alleles in discordance with tested parentage (value between 0 and 1). Depending on the duos tested, different type of parentage could be tested: a haploid gamete restitution (1x), a diploid gamete restitution (2x), a diploid gamete with complete genome restitution (2x^c) or a triploid gamete with complete genome restitution (3x^c). Ploidy of the parental tested gamete is indicated between square brackets and the ploidy of the child is indicated between brackets.

Colour codes are identical to those in Fig. 1. **A, B, C, D**, correspond to four tested duos whose names are indicated in each figure centre.

Fig. 3: Patterns of local proportions of SNP sites in discordance with AAB/ABB tested parentage relationships on chromosome 08. Analysis was performed on deduced gametes from AAB Nendra Padaththi (a), Pome (b) and Nadan (c) after removal of the Mchare genotype that has been found complete in these individuals. These deduced gametes were compared to potential parents Ney Mannan and Monthan (a, b and c), Saba (b and c), Pisang Kelat and Pelipita (b) that showed compatible mosaics with deduced gametes. Black curves represent the local proportion of alleles in discordance with tested parentage (value between 0 and 0.5). Coloured horizontal bars at the top of each subfigure represent the deduced gamete ancestry mosaic for Nendra Padaththi (a), Pome (b) and Nadan (c). The horizontal bars under each curve represent the mosaic from the corresponding tested parent obtained from (Martin *et al.*, 2023). Colour codes are identical to those in Fig. 1.

Fig. 4: Contribution of Mchare to cultivated bananas and associated pedigree relationships. The contribution of identified parents is indicated by black arrows. The type of gamete contribution is indicated: 1x, haploid gamete restitution; 2x, diploid gamete restitution; 2x^c, diploid gamete with complete genome restitution; 3x^c, triploid gamete with complete genome restitution. Chromosome ancestry mosaics of Mchare (Martin *et al.*, 2023) and of the deduced gametes produced by the second parents of Mchare-derived cultivars are shown with bar plots. Colour codes are identical to those in Fig. 1. *, aneuploid chromosome 8 in Nadan.

Table 1: Proposed banana parents-child trios and parent-child duos

	Child	Child group	Parent 1 (P1)	P1 group	Parent 2 (P2)	P2 group	Gamete ploidy		Supported by local analysis
							P1	P2	
Identified trios	IDN 077	AA	Sucrier	AA	Ibota	AAA	1x	1x	yes
	Pisang Berlin	AA	Sucrier	AA	Tjau Lagada	AA	1x	1x	yes
	Pisang Papan	AAA	Tjau Lagada	AA	Sucrier	AA	2x	1x	yes
	Saba	ABB	Monthan	ABB	PKW*	BB (wild)	2x	1x	yes
	Colatina Ouro	AA	Mchare	AA	Sucrier	AA	1x	1x	yes
	Mnalouki	AAB	Mchare	AA	Plantain	AAB	1x	2x	yes
	Gros Michel	AAA	Mchare	AA	Khai Nai On	AA	2x ^c	1x	yes
	<i>Galeo</i>	AA	<i>Palang</i>	AAA	<i>Sinwobogi</i>	AA	1x	1x	yes [!]
	<i>Galeo</i>	AA	<i>Palang</i>	AAA	<i>SF265</i>	AA	1x	1x	yes [!]
	<i>Galeo</i>	AA	<i>Khai Nai On</i>	AA	<i>Palang</i>	AAA	1x	1x	yes [!]
Identified additional duos	Cavendish	AAA	Mchare	AA	?	-	2x ^c	-	yes
	Nadan	AAB	Mchare	AA	?	-	2x ^c	-	yes
	Nendra Padaththi	AAB	Mchare	AA	?	-	2x ^c	-	yes
	Pome	AAB	Mchare	AA	?	-	2x ^c	-	yes
	Koja	AAA	Mchare	AA	?	-	1x	-	yes
	Paka	AA	Mchare	AA	?	-	1x	-	yes
	Pisang Ambon	AAA	Mchare	AA	?	-	1x	-	yes
	<i>Calypso</i>	AAAA	<i>Mchare</i>	AA	?	-	2x ^c	-	yes [!]
	Wh-O-Gu	AAA	Mchare	AA	?	-	1x	-	no
re-analysed ^a	Hom Thong Mokho	AAA	Mchare	AA	?	-	1x	-	no
	Calypso	AAAA	Gros Michel	AAA	?	-	3x ^c	-	yes
	Palang	AAA	Galeo	AA	?	-	2x ^c	-	yes
	Khai Nai On	AA	Galeo	AA	?	-	1x	-	yes
	Sinwobogi	AA	Galeo	AA	?	-	1x	-	yes
	SF265	AA	Galeo	AA	?	-	1x	-	yes

^a: Parentage re-analysed in section "Complete 2x and 3x gamete transmission among cultivated banana" of results.

^c: Indicate that a complete genome of the parent is found in the child

*: PKW, Pisang Klutuk Wulung

!: indicates re-analysed, re-interpreted trios and duos (lines in italics)

Figure 1

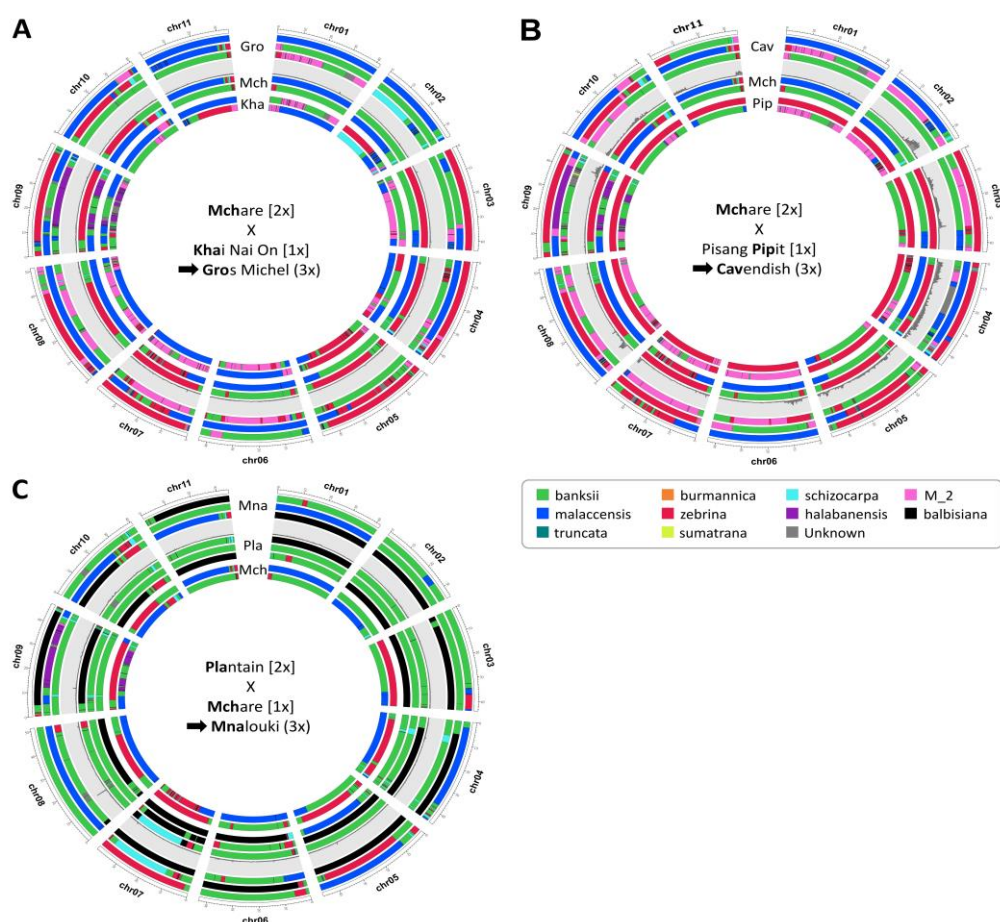


Figure 2

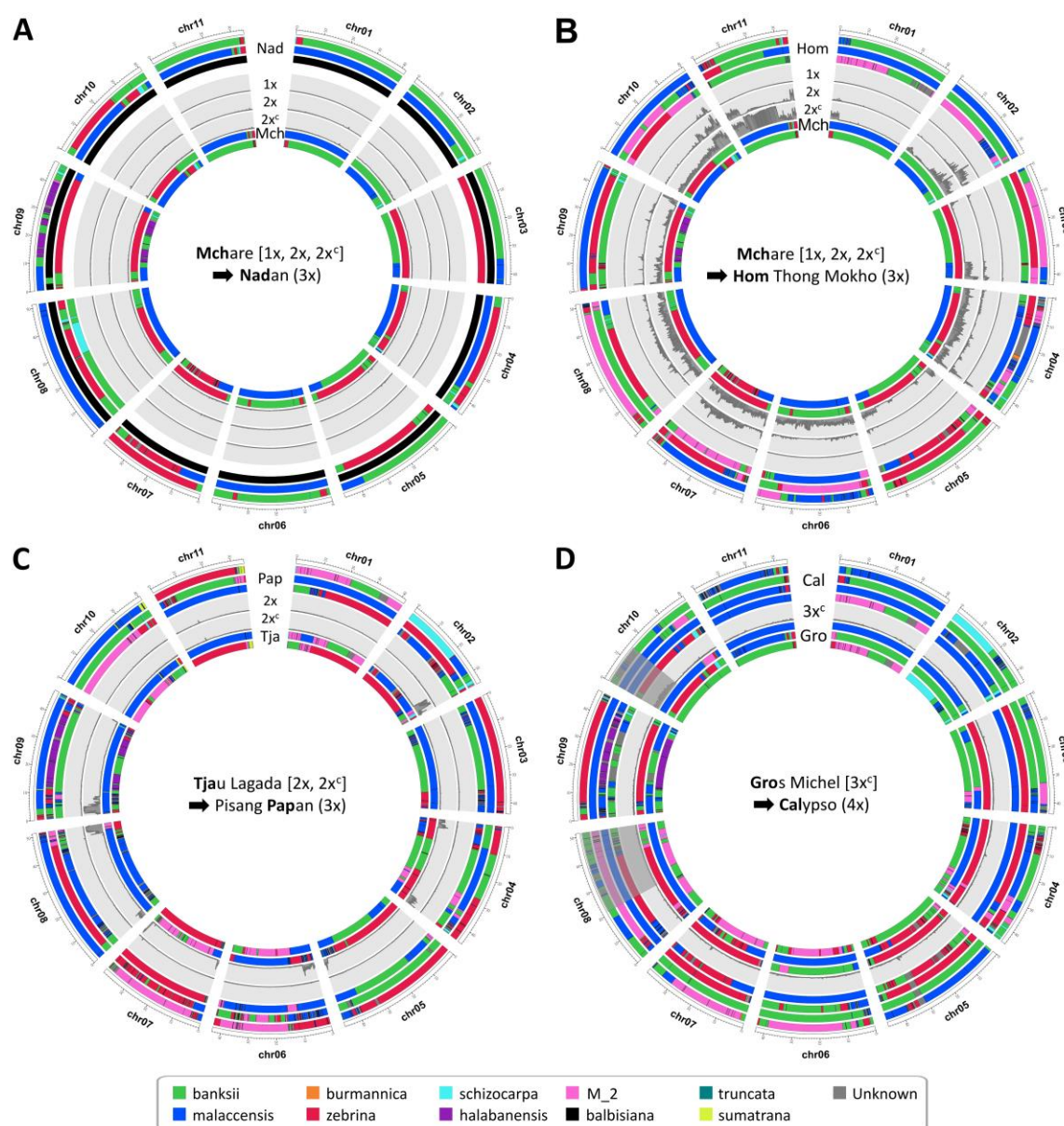


Figure 3

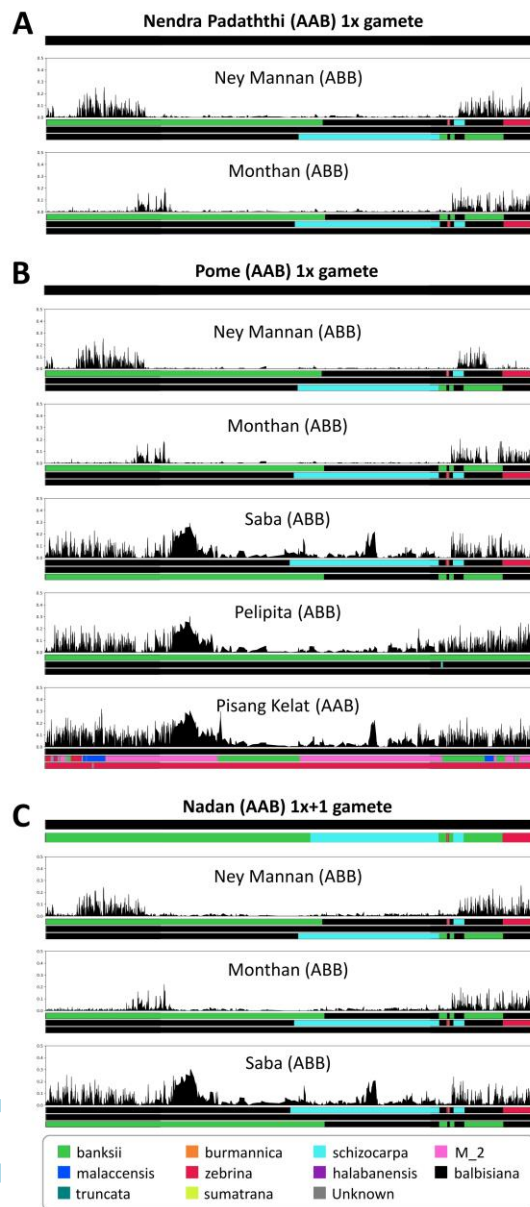


Figure 4

