the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2023) doi: 10.1111/tpj.16086

Interspecific introgression patterns reveal the origins of worldwide cultivated bananas in New Guinea

Guillaume Martin[1,2,*], Aurélien Cottin[1,2], Franc-Christophe Baurens[1,2], Karine Labadie[3], Catherine Hervouet[1,2], Frédéric Salmon[2,4], Nilda Paulo-de-la-Reberdiere[2,5], Ines Van den Houwe[6], Julie Sardos[7], Jean-Marc Aury[8], Angélique D'Hont[1,2] and Nabila Yahiaoui[1,2,*]

[1]*CIRAD, UMR AGAP Institut, Montpellier F-34398, France*
[2]*UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France*
[3]*Genoscope, Institut François Jacob, CEA, Université Paris-Saclay, Evry, France*
[4]*CIRAD, UMR AGAP Institut, F-97130 Capesterre-Belle-Eau, Guadeloupe, France*
[5]*CIRAD, UMR AGAP Institut, CRB-PT, F-97170 Roujol Petit-Bourg, Guadeloupe, France*
[6]*Bioversity International, Willem De Croylaan 42, B-3001, Leuven, Belgium,*
[7]*Bioversity International, Parc Scientifique Agropolis II, 34397, Montpellier, France, and*
[8]*Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France*

Received 23 August 2022; revised 16 December 2022; accepted 23 December 2022.
*For correspondence (e-mail guillaume.martin@cirad.fr; nabila.yahiaoui@cirad.fr).

SUMMARY

Hybridizations between *Musa* species and subspecies, enabled by their transport via human migration, were proposed to have played an important role in banana domestication. We exploited sequencing data of 226 *Musaceae* accessions, including wild and cultivated accessions, to characterize the inter(sub)specific hybridization pattern that gave rise to cultivated bananas. We identified 11 genetic pools that contributed to cultivars, including two contributors of unknown origin. Informative alleles for each of these genetic pools were pinpointed and used to obtain genome ancestry mosaics of accessions. Diploid and triploid cultivars had genome mosaics involving three up to possibly seven contributors. The simplest mosaics were found for some diploid cultivars from New Guinea, combining three contributors, i.e., *banksii* and *zebrina* representing *Musa acuminata* subspecies and, more unexpectedly, the New Guinean species *Musa schizocarpa*. Breakpoints of *M. schizocarpa* introgressions were found to be conserved between New Guinea cultivars and the other analyzed diploid and triploid cultivars. This suggests that plants bearing these *M. schizocarpa* introgressions were transported from New Guinea and gave rise to currently cultivated bananas. Many cultivars showed contrasted mosaics with predominant ancestry from their geographical origin across Southeast Asia to New Guinea. This revealed that further diversification occurred in different Southeast Asian regions through hybridization with other *Musa* (sub)species, including two unknown ancestors that we propose to be *M. acuminata* ssp. *halabanensis* and a yet to be characterized *M. acuminata* subspecies. These results highlighted a dynamic crop formation process that was initiated in New Guinea, with subsequent diversification throughout Southeast Asia.

Keywords: genome ancestry, *Musa* spp., hybridization, domestication, genetic diversity.

INTRODUCTION

Characterizing the contribution of crop plant progenitors is important to understand crop domestication, provide relevant resources for breeding, and inform conservation strategies. Bananas (*Musa* spp.) are cultivated throughout subtropical and tropical regions worldwide, mainly for their fruits, which can be consumed raw as dessert bananas or cooked as a starch-rich staple food. They are one of several crops (including sugarcane [*Saccharum officinarum*] and yams [*Dioscorea* spp.]; Denham et al., 2020) that were domesticated in a region extending from Southeast Asia (SEA) to Melanesia. In banana, the main features selected by humans are related to fruit edibility, i.e., the absence of seeds in the fruit and the development of pulp-enriched fruits without fertilization (parthenocarpy) (Perrier et al., 2011; Simmonds, 1962). Due to their reduced fertility, naturally occurring vegetative propagation has been used to perpetuate and

disperse cultivars (Denham et al., 2020; Perrier et al., 2011).

The genus *Musa* is divided into two clades. One clade is comprised of former sections *Ingentimusa*, *Callimusa*, and *Australimusa* ($2n = 14$, 18, or 20), and the second clade is comprised of former sections *Eumusa* and *Rhodochlamys* ($2n = 22$) (Häkkinen, 2013; Li et al., 2010). Hybridizations between several *Musa* species and subspecies from the second clade were proposed to have played an important role in banana domestication (Perrier et al., 2011; Simmonds, 1962). One of these species, *Musa acuminata* (A genome), is fundamental in the cultivar formation process (Simmonds, 1962). It is sometimes combined with *Musa balbisiana* (B genome), resulting in a conventional cultivar classification based on different global genomic groups (e.g., AA, AAA, AB, AAB, ABB; Baurens et al., 2019, Cenci et al., 2021, Simmonds & Shepherd, 1955). A third species, *Musa schizocarpa* (S genome, $2n = 22$) from New Guinea, was proposed to be involved in a few minor cultivars (AS genome) and in East African Highland Bananas (Carreel et al., 2002; Nemeckova et al., 2018). Fe'i cultivars found in New Guinea and Polynesia and some rare interspecific cultivars are derived from species from the first clade (former *Australimusa* section, T genome, $2n = 20$; De Langhe et al., 2009).

The species *M. acuminata* has diverged into several subspecies, whose numbers and classification status vary throughout the literature and databases (Häkkinen & Väre, 2008; Perrier et al., 2009; Shepherd, 1990; Simmonds, 1962). Different molecular marker analyses have shown the involvement of the subspecies *M. acuminata* ssp. *banksii* from New Guinea, *M. acuminata* ssp. *zebrina* from Java, and *M. acuminata* ssp. *malaccensis* from Malaysia and Sumatra in the genome of banana cultivars, and minor contributions of the *burmannica* group (Boonruangrod et al., 2008; Boonruangrod et al., 2009; Carreel et al., 2002; Christelová et al., 2017; Perrier et al., 2009; Sardos, Perrier, et al., 2016). The *burmannica* group is comprised of the three subspecies *M. acuminata* ssp. *burmannicoides*, *M. acuminata* ssp. *burmannica*, and *M. acuminata* ssp. *siamea* (Southeast India and Myanmar to Vietnam), which have been proposed to correspond to one genetic group (Dupouy et al., 2019; Martin, Cardi, et al., 2020; Perrier et al., 2009). The described *M. acuminata* diversity also includes *M. acuminata* ssp. *truncata* (Malaysia), *M. acuminata* ssp. *microcarpa* (Borneo), *M. acuminata* ssp. *errans* (Philippines), and several subspecies or botanical varieties from Indonesia (Häkkinen & Väre, 2008; Nasution, 1989; Shepherd, 1990). Accessions from the *burmannica* group, *M. acuminata* ssp. *malaccensis*, *M. acuminata* ssp. *zebrina*, and also *M. balbisiana* bear different large chromosomal reciprocal translocations that emerged in these genetic groups and were transmitted to many cultivars (Martin, Baurens, et al., 2020; Shepherd, 1999).

Morphological and molecular marker analyses of banana diversity associated with archeological and linguistic approaches led to a domestication model outlined by Perrier et al. (2011). It proposes that during the Holocene, pre-domesticated forms of banana originating from *M. acuminata* subspecies have been selected for various uses and transported by humans out of their natural range across the New Guinea and SEA region (Kennedy, 2009; Perrier et al., 2011). Hybridizations with local *M. acuminata* subspecies in different contact zones resulted in edible diploids (Perrier et al., 2011). Disturbed meiotic processes in these edible diploids, linked to their hybrid status, sometimes resulted in the production of diploid gametes, which in turn led to the formation of edible triploids. Long-term and large-scale vegetative propagation of the main selected cultivars resulted in a secondary diversification attributed to somaclonal variation (cultivar subgroups).

The observation that some diploid cultivars from Papua New Guinea (PNG) may be directly derived from *M. acuminata* ssp. *banksii* suggested a different domestication process on the island of New Guinea, where the hypothesis of inter(sub)specific hybridization as a main driving force of banana domestication may not apply (Sardos, Perrier, et al., 2016; Sardos, Rouard, et al., 2016).

New approaches based on sequencing data were recently developed to precisely define ancestral contributions in interspecific hybrids through the characterization of ancestry mosaics along the genome (Baurens et al., 2019; Cenci et al., 2021; Martin, Cardi, et al., 2020).

Here, with these approaches, we analyzed whole genome sequencing data from a large and diverse sample of wild and cultivated banana to gain further insight into the genetic make-up of cultivated bananas and their domestication process. We identified single nucleotide polymorphism (SNP) markers that enabled the characterization of ancestral contributions along banana chromosomes, including contributions of two unknown ancestors. Phylogenetic analysis and screening of publicly available sequencing data were carried out to better characterize the two postulated unknown ancestors. The obtained genome mosaics of cultivars, analyzed in relation to known geographical origins, shed new light on cultivated banana domestication and formation processes.

RESULTS

Identifying the ancestral genetic groups that contribute to cultivars and their informative alleles

We generated DNA sequencing data from 150 Musaceae accessions and 12 F1 diploid hybrids. High-coverage genome sequencing data of 25 banana accessions were retrieved from the National Center for Biotechnology Information (NCBI) to complement this dataset (Table S1). The diversity sample of 175 accessions (excluding F1 individuals) included 63

representatives of wild *Musa* species and subspecies (i.e., seedy bananas), 62 diploid and 49 triploid or tetraploid cultivars (i.e., parthenocarpic and seedless) representing different areas of origin or economically important subgroups, and one outgroup species (*Ensete ventricosum*).

A total of 42 011 682 biallelic SNP variant sites were obtained after read mapping on the *M. acuminata* reference genome V4 (Belser et al., 2021) and filtration steps. As segmental aneuploidy can occur in banana accessions (Breton et al., 2022), coverage information of the SNPs was used to detect deviations from expected ploidy along chromosomes. A total of 40 accessions had aneuploid regions (Table S2). Such events may be due to vegetative/*in vitro* propagation or structural differences between ancestral genomes. In one case, triploid accession Lady Finger, an entire supernumerary chromosome 8 was detected.

The SNPs were used to identify genetic groups contributing to cultivars (i.e., ancestral groups) and their corresponding representative alleles in a four-step procedure (Figures S1 and S2; Methods S1–S4). This procedure was set up to take into account the presence of introgressions in some wild representatives, particularly within *M. acuminata*, and the implication of unknown ancestors (Martin, Cardi, et al., 2020).

In a first step, multivariate analysis and allele clustering were performed with SNPs from representatives of *Musa* species and subspecies that were previously reported to be contributors to different banana cultivars. This led to the identification of alleles representing six genetic groups, which we named based on the main (sub) species from which they were derived: *balbisiana* (*M. balbisiana*), *schizocarpa* (*M. schizocarpa*), *zebrina* (*M. acuminata* ssp. *zebrina*), *malaccensis* (*M. acuminata* ssp. *malaccensis*), *burmannica* (*M. acuminata* ssp. *burmannica*, *burmannicoides*, and *siamea*), and *banksii* (*M. acuminata* ssp. *banksii*, *M. acuminata* ssp. *microcarpa* accession Borneo, and *M. acuminata* ssp. *errans* accession Agutay).

In a second step, because the clustering approach could not be applied to larger numbers of genetic groups, the potential contribution of the other sampled *Musa* species and subspecies was evaluated with a different approach. The proportion of alleles present in these accessions but not in the six groups of the first step was calculated. The analysis of shared proportions of these alleles among wild and cultivated accessions allowed the identification of three additional ancestral groups: *Australimusa* (all *Australimusa* spp.), *sumatrana* (*M. acuminata* ssp. or var. *sumatrana*, hereafter referred to as ssp. *sumatrana*), and *truncata* (*M. acuminata* ssp. *truncata*). No contributions have been identified from seven *Musa* spp. (*Musa coccinea*, *Musa itinerans*, *Musa ornata*, *Musa velutina*, *Musa sanguinea*, *Musa rosea*, and *Musa laterita*). In a third step, a private allele approach was used to identify representative alleles, for the three additional ancestral groups, for *E. ventricosum* used as outgroup, and to refine the set of *M. balbisiana* alleles.

A preliminary analysis of ancestry along chromosomes using representative alleles of the nine *Musa* ancestral groups and the outgroup *Ensete* (chromosome ancestry painting) was then performed. It showed that large chromosomal regions not corresponding to aneuploid regions and present in many cultivars and a few hybrid wild *M. acuminata* accessions were not assigned to an ancestral group, thereby supporting the hypothesis of unknown ancestral contributors (Martin, Cardi, et al., 2020).

A fourth step thus involved the identification of representative alleles for these unknown ancestral groups using an iterative approach mainly based on three diploid accessions ('Pisang Jari Buaya', 'Pisang Madu', and 'EN13') that displayed, in the preliminary analysis, chromosome-scale regions of unknown origin (Method S4; Figure S2). This led to the identification of ancestry informative alleles from two unknown ancestries named M_1 and M_2. During this process, we noted that in contrast to M_1, no sampled accession showed a complete haploid chromosome set of M_2 origin.

Ultimately, a total of 6 562 307 alleles representing 12 genetic groups (11 banana ancestral groups and one outgroup) were obtained. They have been used for the determination of ancestry informative allele ratios along chromosomes of the studied accessions and for drawing their chromosome ancestry mosaic. All mosaics are presented in Figure S3 together with identified aneuploid regions.

Chromosome ancestry mosaics of wild accessions

Examples of chromosome ancestry mosaics of wild *Musa* genetic groups that contributed to cultivar genomes are presented in Figure 1a in relation to their geographical distribution. The cumulative proportions of the genome covered by each ancestry for all wild accessions of these groups are presented in Figure 1b. Accessions are numbered (Table S1) and indicated in between brackets in the text. Accessions from *M. balbisiana* or *Australimusa* spp. were found homogenous for their respective genetic groups. Most *M. acuminata* subspecies and *M. schizocarpa* accessions although globally homogenous displayed a few interspersed small segments of different *M. acuminata* origin (e.g., accessions 29, 30, and 37; Figure 1a; Figure S3). In a few cases, introgressed segments of several megabases have been identified (e.g., 26–28 and 34; Figure 1a,b; Figure S3). Some segments in (peri)centromeric regions may be overpredicted due to the complexity of such regions. The *M. acuminata* ssp. *errans* and *M. acuminata* ssp. *microcarpa* accessions (40 and 41; Figure 1b; Figure S3) showed an expected global *banksii* origin with many interspersed small segments from other

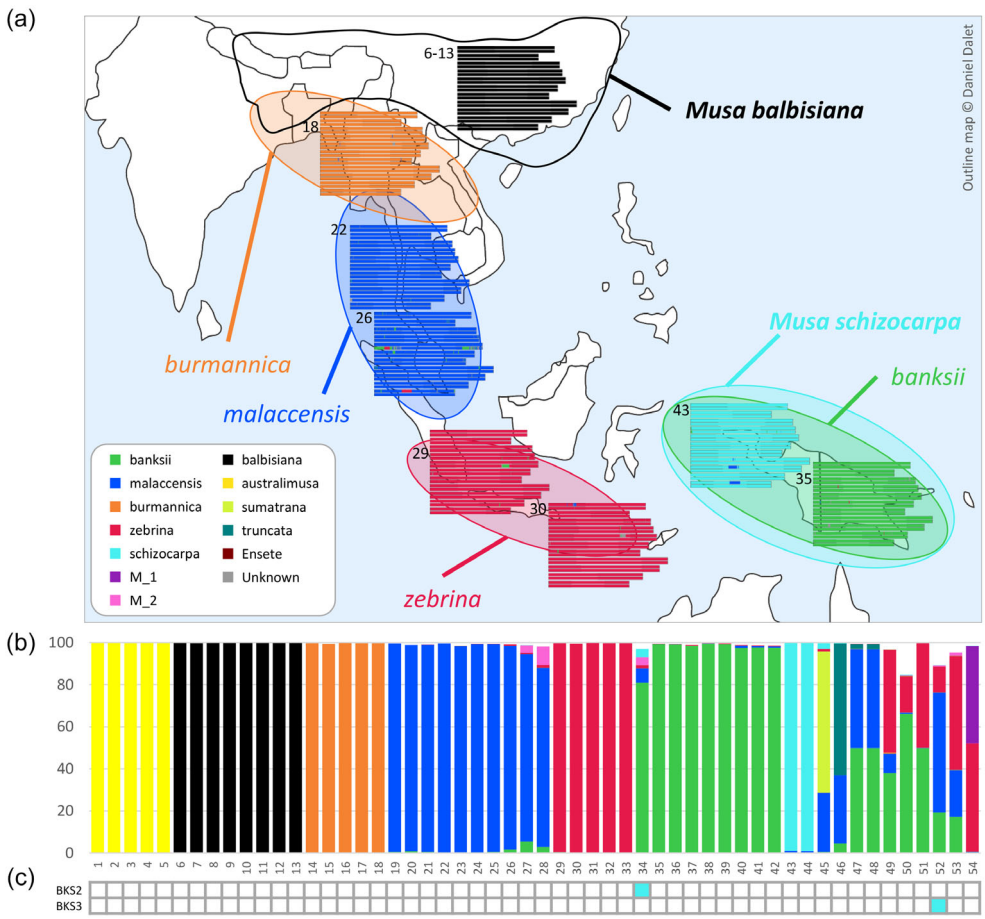

Figure 1. Diversity of wild banana genome ancestry in Southeast Asia (SEA) and New Guinea (NG).
(a) Genome mosaics of several wild accessions representing main genetic groups contributing to banana cultivars. i.e., *M. acuminata* ssp. *banksii*, *M. acuminata* ssp. *zebrina*, *M. acuminata* ssp. *malaccensis*, *M. acuminata* ssp. *burmannica*/*siamea*, *M. balbisiana*, and *M. schizocarpa*. Ancestral group contributions are presented along the 11 *Musa* reference chromosomes by different colors. The representation is based on non-phased genotyping data. Numbers on the left refer to: 6–13, *M. balbisiana* accessions; 18, Pa Rayong; 22, Malaccensis nain; 26, Pisang Segun; 30, Cici Bresil; 29, Pisang Cici Alas ITC0415; 43, *Musa schizocarpa* ITC0599; 35, Banksii H09. The outline map was modified from http://www.histgeo.ac-aix-marseille.fr/ancien_site/carto. (b) Cumulative proportions of the genome covered by each ancestry for each wild accession. The color code is described in the above legend. Numbers at the bottom correspond to accessions listed in Table S1. (c) Results of *schizocarpa*/*banksii* breakpoint analysis for wild accessions. Identified breakpoints are shown on the left. Colored squares indicate presence of the corresponding breakpoint.

M. acuminata subspecies reflecting some divergence compared to *M. acuminata* ssp. *banksii* of New Guinea. The mosaics of *M. acuminata* ssp. *truncata* and *M. acuminata* ssp. *sumatrana* showed large contributions from *malaccensis* (45 and 46; Figure 1b). A few wild accessions had a more complex mosaic resembling cultivars or had mosaics suggesting that they could be hybrids between subspecies (e.g., 47–54; Figure 1b). Among them, wild accession 'EN13' from Indonesia was hybrid between *zebrina* and M_1 (54; Figure 1b).

Chromosome ancestry mosaics of diploid cultivars show shared ancestries and local diversification patterns

The chromosome ancestry mosaics of diploid cultivars involved three up to seven predicted ancestries (Figure 2a, b; Figure S3). A region of origin is known for most diploid cultivars with few exceptions (Table S1), which allows analyzing ancestry patterns in relation to geographical origin (Figure 2a,b). Accessions involving the lowest number of ancestries were six AA cultivars (AAcv) from PNG that showed a dominant *banksii* contribution with introgression from *zebrina* and *schizocarpa* (106–111; Figure 2a,b). A dominant *banksii* contribution always associated with *zebrina* and *schizocarpa* introgressions was predicted for most of the other analyzed AAcv from the New Guinea region (Figure 2b), together with *malaccensis* introgressions or combined with *malaccensis*, M_2, and sometimes M_1 introgressions. Two accessions from PNG showed one complete haplotype of *schizocarpa* origin in addition to one haplotype of *banksii* origin with *zebrina*, *malaccensis*, and *schizocarpa* introgressions (112 and 113; Figure 2a,b; Figure S3).

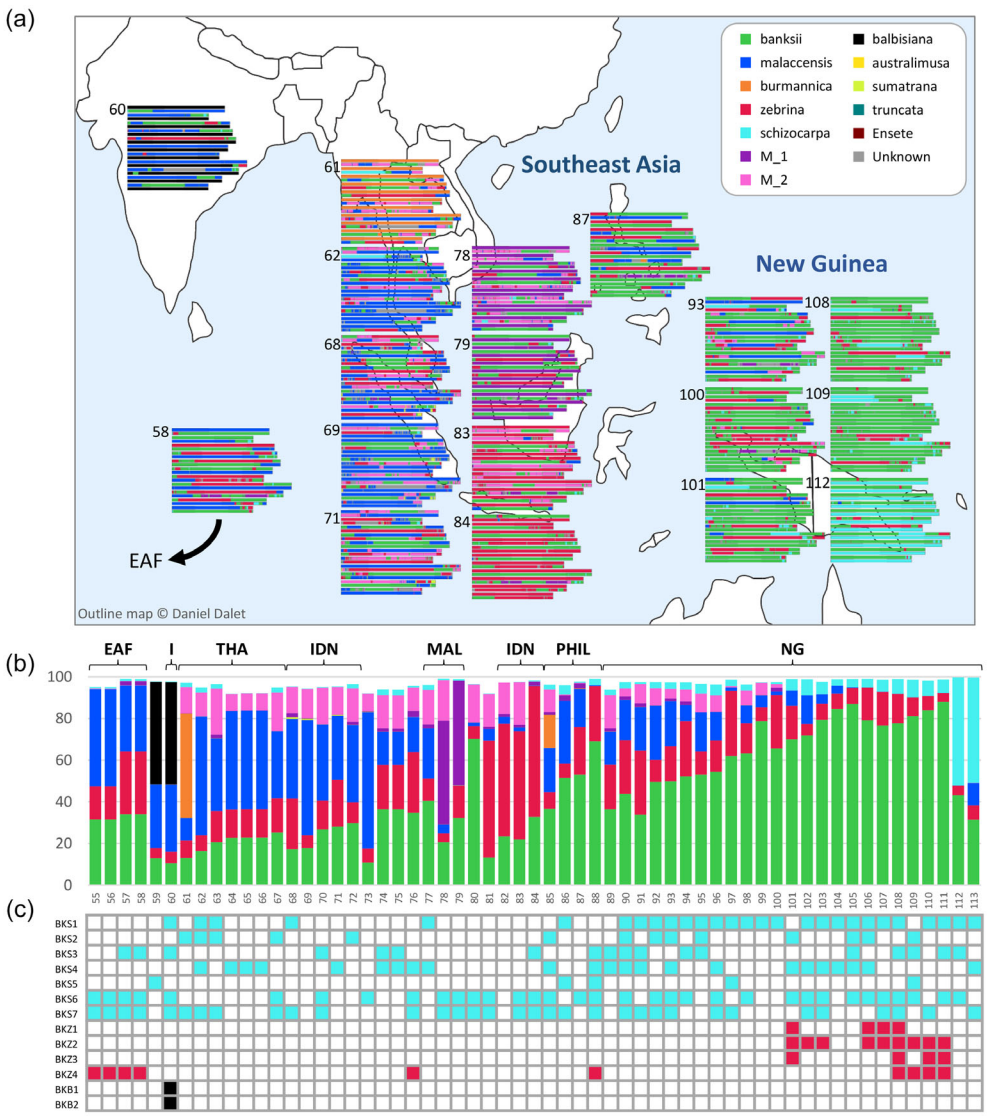

Figure 2. Genome ancestry of diploid banana cultivars from New Guinea (NG) and Southeast Asia (SEA).
(a) Genome mosaics of diploid cultivars showing contrasting mosaics in relation to geographical origin. Ancestral group contributions were represented along the 11 *Musa* reference chromosomes by different colors. The representation is based on non-phased genotyping data. Cultivars were categorized as SEA, NG, or East Africa (EAF) based on available prospection information or based on provenance (collections, Table S1). Numbers on the left refer to accessions listed in Table S1 and present in Figure 2b. The outline map was modified from http://www.histgeo.ac-aix-marseille.fr/ancien_site/carto. (b) Cumulative proportions of the genome covered by each ancestry for each diploid cultivar accession. NG, New Guinea (Papua New Guinea and Indonesian New Guinea); PHIL, Philippines; IDN, Indonesia (excluding Indonesian NG); MAL, Malaysia; THA, Thailand; I, India; EAF, East Africa. The color code is described in the above legend. Numbers at the bottom correspond to accessions listed in Table S1. (c) Results of *schizocarpa*/*banksii* (BKS1–BKS7), *zebrina*/*banksii* (BKZ1–BKZ4), and *balbisiana*/*banksii* (BKB1 and BKB2) breakpoint analysis for diploid cultivars. Identified breakpoints are shown on the left. Colored squares indicate presence of the corresponding breakpoint.

Diploid AAcv from SEA (Thailand, Malaysia, Philippines, and Indonesia [excluding the New Guinea part]) generally showed mosaics with four up to seven origins (Figure 2a,b). The three ancestries *banksii*, *zebrina*, and *schizocarpa* were present in all of them except for one accession, 'Pisang Jaran' from Indonesia (82; Figure 1b), which showed no *schizocarpa* ancestry. These three ancestries were often together with *malaccensis* and/or M_2, M_1, and in a few cases *burmannica* or *sumatrana*. The regional features illustrating contributions from local or geographically close subspecies were observed for SEA accessions, with a dominant *zebrina* contribution for a few accessions from Indonesia (e.g., 82–84; Figure 2a,b), a large *malaccensis* contribution in some accessions from Thailand or Indonesia (e.g., 62–72; Figure 2a,b), and a *burmannica* contribution in two accessions from Thailand (61; Figure 2a, b) and the Philippines (85; Figure 2b). Three diploid accessions from Indonesia (68, 69, and 71; Figure 2a,b) showed a

few regions assigned to *sumatrana*. The *truncata* contribution was limited to very small segments (Figure S3).

The 'M_1' unknown ancestral group was apparently present as a complete haplotype in one cultivar from continental Malaysia ('Pisang Jari Buaya', 79) and one from the island of Borneo ('Pisang Madu', 78; Figure 2a,b). A centromeric region on chromosome 9 was also assigned to M_1 in several cultivars from different areas (Figure S3).

The M_2 unknown ancestral group contributed large chromosomal regions in a majority of AAcv from SEA and in some accessions from New Guinea (Figure 2b). It is also present in the Sucrier dessert banana (74, 75, and 89; Figure 2b), the only globally distributed diploid subgroup (Stover & Simmonds, 1987). In contrast, its contribution was more limited or absent in East African accessions (Mchare subgroup and Paka, 55–58; Figure 2a,b) and a few AAcv from the Philippines, Malaysia, and Indonesia (Figure 2b).

Finally, AB interspecific cultivars (59 and 60; Figure 2a, b; Figure S3) thought to be of Indian origin displayed some A/B recombined chromosomes. Their A genome was predominantly of *malaccensis* origin, with large *banksii* and *zebrina* regions and a few s*chizocarpa* introgressions.

Common features and diversity of triploid cultivar mosaics

In contrast to most diploids, triploid cultivars have diffused in different regions of the world and their geographical origins are less precisely known. Figure 3a displays the mosaics of some popular cultivars worldwide or in tropical/subtropical regions (i.e., Cavendish, Plantains, Pome, East African Highland Banana, Ney Mannan, Pisang Awak), which illustrates the diversity of obtained ancestry profiles. The analysis of ancestral contributions (Figure 3b; Figure S3) showed that all triploid cultivars involving *M. acuminata* (AAA, AAB, ABB, and AAT) had contributions from *banksii*, *zebrina*, and *schizocarpa* and that some contributors were more dominant in some accessions (Figure 3b, Figure S3).

The majority of AAA cultivars showed mixed mosaic profiles with regions of *banksii*, *zebrina*, *schizocarpa*, *malaccensis*, M_2, and sometimes M_1 on chromosome 9 (Figure 3b; Figure S3). Commercial Cavendish dessert banana cultivars came under this category (Figure 3a), in addition to accessions representing the Ambon (117), Orotava (121), Gros Michel (124), and Red (125 and 126) subgroups (Figure 3b; Figure S3).

A dominant *zebrina* contribution was found for East African Highland Bananas, also with large *banksii* regions and *schizocarpa* introgressions (132–134; Figure 3a,b), while a dominant *malaccensis* contribution was observed for Ibota (115 and 116; Figure 3b).

The A/B interspecific triploid cultivars showed a global AAB or ABB structure with some recombined A/B chromosomes (Figure 3a; Figure S3). Profiles of mixed or different dominant ancestries for their A genomes were observed (Figure 3b). The A genomes of AAB cultivars representing the Mysore (137), Pome (141), Nadan (142), and Nendra Padathti (143) subgroups were mixed, with at least four to six origins (Figure 3a,b; Figure S3). A dominant *banksii* contribution with introgressions from *zebrina* and s*chizocarpa* was observed for AAB accessions representing Maia Maoli/Popoulu, which are mainly found in the Pacific Islands (151 and 152; Figure 3b), and for ABB accessions representing the subgroups Bluggoe, Monthan, Ney Mannan, Saba, and Pelipita (153–157; Figure 3a,b; Figure S3). A similar profile, but with additional *malaccensis* introgression, was obtained for AAB Plantains, which are major crops in West Africa (146–149; Figure 3), for Laknao (150) from the Philippines, and for Iholena (145). Finally, a dominant *malaccensis* origin with *banksii*, *zebrina*, and *schizocarpa* regions was observed for representatives of ABB Pisang Awak (158–160; Figure 3) and AAB Silk (138–140; Figure 3b).

No major *burmannica* contributions were predicted in our set of triploids, and *sumatrana* regions were found only in the AAA Indonesian accession 'Pisang Papan' (122; Figure S3).

We also showed that Fe'i cultivars were of homogenous *Australimusa* ancestry (Figure S3), and the mosaic of cultivar 'Karoina' (161; Figure 3b) revealed an AAT composition consistent with cytogenetic analysis findings of (D'Hont et al., 2000) and suggesting hybridization between an introgressed *banksii*-rich individual and an *Australimusa* species.

***Musa schizocarpa* introgressions are conserved among cultivars**

The observed cultivar genome mosaic patterns suggested that some breakpoints between segments of different origins may be shared between accessions. Notable patterns included several *schizocarpa*/*banksii* introgressions in many cultivars, two *banksii*/*balbisiana* recombination breakpoints on chromosome 9 in some A/B genome hybrids, and a few *banksii*/*zebrina* recombination breakpoints in East African AAcv Mchare and some AAcv from PNG. To validate these observations, the precise position of several of these recombination breakpoints was determined and scored.

Seven *schizocarpa*/*banksii* breakpoints on chromosomes 1, 2, 4, and 9 were selected (BKS1–BKS7; Table 1; Table S3). The breakpoints were found distributed in 21 to 69 accessions (Figures 1c, 2c, and 3c; Table 1; Table S3). Each breakpoint was present in at least one of the six PNG cultivars with the simplest mosaic (106–111; Figure 2; Figure S4). All cultivars of our sample (excluding Fe'i) had at least one of these seven breakpoints, with two exceptions: 'Pisang Jaran' (82; Figure 2; no *schizocarpa* introgression) and 'IDN110' (69; Figure 2). The 'IDN110' cultivar had only one *M. schizocarpa* introgression, present on chromosome 2 and visible in other cultivars from different origins but

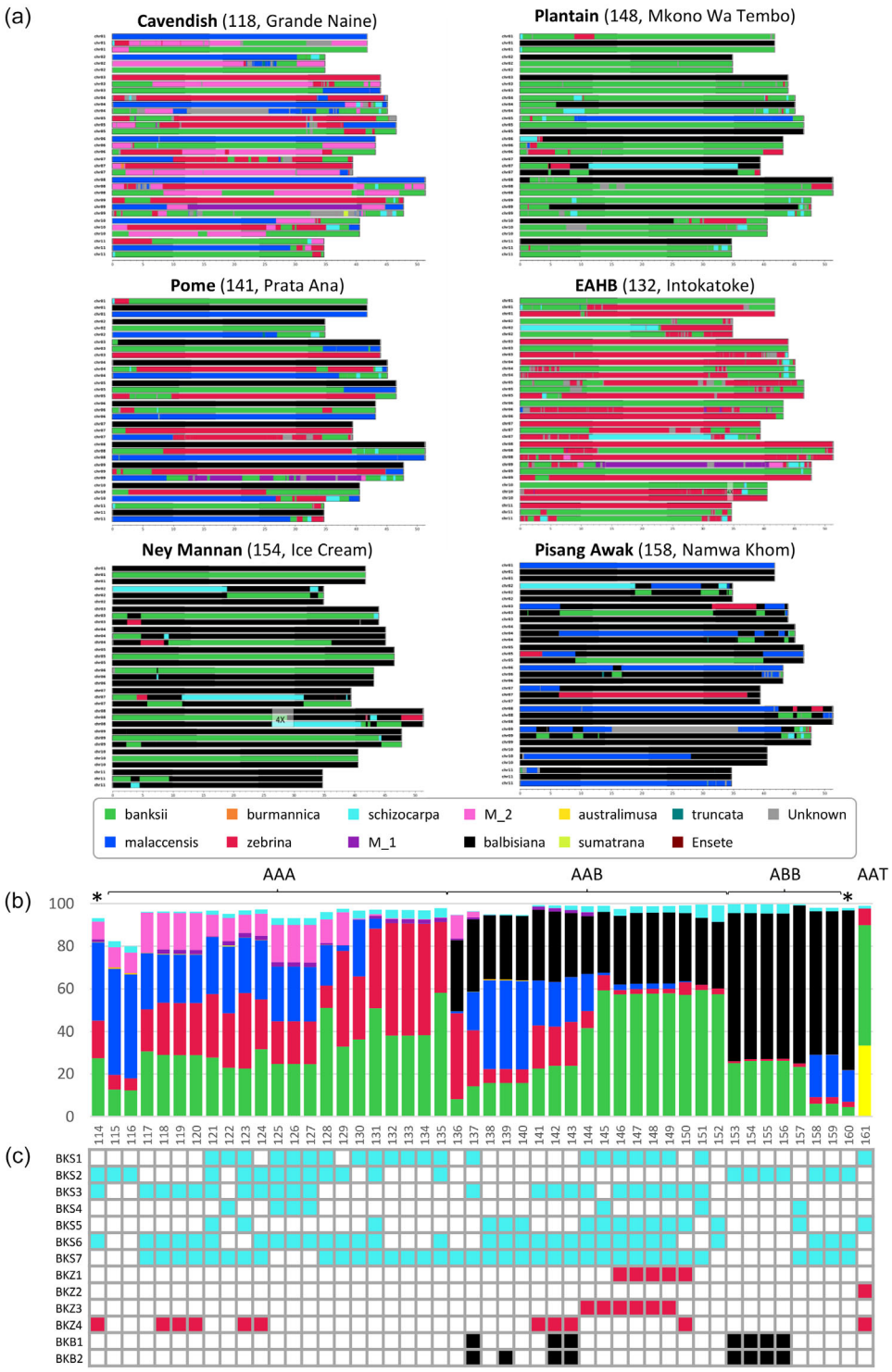

Figure 3. Genome ancestry of triploid banana cultivars.
(a) Contrasting genome ancestry mosaics of some popular triploid banana cultivars. Ancestral group contributions are presented along the 11 *Musa* reference chromosomes by different colors. The representation is based on non-phased genotyping data. Banana subgroups are indicated in bold, accession names are shown in between brackets. Segmental aneuploidy regions are indicated by whitish squares. (b) Cumulative proportions of the genome covered by each ancestry for each triploid cultivar are presented. Main banana genomic groups are indicated on top. A, *M. acuminata*; B, *M. balbisiana*; T, *Australimusa*. The color code is described in the above legend. Numbers at the bottom correspond to accessions listed in Table S1. Accessions are mainly triploids, and * indicates tetraploid accessions Calypso (114) and Pisang Awak (160). (c) Results of *schizocarpa*/*banksii* (BKS1–BKS7), *zebrina*/*banksii* (BKZ1–BKZ4), and *banksii*/*balbisiana* (BKB1 and BKB2) breakpoint analysis for polyploid cultivars. Identified breakpoints are shown on the left. Colored squares indicate presence of the corresponding breakpoint.

The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd.,
The Plant Journal, (2023), doi: 10.1111/tpj.16086

Table 1 Statistics on the presence of selected breakpoints of *M. schizocarpa* introgressions

	BKS1	BKS2	BKS3	BKS4	BKS5	BKS6	BKS7	No BK_S1_7
Chromosome	1	2	4	4	9	9	9	-
Nb. of PNG AAcv[a] (out of six)	5	2	3	1	1	5	3	0
Nb. of other cultivars (out of 105)	45	32	37	29	20	60	66	2
Nb. of wild accessions (out of 63)	0	1	1	0	0	0	0	61

[a]AA cultivars with the simplest mosaic content (Gulum, Manameg Red, Papat, Sena, Spiral, Tomolo); BKS, breakpoint of *schizocarpa* introgression into *banksii*; Nb., number.

absent from the six cultivars with the simplest mosaic (69; Figure S3). Two of these breakpoints were also present in two introgressed wild accessions (Figure 1c).

Four *zebrina*/*banksii* breakpoints present in PNG accessions with the simplest mosaics were also selected. They were shared between 9 to 21 cultivars (BKZ1–BKZ4; Figures 2c and 3c; Figure S5; Table S3). One of them, BKZ2, was shared between all six PNG accessions and four other PNG accessions, two, BKZ1 and BKZ3, were shared between some of the six PNG accessions and mostly Plantains (146–149), and the fourth one, BKZ4, was shared between four of the six PNG accessions and 17 other diploid and triploid cultivars, including AA Mchare (57–58) and AAA Cavendish (118–120).

Using the same approach, we also validated that the two *banksii*/*balbisiana* breakpoints on chromosome 9 were shared between eight and nine accessions, respectively, corresponding to the AB diploid cultivar 'Kunnan' (60; Figure 2c) and accessions representing four AAB subgroups (Mysore [137], Silk [139], Nadan [142], and Nendra Padathti [143]; Figure 3c) and four ABB subgroups (Saba [153], Ney Mannan [154], Monthan [155], and Bluggoe [156]; Figure 3c; Figure S5; Table S3).

Phylogenetic positioning and search for M_1 and M_2

To investigate relationships between the two unknown contributors M_1 and M_2 and the other ancestral groups that contributed to cultivars, we performed maximum likelihood phylogenetic analysis on one region for each of the 11 reference chromosomes. These regions were those where M_1 and M_2 haplotypes could be defined and where either haplotypes or consensuses from wild accessions could be derived from concatenated SNPs.

Four monophyletic groups were identified on all 11 phylogenies (Figure 4; Figure S6). Group I was comprised of *M. coccinea*, *Musa textilis*, *Musa lolodensis*, *Musa maclayi*, and *Musa peekelii*. Group II corresponded to *M. balbisiana* accessions. Group III was comprised of *M. ornata*, *M. velutina*, and *M. sanguinea*. Group IV was comprised of all accessions from *M. acuminata* subspecies, as well as *M. schizocarpa*, *M. rosea*, *M. laterita*, and haplotypes from unknown contributors M_1 and M_2. The *M. itinerans* accession generally branched at the basis of the monophyletic group formed by groups III and IV (10/11 phylogenies).

Within group IV, accessions from the same (sub)species consistently formed monophyletic groups according to their classification, with the exception of the *M. acuminata* ssp. *malaccensis* accession 'DB_malaccensis_2012_1154' (Figure S6e,f,j). The *M. laterita* and *M. rosea* accessions formed a monophyletic group that always clustered with the group formed by *M. acuminata* ssp. *burmannica* and *M. acuminata* ssp. *siamea* accessions. *Musa schizocarpa* generally grouped with the remaining *M. acuminata* subspecies (10/11 phylogenies). Accessions/haplotypes of *M. acuminata* ssp. *errans* and *M. acuminata* ssp. *microcarpa* 'Borneo' were always grouped with *M. acuminata* ssp. *banksii*.

In 10 phylogenies, the haplotype representative of the M_1 unknown origin (purple in Figure 4) was placed at the base of a monophyletic group formed by itself and *M. schizocarpa* accessions. *Musa acuminata* ssp. *sumatrana* usually clustered at the base of this group (nine phylogenies; Figure 4; Figure S6). In 10 phylogenies, the haplotype representative of the M_2 unknown origin was placed at the base of a monophyletic group formed by itself and *M. acuminata* ssp. *zebrina* accessions. In one phylogeny (chromosome 5; Figure S6d), both M_1 and M_2 haplotypes formed a monophyletic group with *M. schizocarpa* accessions as well as *M. acuminata* ssp. *zebrina* and *M. acuminata* ssp. *sumatrana*.

The *M. acuminata* ssp. *truncata* accession usually clustered at the base of the group comprised of *M. acuminata* ssp. *banksii*, *M. acuminata* ssp. *microcarpa* ('Borneo'), and *M. acuminata* ssp. *errans* (Figure 4; Figure S6). For one phylogeny on chromosome 10, it clustered at the base of the group containing *M. schizocarpa* accessions and the M_1 haplotype.

We extended our search for M_1 and M_2 wild representatives by using low-coverage genome skimming sequencing data for 39 *Musa* species or subspecies available in the NCBI database (Table S1). The analysis of profiles of normalized ancestry informative allele ratios obtained for the 39 accessions suggested that M_1 corresponded to *M. acuminata* ssp. *halabanensis* (Figure S7). No wild representative for M_2 was identified.

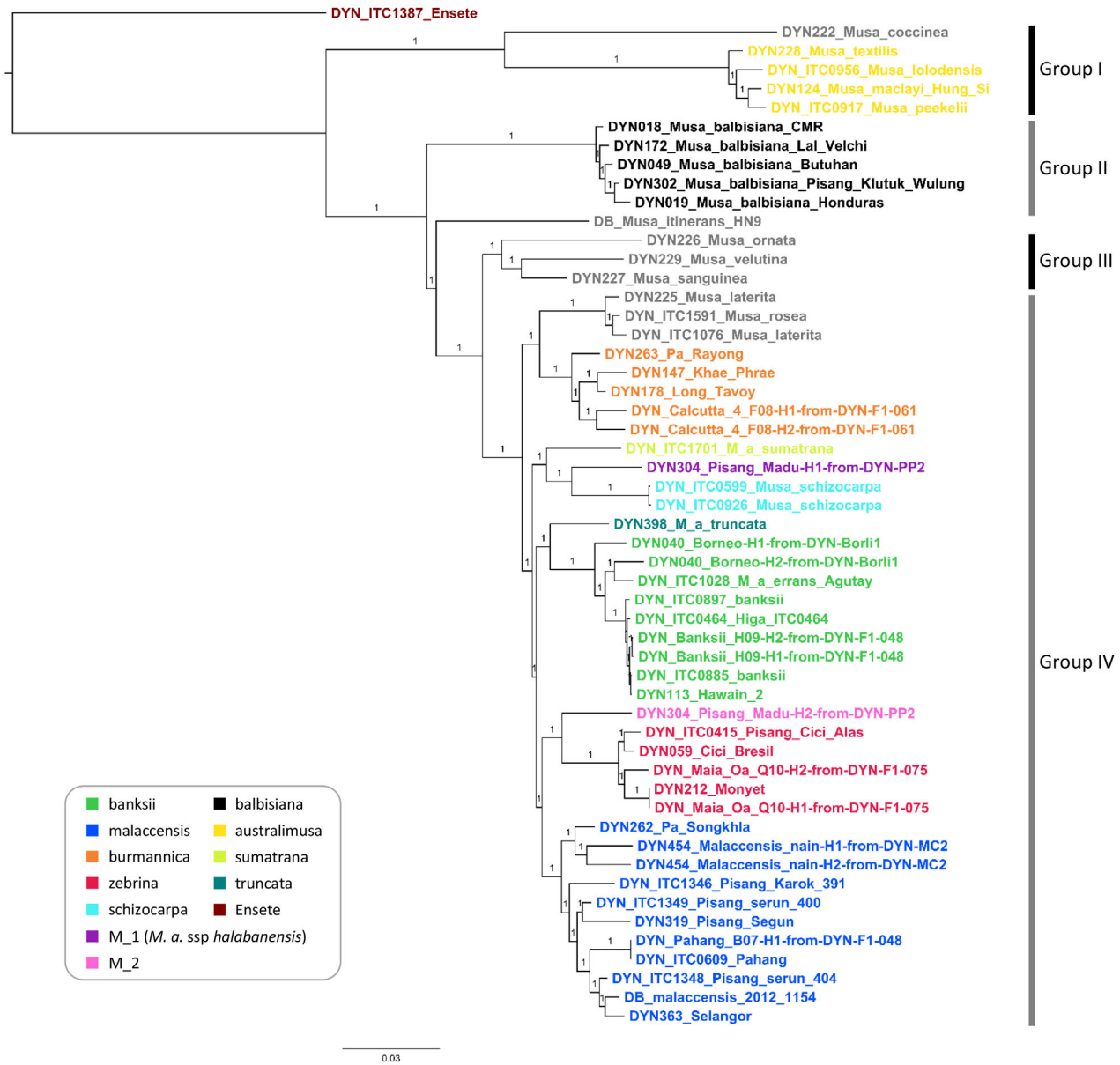

Figure 4. Maximum likelihood phylogenetic tree of *Musa* haplotypes corresponding to reference chromosome 2.
Phylogenetic analysis was performed using PHYML v3.1 on consensus or haplotypes of different accessions, reconstructed from concatenated polymorphous sites and involving a region of chromosome 2 of approximately 4.48 Mb, with a total of 123 896 polymorphic sites. Branch support was estimated with the 'approximate Bayes branch supports' algorithm of PHYML v3.1 and rounded to three decimals. Colors correspond to ancestral origins contributing to cultivated bananas as determined via chromosome painting. Main phylogenetic clusters are indicated on the right. Some haplotypes were derived using parent–child trios with the F1 name indicated in the labels. The tree was rooted on *Ensete*.

DISCUSSION

Multiple and unknown wild ancestral contributors to cultivated bananas

We analyzed a sample of 175 *Musaceae* accessions in which we identified 11 genetic groups that contributed to banana cultivars. Previous studies had already identified *M. acuminata* ssp. *banksii*, *M. acuminata* ssp. *zebrina*, *M. acuminata* ssp. *malaccensis*, *M. acuminata* ssp. *burmannica*, *M. balbisiana*, and, in a few cases, *Australimusa* and *M. schizocarpa* as contributors to cultivar genomes (e.g., Boonruangrod et al., 2008, Boonruangrod et al., 2009, Carreel et al., 2002, Perrier et al., 2009, Simmonds & Shepherd, 1955). The contribution of the species *M. schizocarpa* was until now thought to be restricted to a few AS cultivars (Carreel et al., 1994) and suspected to be present in East

African highland bananas (Nemeckova et al., 2018). We showed here that *M. schizocarpa* was a contributor to all cultivars. The only exception, 'Pisang Jaran', is quite seedy (https://www.crop-diversity.org/mgis) and thus may be considered as a non-typical cultivar. To these groups, we also added the contributions of two unknown ancestries, M_1 and M_2, one of which we identified as *M. acuminata* ssp. *halabanensis*, and identified local *sumatrana* and possibly *truncata* contributions.

The informative alleles of these 11 genetic groups allowed precise evaluation of their contributions to cultivated banana genomes. They also revealed the presence of introgressions in wild (seedy) individuals, particularly within *M. acuminata*, which may be remnants of hybridizations between geographically close or introduced subspecies (e.g., Rouard et al., 2018) or may have resulted from gene flow from partially fertile cultivars.

For two ancestral groups, i.e., *banksii* and *burmannica*, the representative alleles were identified from accessions morphologically described as belonging to several *M. acuminata* subspecies. The close relationship between *M. acuminata* ssp. *burmannica*, *M. acuminata* ssp. *burmannicoides*, and *M. acuminata* ssp. *siamea* forming the *burmannica* ancestral group was supported by our phylogeny results and previous research (Boonruangrod et al., 2009; Christelová et al., 2017; Dupouy et al., 2019; Martin, Cardi, et al., 2020; Perrier et al., 2009). This genetic group was also found to be phylogenetically related to *M. laterita* and *M. rosea*, which was in line with the geographical distribution of these species (Janssens et al., 2016). The phylogenetic analysis also supported the close relationships between *M. acuminata* ssp. *banksii*, *M. acuminata* ssp. *microcarpa* accession 'Borneo', and *M. acuminata* ssp. *errans* accession 'Agutay' forming our *banksii* ancestral group. Recently, a common genetic background was suggested between these three subspecies with, however, some variability specific to *M. acuminata* ssp. *errans*/*M. acuminata* ssp. *microcarpa* that may have contributed to cultivars (Sardos et al., 2022).

The subspecies *M. acuminata* ssp. *sumatrana* and *M. acuminata* ssp. *truncata* were, each, only represented by one individual from Northwestern Sumatra and the Malay Peninsula, respectively. These areas are close to that of *M. acuminata* ssp. *malaccensis*. Their mosaics suggested that they might have been introgressed by *M. acuminata* ssp. *malaccensis* or are more related to this subspecies. The phylogenetic analysis did not help clarify their relationships and will require a larger sample.

We previously proposed that two unknown ancestral groups contributed to banana cultivars (Martin, Cardi, et al., 2020). Further studies based on ribosomal DNA polymorphism (Jeensae et al., 2021) or global population structure approaches (Sardos et al., 2022) supported this assumption. We postulated two ancestors of unknown origin, i.e., M_1 and M_2, based on the most parsimonious interpretation of the observed mosaic profiles in our iterative approach to identify informative alleles for unknown ancestors. The M_1 and particularly M_2 ancestral groups explained the origin of a majority of chromosomal segments not attributed to the nine groups of known origin in cultivars. However, the origins of a few chromosomal segments could not be determined, particularly in accessions of a dominant *malaccensis*, *zebrina*, or M_2 origin. Some might originate from more divergent forms of these contributors, while others could correspond to M_2 regions not represented in the sample used to determine M_2 alleles.

Phylogenetic analysis suggested a close relationship between M_1 and two geographically distant groups *M. schizocarpa* (PNG) and *M. acuminata* ssp. *sumatrana* (Sumatra, Indonesia). The analysis of low-coverage sequencing samples from public databases suggested that M_1 corresponded to *M. acuminata* ssp. *halabanensis*. This wild banana was described in Sumatra (Hotta, 1989; Meijer, 1961; Nasution, 1989) and was accepted as an *M. acuminata* subspecies (Häkkinen & Väre, 2008). It was recently proposed to be present in cultivar 'Pisang Jari Buaya' (Ahmad, 2021). The identification of *M. acuminata* ssp. *halabanensis* as representative for M_1 validates the hypothesis that M_1 represents a single genetic pool and demonstrates the contribution of additional wild *Musa* to the formation of cultivars.

The M_2 contributor was present in many cultivars, particularly accessions from SEA. Alleles representing M_2 were defined with different accessions based on chromosomal regions that were often, but not always, overlapping. The consistency of 10 out of 11 phylogenies that grouped M_2 haplotypes close to *M. acuminata* ssp. *zebrina* (from Java, Indonesia) supports the hypothesis that M_2 is a single missing genetic group. A large reciprocal translocation involving chromosomes 1 and 7 was previously identified in cultivar 'Pisang Madu' (Martin, Cardi, et al., 2020), which bears both M_1 and M_2 contributions for these chromosomes. Comparing the distribution of this 1/7 translocation in cultivars (Martin, Baurens, et al., 2020) in relation to the observed genome mosaics supports an M_2 origin for this translocation.

In Indonesia, 15 morphological types were described as botanical varieties of *M. acuminata* (Nasution, 1989; Nasution 1991; Pollefeys et al., 2004), some of which are generally considered to be subspecies (*malaccensis*, *zebrina*, *microcarpa*, *halabanensis*). The substantial genetic diversity of *M. acuminata* in this area has been confirmed by molecular marker analysis (Poerba et al., 2019). The central zone of *Musa* diversity is thus a good starting point to look for the M_2 ancestor.

A complex inter(sub)specific hybrid origin of banana cultivars involving core and local contributors

The genome mosaics results demonstrated the complex inter(sub)specific hybrid origins of all sampled cultivars involving at least three to seven ancestries. The only exception concerned Fe'i cultivars, which were found to be solely derived from *Australimusa* ancestry. Their possible hybrid status within this genetic group has yet to be investigated.

We found that the three ancestries *banksii*, *schizocarpa*, and *zebrina* were present in all of the other sampled cultivars (except 'Pisang Jaran') and thus can be considered as a 'core' contribution to cultivated bananas. The *banksii*, *zebrina*, *schizocarpa*, and *malaccensis* genetic groups were also found to be dominant ancestral contributions in cultivars from regions where these wild species or subspecies are present. For example, the increased *malaccensis* contributions in some cultivars from SEA were consistent with the *M. acuminata* ssp. *malaccensis* distribution from the Malay Peninsula to Sumatra. This consistency was also observed for major contributions of *banksii* and *schizocarpa* in New Guinea or *zebrina* in Indonesia. Contributions of *burmannica* were also in line with the cultivar origins, but were limited in number, as previously noted with samples of similar diversity (Perrier et al., 2009). *Musa acuminata* ssp. *sumatrana* and *M. acuminata* ssp. *truncata* contributions may have occurred locally, in a limited fashion.

The *M. acuminata* ssp. *halabanensis* (M_1 ancestry) contribution did not seem widespread in sampled cultivars and may have only occurred in a few Indonesian or Malaysian cultivars. The *halabanensis* origin assigned to the (peri)centromeric region of chromosome 9 in several cultivars will need confirmation due to the more limited number of markers and/or possible *halabanensis* overprediction in this region. In contrast, M_2 was a major contributor to different banana cultivars, including those producing dessert bananas such as Cavendish, which accounts for around half of the current world banana production, as well as its predecessor Gros Michel, which was almost completely wiped out by *Fusarium* wilt. Many cultivars used as dessert bananas (e.g., Sucrier, Cavendish, Gros Michel, Ambon, Mysore, and Pome) had mixed mosaics with large *malaccensis* and often M_2 regions, in addition to the 'core' contributors. Recently, a quantitative trait locus contributing to banana pulp acidity was located on one haplotype of M_2 origin in 'Pisang Madu' (Biabiany et al., 2022). This suggests that M_2 could be an important contributor to fruit quality traits in dessert cultivars such as Cavendish.

Different types of A/B interspecific triploid cultivars, mainly belonging to the starchy banana subgroups Plantain, Iholena, Maia Maoli, Popoulou, Laknau, Bluggoe, Ney Mannan, and Monthan, were previously predicted to be derived from *M. acuminata* ssp. *banksii* for their A genome (Hippolyte et al., 2012; Lebot et al., 1993; Perrier et al., 2009). We found that their A genomes were more complex than expected and quite similar to those of diploid *banksii*-rich New Guinea hybrid cultivars. This supports the use of such diploids in Plantain and starchy banana breeding programs.

Finally, recombination between A and B genomes was visible in almost all of the analyzed A/B interspecific hybrids, thus confirming that there were several interspecific hybridization steps at their origin, as previously suggested (Baurens et al., 2019; Cenci et al., 2021; De Langhe et al., 2010). The conserved A/B recombination breakpoints observed for some accessions representing different subgroups (Kunnan, Nadan, Nendra Padaththi, Silk, Mysore, Bluggoe, Monthan, Ney Mannan, and Saba) indicated a shared origin of these accessions. This is in line with the proposed common region of origin of these subgroups in India (De Langhe et al., 2010; Stover & Simmonds, 1987) and illustrates shared pedigree relationships between some cultivars.

In general, the observed mosaics together with available geographical information suggest that banana cultivars bear core ancestral contributors but also went through a diversification process in different regions, by hybridizations/introgressions with local wild genetic pools.

Widespread and shared *M. schizocarpa* introgressions in cultivated banana suggest a common origin in New Guinea

Previous molecular marker studies have identified a pool of AA cultivars from PNG that were closely related to *M. acuminata* ssp. *banksii* (Carreel et al., 2002; Perrier et al., 2009; Sardos, Perrier, et al., 2016), supporting the idea that this subspecies had a pivotal role in banana domestication, while suggesting that domestication in New Guinea might not have been based on inter(sub) specific hybridization (Carreel et al., 2002; Sardos, Perrier, et al., 2016). Our ancestry analysis at the chromosome level brought more precise information on AAcv from PNG. We found that the cultivars with the simplest mosaics were indeed from PNG and displayed a major *banksii* contribution. However, this *banksii* contribution was always accompanied by introgressed regions from *schizocarpa*. *Musa acuminata* ssp. *banksii* and *M. schizocarpa* are sympatric in PNG (Argent, 1976; Arnaud & Horry, 1997). Both were described as preferentially autogamous, but natural hybrids between them were observed (Argent, 1976; Eyland et al., 2020) and confirmed in molecular studies (Carreel et al., 2002). We showed that several *schizocarpa* introgression breakpoints into *banksii* were shared between New Guinea accessions and all of the diploid and polyploid cultivars of our sample, but with one exception. This indicates that individuals bearing these

schizocarpa/banksii breakpoints are common ancestors to current cultivars. In addition, since *M. schizocarpa* has not been reported outside of New Guinea, this suggests that *schizocarpa* introgressions into *banksii* occurred on the island of New Guinea and that the resulting introgressed individuals were selected by humans in the New Guinea region, were transported across SEA, and gave rise to current cultivars.

In addition to *schizocarpa*, the New Guinea accessions we analyzed also always had introgressions from *zebrina*. It is unclear whether introgressions of *M. acuminata* ssp. *zebrina*, native to the island of Java, directly occurred in New Guinea. Movements of populations in the Wallacea and New Guinea area during the Holocene (Donohue & Denham, 2009; Pedro et al., 2020; Soares et al., 2011) favored plant exchanges in the area, and these may have included *M. acuminata* ssp. *zebrina*, which has a high ornamental value. Yet cultivars resulting from hybridizations between *banksii*/*schizocarpa* early cultivars and *zebrina* in Indonesia could have been brought back to New Guinea.

Archeological investigations at the Kuk Swamp site in PNG suggested that bananas were intensively cultivated there at least 6950 to 6440 years ago (Denham et al., 2003). In addition, Mchare and Plantain cultivars are believed to have been transported from the SEA–New Guinea region to Africa in a proposed time frame of 1000 to 3000 years ago, depending on evidence and sources (Grimaldi et al., 2022; Perrier et al., 2019). The conserved *schizocarpa*/*banksii* and *zebrina*/*banksii* breakpoints between some PNG diploid cultivars and Mchare or Plantains cultivars suggest that introgression events associated with these breakpoints took place at least 1000 years ago.

The fact that the simplest mosaics were found in diploid cultivars from New Guinea, that they involved two contributors originating from New Guinea (*banksii* and *schizocarpa*) that were core contributors to banana cultivars, and that *banksii*/*schizocarpa* introgression breakpoints were shared between these simplest diploid cultivars and with the other cultivars leads us to propose that banana domestication was initiated in the New Guinea region.

This domestication process was based on interspecific hybridizations involving *banksii*, *schizocarpa*, and potentially *zebrina* ancestral groups. The fertility of hybrids between *banksii* and *schizocarpa* (and *zebrina*) must have been sufficient to allow backcrosses with *M. acuminata* ssp. *banksii*, resulting in the observed introgression patterns in the PNG accessions. Such introgression events may have been rare and/or particular hybrids may have been selected by humans for their traits of interest, thereby explaining the shared introgression patterns. Inter(sub) specific hybridizations may have disrupted finely regulated fertility and fruit development mechanisms, resulting in parthenocarpic banana fruit, as proposed for natural sources of parthenocarpy in tomato (*Solanum lycopersicum*) (Picarella & Mazzucato, 2018). Some resulting individuals bearing more pulp-rich edible fruits were selected, disseminated, and transported westwards. Further diversification occurred thanks to residual fertility, along diffusion paths in SEA, through hybridizations involving local *Musa* individuals, i.e., wild or selected for various uses. These hybridizations, combining multiple ancestries and accompanied by polyploidization, resulted in diverse diploid and triploid cultivars. Further movements of wild *Musa* or of partially fertile hybrids from SEA regions towards New Guinea might explain the introgressions of Western origin in *banksii*-rich New Guinea cultivars.

Our approach revealed more complex banana cultivar origins than previously thought with at least 11 genetic groups involved and several crossing steps. Yet shared recombination breakpoints between cultivars of diverse origins suggest that initially a rather limited number of successful combinations from New Guinea was disseminated, thereby participating in the formation of the current pool of cultivars.

These findings and resulting hypotheses are important to help broaden or more precisely target useful breeding resources. Identifying the M_2 unknown ancestor will be essential in that respect.

EXPERIMENTAL PROCEDURES

Plant material and sequencing

A set of 226 Musaceae accessions was used for this analysis (Table S1). It included 111 accessions from the Guadeloupe CIRAD field collection hosted by CRB Plantes Tropicales Antilles CIRAD-INRAe (http://crbtropicaux.com/Portail) and 37 accessions from the *in vitro* Bioversity International Transit Center in Leuven (Belgium) (https://www.bioversityinternational.org/banana-genebank). DNA from two accessions was provided by X. Perrier (CIRAD) from material described in (Perrier et al., 2019). The 12 F1 individuals were provided by the CIRAD banana breeding platform in Guadeloupe. For these 162 individuals, total DNA was sequenced using an Illumina HiSeq 4000 platform at Genoscope (https://jacob.cea.fr/drf/ifrancoisjacob/english/Pages/Departments/Genoscope.aspx), except for one accession for which DNA was processed with the Dovetail™ Hi-C protocol and sequenced using HiSeqX at Dovetail Genomics (https://dovetailgenomics.com/). For 64 additional accessions, whole genome sequencing data were collected from the NCBI public database (Table S1), including 39 accessions with low sequencing coverage that we only used to search for wild representatives of ancestries of unknown origins (Table S1).

Reads were quality-filtered and adapters were removed using Cutadapt (Martin, 2011). Quality filtering and adapter removal were already performed for reads provided by Genoscope.

Variant calling and vcf filtration

Variant calling was performed using *M. acuminata* reference sequence V4 (Belser et al., 2021) using the vcfhunter toolbox (https://github.com/SouthGreenPlatform/VcfHunter) as described

in (Baurens et al., 2019) for whole genome sequencing reads. The vcf file was filtered to remove variant sites found on annotated transposable elements and also on tandem repeats identified by Mreps (Kolpakov et al., 2003). Biallelic sites with no insertions/deletions (Indels) were selected from the resulting vcf file. In addition, for each accession, sites supported by less than 10 reads or more than 1000 reads and heterozygous sites with alleles supported by less than three reads or less than 10% reads were converted to missing data.

Detection of aneuploidy

To detect large Indels or potential aneuploidy, the SNP coverage was plotted for each position in the vcf file along chromosomes of *M. acuminata* reference sequence V4 (Belser et al., 2021) using vcf2allPropAndCov.py and vcf2allPropAndCovByChr.py from the vcf toolbox (https://github.com/SouthGreenPlatform/VcfHunter) (Baurens et al., 2019). The size of large regions showing dosage change was estimated visually based on the graphical representation. The most common event was a 4-Mb duplication in the pericentromeric region of chromosome 8 that was shared by 23 accessions and considered as a single event (Table S2). Thus, a total of 21 deletions (estimated size range 0.1–25 Mb, average 8.3 Mb) and 10 duplications (estimated size range 0.1–4 Mb, average 3.8 Mb) were identified. One accession (Lady Finger) showed one entire supernumerary chromosome 8. Of the aneuploid segments, 14 were at least partly located in the pericentromeric regions of chromosomes (defined as a 10-Mb region centered on the average location of the centromeric repeat Nanica). Only the aneuploidy of the complete chromosome 8 in Lady Finger was considered for final chromosome painting. Large aneuploid segments are indicated with shaded boxes on the final painting representations (Figure S3) and the genome mosaic interpretation in these regions should be based on allele ratio profiles.

Identification of ancestry informative alleles

The process of identification of ancestry informative alleles is summarized in Figures S1 and S2 and detailed in Supplementary Information files (Methods S1–S4; Tables SM1–SM13; Figures SM1–SM11). The first step (Figure S1; Method S1) was to identify specific alleles from predicted *Musa* contributors: *M. balbisiana*, *M. schizocarpa*, and *M. acuminata* subspecies (i.e., ssp. *banksii*, ssp. *errans*, ssp. *microcarpa* 'Borneo', ssp. *burmannicoides*, ssp. *burmannica*, ssp. *siamea*, ssp. *malaccensis*, and ssp. *zebrina*). As wild accessions can be introgressed (Martin, Cardi, et al., 2020), it was necessary to select the most homogenous representatives. For this, private alleles of the most homozygous and the second most homozygous accession of each genetic group were identified using IdentPrivateAllele.py (i.e., alleles present in the most homozygous accession and not in members of other groups). They were independently used to generate local ancestry plots, representing along chromosomes of each accession normalized values of the proportion of reads supporting private alleles of each genetic group. The process was performed using allele_ratio_group.py, allele_ratio_per_acc.py, PaintArp.py, and plot_allele_normalized_mean_ratio_per_acc.py. Selected homogenous accessions and chromosomes were used to identify informative alleles using correspondence analysis (CoA) for each of the 11 banana chromosomes and subsequent meanshift allele clustering (vcf2struct.py; Martin, Cardi, et al., 2020). Alleles representing six genetic groups were obtained (*banksii*, *zebrina*, *malaccensis*, *burmannica*, *M. schizocarpa*, and *M. balbisiana*). The second step was to search for additional contributors in a larger sample of wild *Musa* species (Figure S1; Method S2). Using IdentOtherAncestry.py, for each banana accession, two values were calculated: (i) the proportion of alleles that were not found in any of the accessions used as ancestors in the multivariate approach and (ii) the proportion of these alleles present in other *Musa* species and *M. acuminata* subspecies. The third step was to identify private alleles of identified additional contributors as well as improving the set of private alleles from *M. balbisiana* (Figure S1; Method S3). The process was performed for all genetic groups using IdentPrivateAllele.py. The level of fixation of private alleles was estimated as the average of read proportion values supporting each private allele, calculated in accessions of the corresponding genetic group, using allele_ratio_group.py. This added a set of private alleles for *M. acuminata* ssp. *sumatrana*, *M. acuminata* ssp. *truncata*, *Australimusa*, and *Ensete* as outgroup. For the *M. acuminata* subspecies of step one and *M. schizocarpa*, the alleles attributed using the CoA approach were kept because they introgress each other and thus a private allele approach is not appropriate. The *M. balbisiana* alleles that were kept, were those of the private allele strategy that allowed eliminating basal alleles shared with *Australimusa* and *Ensete*.

Normalized ancestral allele ratios were calculated for each private allele of the 10 resulting genetic groups in the dataset of 187 individuals and a mean value was calculated on sliding windows of 5601 SNPs (allele_ratio_per_acc.py and PaintArp.py). The mean normalized ratios along chromosomes were drawn to visualize local ancestry profiles (plot_allele_normalized_mean_ratio_per_acc.py).

In the fourth step, iterative approaches were used to select private alleles of the M_1 and then the M_2 unknown contributors (Figure S2; Method S4). Phased data (haplotypes) used in this process were obtained from parent–child trios with PhaseInVcf.py. Globally, iterations were a three-stage process (Figure S2): (i) identification and selection of accessions/haplotypes with large regions of unknown ancestry, (ii) determination of private alleles for M_1 or M_2 from selected accessions/haplotypes using all defined ancestral genetic groups, (iii) the arbitrary attribution or calculation of a level of allele fixation for the M_1 or M_2 set of alleles, (iv) the calculation of normalized allele ratio profiles on the complete dataset with all defined ancestral groups, and (v) drawing of ancestry plots to identify other accessions or parent–child trios that would allow improvement of the M_1 or M_2 private allele set. The location of regions with two haplotypes of known origin that could not be used to identify M_1- or M_2-specific alleles was identified using DrawRatioDetailInterractive.py. For M_1, private alleles were first identified in iteration 1 using 'Pisang Jari Buaya' that had a complete haplotype of unknown origin, then in iteration 2 using accession 'EN13', and in iteration 3 using two parent–child trios involving 'Pisang Madu'. The final set of M_1 private alleles was based on all accessions of the three iterations. For M_2, private alleles were identified in iteration 1 using 'Pisang Madu' haplotypes from the M_1 iteration 3 step, in iteration 2 they were identified using five other diploid cultivars with M_2 regions, and in iteration three they were identified using eight different parent–child trios that allowed access to M_2 haplotypes. The final set of M_2 private alleles was based on alleles derived in iteration 3. New scripts developed to perform this analysis were added to the vcfhunter toolbox (https://github.com/SouthGreenPlatform/VcfHunter).

Chromosome painting

Identified ancestry informative alleles were used to analyze ancestral contributions along chromosomes of the 226 accessions. For this, representative accessions of ancestral groups have been used to infer, for each position *j* corresponding to an informative allele, the observed proportion A_{ijk} of reads supporting an allele

from an accession *i* from a given ancestral group *k*. These values were used to calculate P_{jk} as the mean proportion of reads supporting the given allele observed in the *n* accessions representative of the group *k* that were not introgressed from another group in this region ($P_{jk} = \frac{1}{n}\Sigma A_{ijk}$). This was performed using the allele_ratio_group.py script added to vcfhunter toolbox. The P_{jk} value was arbitrarily set to 0.47 or 0.66 for the M_1 origin in regions in which three or two accessions were used to identify specific alleles of this origin (Method S4 for detailed calculation). The P_{jk} value was arbitrarily set to 0.8 for M_2 origins.

The observed proportion of reads (O_{ijk}, similar to A_{ijk} but for accessions to be characterized) was then calculated using the script allele_ratio_per_acc.py. For 187 individuals that did not show many missing data, data points that showed SNP coverage equal to or greater than 10 were used. For 39 individuals that were less covered, data points that showed SNP coverage equal to or greater than 3 were used.

These proportions were then normalized (N_{ijk} value) for each *j* SNP position on sliding windows of 5801 SNPs by dividing, for each ancestral origin, the sum of proportions of observed ratios by the sum of the expected ratios:

$$N_{ijk} = \frac{\sum_{l=j-2901}^{j+2900} O_{ilk}}{\sum_{l=j-2901}^{j+2900} P_{lk}}.$$

The value obtained was generally between 0 and 1 and was considered as 1 in the rare cases where it exceeded 1. These values were used to calculate the ancestral contributions to a given position in terms of ancestral haplotype composition by calculating the haplotype composition probabilities based on binomial distributions. Then, as the SNPs were not phased, a minimal number of recombinations was assumed for each accession and therefore contiguous regions of the same ancestry were arbitrarily attributed to one of the homologous chromosomes, resulting in pseudo-haplotype representations. These steps were performed using the PaintArp.py script.

Normalized values (N_{ijk}) were visualized along chromosomes using the plot_allele_normalized_mean_ratio_per_acc.py script. Deduced pseudo-haplotypes were drawn using the haplo2kar.1.0.py script. Due to low-coverage sequencing, only the normalized mean ratio plots were analyzed for the 39 least-covered individuals. To summarize mosaic information, cumulative proportions of the genome covered by each ancestry were calculated for each accession based on reference genome size and global ploidy.

Breakpoint determination

Recombination/introgression breakpoints, corresponding to regions along chromosomes where there is a switch between two distinct ancestral origins, were determined by visual inspection of graphic representations depicting, for each position with an assigned ancestral origin, the proportion of reads supporting this origin in the studied accession. This was done using the DrawRatioDetailInterractive.py script with data generated during the chromosome painting step. A graph was generated for each origin along each chromosome of the accession. These graphs showed a succession of clusters of colored dots corresponding to each origin, the ends of cluster blocks thus locating the position of recombination regions (Figure S4). By clicking on the first or the last point of the introgressed block, we obtained breakpoint coordinates which were then compared between accessions. The *schizocarpa*, *zebrina*, and *banksii* allele coordinates were noted, respectively, for *schizocarpa*/*banksii*, *zebrina*/*banksii*, and *banksii*/*balbisiana* breakpoints. Described breakpoints were selected based on their distribution in the sample and on the breakpoint determination precision.

Phylogenetic analysis of ancestral groups

Phylogenetic analysis was performed using wild representatives of genetic groups used for genome mosaic determination, accessions representing *Ensete* and other *Musa* species, and haplotypes from parent–child trios. The phylogenetic analysis consisted of four main steps (detailed in Method S5, Tables SM14 and SM15, and Figure SM12): (i) Identification for each chromosome of the largest region in which haplotypes of M_1 and M_2 could be defined using DrawRatio.py. These M_1 and M_2 regions were defined in accessions 'Pisang Madu' or 'Manang'. (ii) Identification, in selected regions, of introgressions in accessions representative of genetic groups. (iii) Production of a multifasta file from the vcf with identified non-introgressed accessions and haplotypes using PhaseInVcfToFasta.2.0.py (Table S4). (iv) Phylogenetic analysis using PHYML v3.1 (Guindon et al., 2010) with the GTR + Γ model (–m GTR –b – 5 –v e –c 4 –a e –s BEST options) and tree visualization using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) with trees manually re-rooted with *E. ventricosum* and colored using ReformatTree.py. Maximum likelihood phylogenetic analysis was performed based on chromosomal regions of 52 to 55 accessions from which representative consensus or haplotypes were reconstructed from concatenated polymorphous sites. Regions of 2.5 to 15.0 Mb were used, with a total of 87 827 to 498 129 sites and between 72 785 and 410 907 polymorphic sites (Table S4).

ACKNOWLEDGMENTS

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), and the Agropolis Fondation (ID 1504-006) 'GenomeHarvest' project through the French Investissements d'Avenir program (Labex Agro: ANR-10-LABX-0001-01). The authors wish to thank Françoise Carreel and Xavier Perrier for critical reading of the manuscript, Céline Cardi for technical assistance, Xavier Perrier for providing two DNA samples, and Christophe Jenny for discussions on different banana accessions.

AUTHOR CONTRIBUTIONS

GM designed and performed the bioinformatics analysis and developed bioinformatics programs, AC contributed to the methodology design and developed bioinformatics programs, FCB performed, interpreted, and wrote about the aneuploidy analysis, CH extracted DNA and validated accessions, NPdlR and IvdH provided plant materials, JS contributed data and information on plant materials and edited the manuscript, FS generated F1 hybrids, KL supervised sequencing, JMA supervised the sequencing project, ADH coordinated the DYNAMO banana project and edited the manuscript, NY performed the introgression analysis, and GM and NY designed the study, analyzed the results, and wrote the paper.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The vcf file will be available in the Banana Genome Hub (https://banana-genome-hub.southgreen.fr/gigwa). Chromosome painting results will be available at https://gemo.southgreen.fr/. The vcf file, allele ratio profiles, and chromosome painting results will also be available from https://banana-genome-hub.southgreen.fr/ in the Download section. The scripts were added to the vcfhunter toolbox (https://github.com/SouthGreenPlatform/VcfHunter). The Illumina reads are available under Bioprojects PRJEB58004 and PRJEB28077.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Supporting Information.

Figure S1. Methodological process for the identification of ancestry informative alleles with identified wild *Musa* representatives (Steps I, II, and III).

Figure S2. Step IV: Methodological process for the identification of ancestry informative alleles for the M-1 and M-2 ancestries of unknown origin and final chromosome painting.

Figure S3. Representations of chromosome ancestry painting of 189 wild and cultivated bananas and 12 F1 hybrids.

Figure S4. Characterization of interspecific recombination breakpoints.

Figure S5. Localization of analyzed *M. acuminata* ssp. *zebrina* and A/B genome recombination breakpoints.

Figure S6. Phylogenetic trees.

Figure S7. Normalized allele ratio profiles in *M. acuminata* ssp. *halabanensis*.

Table S1. Accession information.

Table S2. Chromosome localisation and size of aneuploidy or large insertion/deletions fragments on 2x and 3x accessions.

Table S3. Distribution of selected recombination breakpoints of *M. schizocarpa*, *M. a.* ssp. *zebrina* and A/B genomes in analysed banana diversity.

Table S4. Regions used for phylogenetic analysis.

Method S1. Identification of specific alleles for *M. balbisiana*, *M. schizocarpa*, and four *M. acuminata* genetic groups through a multivariate approach and allele clustering 2.

Method S2. Identification of additional contributors.

Method S3. Identification of the *Australimusa*-, *M. acuminata* ssp. *truncata*-, *M. acuminata* ssp. *sumatrana*-, *Ensete*-, and *M. balbisiana*-specific alleles.

Method S4. Identification of specific alleles for two unknown contributors.

Method S5. Phylogenetic analysis of ancestral groups.

Figure SM1. Neighbor joining tree on wild *Musa* accessions.

Figure SM2. Normalized ratio profiles in three accessions after allele attribution of known origins.

Figure SM3. Normalized ratio profiles in three accessions after allele attribution of unknown origin M_1 iteration 1.

Figure SM4. Normalized ratio profiles in three accessions after allele attribution of unknown origin M_1 iteration 2.

Figure SM5. Detailed normalized ratio profiles in Pisang Madu haplotypes after allele attribution of unknown origin M_1 iteration 2.

Figure SM6. Normalized ratio profiles in 11 accessions after allele attribution of unknown origin M_1 iteration 3.

Figure SM7. Detailed normalized ratio profiles in Pisang Madu haplotypes after allele attribution of unknown origin M_1 iteration 3.

Figure SM8. Normalized ratio profiles in six accessions after allele attribution of unknown origin M_2 iteration 1.

Figure SM9. Detailed normalized ratio profiles in five accessions after allele attribution of unknown origin M_2 iteration 1.

Figure SM10. Detailed normalized ratio profiles in 26 haplotypes from parent–child trios after allele attribution of unknown origin M_2 iteration 2.

Figure SM11. Detailed normalized ratio profiles in Pisang Madu and haplotypes from parent–child trios after allele attribution of unknown origin M_2 iteration 3.

Figure SM12. Example of allele ratio painting used to identify introgression in selected regions for phylogenetic analysis.

Table SM1. Accessions used to select private alleles in order to identify accessions for the multivariate approach.

Table SM2. List of accessions and chromosomes used for the identification of group-specific alleles through a multivariate approach.

Table SM3. Number of alleles per chromosome used for the multivariate analysis.

Table SM4. Number of alleles attributed to ancestral groups after each step of allele attribution.

Table SM5. Global analysis of specific alleles.

Table SM6. Accessions and origins used for the private allele approach described in Method S3.

Table SM7. Accessions and origins used for identification of unknown contributor M_1.

Table SM8. Location of regions that could not be used for unknown contributor M_1 in DYN078_EN_13_IDN075.

Table SM9. Location of regions that could not be used for unknown contributor M_1 allele identification in DYN304_Pisang_-Madu haplotypes.

Table SM10. Location of regions that could not be used for unknown contributor M_2 allele identification in DYN304_Pisang_-Madu haplotypes.

Table SM11. Accessions used for identification of unknown contributor M_2.

Table SM12. Location of regions that could not be used for unknown contributor M_2 (in five additional accessions).

Table SM13. Location of regions that could not be used for unknown contributor M_2 (in accessions in which we can access haplotypes).

Table SM14. Accessions and chromosomes selected for the phylogenetic analysis.

Table SM15. Haplotypes used in the phylogenetic analysis according to studied chromosomal regions.

REFERENCES

Ahmad, F. (2021) *Genetics and diversity of Indonesian bananas. PhD thesis*. Wageningen, The Netherlands: Wageningen University, p. 210.

Argent, G. (1976) The wild bananas of Papua New Guinea. *Notes from The Royal Botanic Garden*, **35**, 77–114.

Arnaud, E. & Horry, J.P. (1997) *Musalogue: a catalogue of Musa germplasm. Papua New Guinea collecting missions, 1988–1989*. Montpellier, France: INIBAP.

Baurens, F.C., Martin, G., Hervouet, C., Salmon, F., Yohome, D., Ricci, S. *et al.* (2019) Recombination and large structural variations shape

interspecific edible bananas genomes. *Molecular Biology and Evolution*, **36**, 97–111.

Belser, C., Baurens, F.C., Noel, B., Martin, G., Cruaud, C., Istace, B. *et al.* (2021) Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. *Communications Biology*, **4**, 1047.

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

Boonruangrod, R., Desai, D., Fluch, S., Berenyi, M. & Burg, K. (2008) Identification of cytoplasmic ancestor gene-pools of *Musa acuminata* Colla and *Musa balbisiana* Colla and their hybrids by chloroplast and mitochondrial haplotyping. *Theoretical and Applied Genetics*, **118**, 43–55.

Boonruangrod, R., Fluch, S. & Burg, K. (2009) Elucidation of origin of the present day hybrid banana cultivars using the 5′ETS rDNA sequence information. *Molecular Breeding*, **24**, 77–91.

Breton, C., Cenci, A., Sardos, J., Chase, R., Ruas, M., Rouard, M. *et al.* (2022) A protocol for detection of large chromosome variations in banana using next generation sequencing. In: Jankowicz-Cieslak, J. & Ingelbrecht, I.L. (Eds.) *Efficient Screening Techniques to Identify Mutants with TR4 Resistance in Banana: Protocols*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 129–148.

Carreel, F., Fauré, S., Gonzalez de Leon, D., Lagoda, P., Perrier, X., Bakry, F. *et al.* (1994) Evaluation de la diversite génétique chez les bananiers diploïdes (*Musa* sp). *Genetics, Selection, Evolution*, **26**, 125s–136s.

Carreel, F., Gonzalez de Leon, D., Lagoda, P., Lanaud, C., Jenny, C., Horry, J.P. *et al.* (2002) Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. *Genome*, **45**, 679–692.

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

Christelová, P., De Langhe, E., Hřibová, E., Čížková, J., Sardos, J., Hušáková, M. *et al.* (2017) Molecular and cytological characterization of the global *Musa* germplasm collection provides insights into the treasure of banana diversity. *Biodiversity and Conservation*, **26**, 1–24.

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**, 849–857.

De Langhe, E., Vrydaghs, L., De Maret, P., Perrier, X. & Denham, T. (2009) Why bananas matter: an introduction to the history of banana domestication. *Ethnobotany Research and Applications*, **7**, 165–177.

Denham, T., Barton, H., Castillo, C., Crowther, A., Dotte-Sarout, E., Florin, S.A. *et al.* (2020) The domestication syndrome in vegetatively propagated field crops. *Annals of Botany*, **125**, 581–597.

Denham, T.P., Haberle, S.G., Lentfer, C., Fullagar, R., Field, J., Therin, M. *et al.* (2003) Origins of agriculture at Kuk Swamp in the Highlands of New Guinea. *Science*, **301**, 189–193.

D'Hont, A., Paget-Goy, A., Escoute, J. & Carreel, F. (2000) The interspecific genome structure of cultivated banana, *Musa* spp. revealed by genomic DNA *in situ* hybridization. *Theoretical and Applied Genetics*, **100**, 177–183.

Donohue, M. & Denham, T. (2009) Banana (*Musa* spp.) Domestication in the Asia-Pacific Region: Linguistic and archaeobotanical perspectives. *Ethnobotany Research and Applications*, **7**, 40.

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

Eyland, D., Breton, C., Sardos, J., Kallow, S., Panis, B., Swennen, R. *et al.* (2020) Filling the gaps in gene banks: Collecting, characterizing, and phenotyping wild banana relatives of Papua New Guinea. *Crop Science*, **61**, 137–149.

Grimaldi, I.M., Van Andel, T.R. & Denham, T.P. (2022) Looking beyond history: tracing the dispersal of the Malaysian complex of crops to Africa. *American Journal of Botany*, **109**, 193–208.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Häkkinen, M. (2013) Reappraisal of sectional taxonomy in *Musa* (*Musaceae*). *Taxon*, **62**, 809–813.

Häkkinen, M. & Väre, H. (2008) Typification and check-list of *Musa L.* names (Musaceae) with nomenclatural notes. *Adansonia*, **30**, 63–112.

Hippolyte, I., Jenny, C., Gardes, L., Bakry, F., Rivallan, R., Pomies, V. *et al.* (2012) Foundation characteristics of edible *Musa* triploids revealed from allelic distribution of SSR markers. *Annals of Botany*, **109**, 937–951.

Hotta, M. (1989) Identification list of *Ensete* and *Musa* (*Musaceae*)in SE Asia and West Malesia. *Occasional Papers of the Kagoshima University Research Center for the South Pacific*, **16**, 67–75.

Janssens, S.B., Vandelook, F., De Langhe, E., Verstraete, B., Smets, E., Vandenhouwe, I. *et al.* (2016) Evolutionary dynamics and biogeography of Musaceae reveal a correlation between the diversification of the banana family and the geological and climatic history of Southeast Asia. *The New Phytologist*, **210**, 1453–1465.

Jeensae, R., Kongsiri, N., Fluch, S., Burg, K. & Boonruangrod, R. (2021) Cultivar specific gene pool may play an important role in *Musa acuminata* Colla evolution. *Genetic Resources and Crop Evolution*, **68**, 1589–1601.

Kennedy, J. (2009) Bananas and people in the homeland of genus *Musa*: Not just pretty fruit. *Ethnobotany Research & Applications*, **7**, 19.

Kolpakov, R., Bana, G. & Kucherov, G. (2003) mreps: Efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Research*, **31**, 3672–3678.

Lebot, V., Aradhya, K.M., Manshardt, R. & Meilleur, B. (1993) Genetic relationships among cultivated bananas and plantains from Asia and the Pacific. *Euphytica*, **67**, 163–175.

Li, L.F., Hakkinen, M., Yuan, Y.M., Hao, G. & Ge, X.J. (2010) Molecular phylogeny and systematics of the banana family (Musaceae) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. *Molecular Phylogenetics and Evolution*, **57**, 1–10.

Martin, G., Baurens, F.-C., Hervouet, C., Salmon, F., Delos, J.-M., Labadie, K. *et al.* (2020) 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. *et al.* (2020) Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *The Plant Journal*, **102**, 1008–1025.

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

Meijer, W. (1961) Notes on wild species of *Musa* from Sumatra. *Acta Botanica Neerlandica*, **10**, 248–256.

Nasution, R.E. (1989) Wild bananas of Indonesia. In: Siemonsma, J.S. & Wulijarni, S.N. (Eds.) *Plant resources of South-East Asia. Proceedings of the first PROSEA International Symposium, Jakarta, Indonesia*. Wageningen, Netherlands: Pudoc, pp. 281–283.

Nemeckova, A., Christelova, P., Cizkova, J., Nyine, M., Van den Houwe, I., Svacina, R. *et al.* (2018) Molecular and cytogenetic study of East African Highland banana. *Frontiers in Plant Science*, **9**, 1371.

Pedro, N., Brucato, N., Fernandes, V., André, M., Saag, L., Pomat, W. *et al.* (2020) Papuan mitochondrial genomes and the settlement of Sahul. *Journal of Human Genetics*, **65**, 875–887.

Perrier, X., Bakry, F., Carreel, F., Jenny, C., Horry, J.P., Lebot, V. *et al.* (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. *et al.* (2011) Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 11311–11318.

Perrier, X., Jenny, C., Bakry, F., Karamura, D., Kitavi, M., Dubois, C. *et al.* (2019) East African diploid and triploid bananas: a genetic complex transported from South-East Asia. *Annals of Botany*, **123**, 19–36.

Picarella, M.E. & Mazzucato, A. (2018) The occurrence of seedlessness in higher plants; insights on roles and mechanisms of parthenocarpy. *Frontiers in Plant Science*, **9**, 1997.

Poerba, Y.S., Martanti, D. & Ahmad, F. (2019) Genetic variation of wild *Musa acuminata* Colla from Indonesia. *Biotropia*, **26**, 115–126.

Pollefeys, P., Sharrock, S. & Arnaud, E. (2004) *Preliminary analysis of the literature on the distribution of wild Musa species using MGIS and DIVA-GIS.(INIBAP ed.)*. Rome, Italy: IPGRI.

Rouard, M., Droc, G., Martin, G., Sardos, J., Hueber, Y., Guignon, V. *et al.* (2018) Three new genome assemblies support a rapid radiation in *Musa acuminata* (wild banana). *Genome Biology and Evolution*, **10**, 3129–3140.

Sardos, J., Breton, C., Perrier, X., Van den Houwe, I., Carpentier, S., Paofa, J. *et al.* (2022) Hybridization, missing wild ancestors and the domestication of cultivated diploid bananas. *Frontiers in Plant Science*, **13**. https://doi.org/10.3389/fpls.2022.96922

Sardos, J., Perrier, X., Dolezel, J., Hribova, E., Christelova, P., Van den Houwe, I. *et al.* (2016) DArT whole genome profiling provides insights on the evolution and taxonomy of edible Banana (*Musa spp.*). *Annals of Botany*, **118**, 1269–1278.

Sardos, J., Rouard, M., Hueber, Y., Cenci, A., Hyma, K.E., van den Houwe, I. *et al.* (2016) A genome-wide association study on the seedless phenotype in banana (*Musa* spp.) reveals the potential of a selected panel to detect candidate genes in a vegetatively propagated crop. *PLoS One*, **11**, e0154448.

Shepherd, K. (1990) Observations on Musa taxonomy. In: Jarret, R.L. (Ed.) *Identification of genetic diversity in the genus Musa*. Los Banos, Philippines: INIBAP, pp. 158–165.

Shepherd, K. (1999) *Cytogenetics of the genus Musa*. Montpellier France: INIBAP.

Simmonds, N.W. (1962) *The evolution of the bananas*. London Great Britain: Longmans.

Simmonds, N.W. & Shepherd, K. (1955) The taxonomy and origins of the cultivated banana. *Journal of the Linnean Society of London Botany*, **55**, 302–312.

Soares, P., Rito, T., Trejaut, J., Mormina, M., Hill, C., Tinkler-Hundal, E. *et al.* (2011) Ancient voyaging and Polynesian origins. *American Journal of Human Genetics*, **88**, 239–247.

Stover, R.H. & Simmonds, N.W. (1987) *Bananas Harlow*. UK: Longman Scientific & Technical.