

Variations structurales chez le bananier caractérisation et impact sur les ségrégations chromosomiques de deux translocations réciproques originaires du groupe M. acuminata burmannica

Marion Dupouy

▶ To cite this version:

Marion Dupouy. Variations structurales chez le bananier caractérisation et impact sur les ségrégations chromosomiques de deux translocations réciproques originaires du groupe M. acuminata burmannica. Amélioration des plantes. Montpellier SupAgro, 2019. Français. NNT: 2019NSAM0054. tel-04106675

HAL Id: tel-04106675 https://theses.hal.science/tel-04106675

Submitted on 25 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Génétique et Amélioration des Plantes

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau

Portée par

Unité de recherche AGAP

Variations structurales chez le bananier : Caractérisation et impact sur les ségrégations chromosomiques de deux translocations réciproques originaires du groupe *M. acuminata burmannica*

Présentée par Marion DUPOUY Le 19 décembre 2019

Sous la direction d'Angélique D'Hont

Devant le jury composé de

Mme Anne-Marie CHEVRE, Directrice de Recherche, HDR, INRA M. Patrick WINCKER, Cadre scientifique, HDR, CEA M. Vincent RANWEZ, Professeur, HDR, Montpellier SupAgro Mme Angélique D'HONT, Cadre scientifique, HDR, CIRAD Rapporteur Rapporteur Examinateur Directeur de thèse





REMERCIEMENTS

Quelques semaines après la soutenance, il est temps, enfin, d'écrire ces lignes.

Ce travail n'aurait évidemment pas été possible sans la confiance accordée par ma directrice de thèse, Angélique D'Hont. Je la remercie pour son accueil dans l'équipe SEG, sa confiance renouvelée même dans les moments où le doute me submergeait, et son encadrement scientifique régulier et attentif.

Je remercie également Guillaume Martin pour son encadrement, sa compréhension, sa disponibilité. Dans le même élan, je remercie Franc-Christophe Baurens pour avoir toujours plongé avec enthousiasme dans les questions que je lui apportais, même lorsqu'elles me semblaient futiles. Ces discussions ont été des moments très valorisants pendant cette expérience.

Je remercie mon jury de thèse, Anne-Marie Chèvre, Patrick Wincker et Vincent Ranwez, pour le temps qu'ils ont accordé à mes travaux, leurs remarques et la discussion stimulante qui a suivi la soutenance.

Un des maillons essentiels d'une thèse est son comité de suivi. Le mien a su m'accompagner dans les moments compliqués, en m'encourageant à regarder les problèmes en face, sans complaisance mais toujours avec bienveillance. Pour cela, je remercie François Sabot, Hélène Bergès, Eric Jenczewski, Patrice This et Marc Seguin.

Merci à l'équipe SEG, particulièrement Nabila Yahiaoui, pour son accueil, l'intérêt porté à mes travaux, et nos discussions. Merci également aux autres experts « Banane » avec qui j'ai pu interagir, et aux membres du projet Genome Harvest.

Je remercie la cellule de direction d'AGAP, en particulier Patrice This, pour son accueil dans l'unité. Je me réjouis de l'accompagnement et l'encouragement de cette cellule de direction pour les activités proposées par l'association Graines d'AGAP. Je crois que la vie doctorante d'AGAP peut être une fierté pour l'unité, et que réciproquement, la bonne volonté active de l'unité bénéficie à la bonne intégration et à l'émulation des doctorants et stagiaires. J'espère que le CIRAD saura s'inspirer de ce mutualisme pour le déploiement d'une association des doctorants du CIRAD. Je remercie par ailleurs mes camarades de Graines d'AGAP, la mini-parenthèse et la Confédération des Jeunes Chercheurs pour les actions qu'ils mènent en faveur du bon déroulement des thèses de doctorat, ainsi que l'association Femmes & Science pour son programme de mentorat.

Vient le temps des remerciements individuels. Je ne saurai être exhaustive, et pour ça je présente d'avance mes excuses. Merci à Aurélien, Lauriane, Clara, Clément, Lisa, Kelly, Aurélie, Joao, Stella, Ian, Thibault et Jeanne, mes chers contemporains doctorants du CIRAD. On aura traversé ça ensemble, pour le meilleur, et pour le « bof bof ». Moments « bof bof » qui auraient pu avoir raison de ma persévérance sans l'écoute et le positivisme de Catherine Desrumaux, d'Hélène Joly et d'Emilie Benoit-Rivier. Merci à Jean-François, Maryline, Frédéric, Gautier et toutes celles et ceux qui ont accompagné les pauses animées au coin café.

Je n'en serai pas là sans les chercheuses et chercheurs qui ont accompagné mon apprentissage universitaire du métier de la recherche et ma motivation à poursuivre dans cette voie depuis la Licence : Céline Brochier-Armanet, Alexandra Louis et Eric Pelletier en particulier, ainsi que mes différents professeurs du Master BIMS de Rouen, de la Licence MIV de Lyon et du DUT bioinfo d'Aurillac.

Avoir un objectif à la fin de la thèse m'a aidé à traverser le long tunnel de la rédaction et de la formalisation des travaux. Pour cela, je remercie Daniel Cabero, Bruno Claustres et Delphine Hourcade de m'avoir choisie pour assumer la responsabilité du poste que j'occupe aujourd'hui.

Enfin, merci du plus profond de mon cœur à ma famille, particulièrement mes parents, pour leurs encouragements, leur fierté, leur soutien, leur présence et la confiance qu'ils m'ont accordée dans mon parcours post-bac. Ni vous ni moi ne savions vraiment où j'allais, mais j'y allais sereinement grâce à vous. Merci à Mirri, mon chat. C'est futile, mais tellement important... Et merci à Aurélien, d'avoir traversé ça avec moi. Tu y es presque.

Le mérite n'est qu'un mirage, cachant une réalité faite de contexte et de chance. J'ai eu la chance d'être accompagnée au long de ma vie par toutes ces personnes qui ont fait de moi qui je suis, compensant un contexte qui ne permettait pas de prédire ce parcours. Je souhaite à toutes celles et ceux à qui on a un jour dit que ça ne serait pas possible d'avoir le même « mérite » que moi.

TABLE DES MATIERES

	ments1	
Liste des Figures5		
Liste des Tableaux6		
Résumé7		
Abstract8		
Introduct	ion11	
A. Le	bananier11	
1.	Morphologie des bananiers11	
2.	Systématique et origine géographique des bananiers13	
3.	Domestication et origine des bananiers cultivés15	
4.	Le bananier, stratégie de culture actuelle et enjeux agronomiques17	
5.	Stratégies d'amélioration variétale19	
B. Le	s grandes variations structurales et leur impact sur les ségrégations chez les	
plante	5	
plante 1.	23 Les types de variations structurales	
plante 1. 2.	23 Les types de variations structurales23 Méthodes bioinformatiques de détection à partir de données génomiques.25	
plante 1. 2. 3.	23 Les types de variations structurales	
plante 1. 2. 3. C. Ré	23 Les types de variations structurales	
plante 1. 2. 3. C. Ré chez le	23 Les types de variations structurales	
plante 1. 2. 3. C. Ré chez le 1.	23 Les types de variations structurales	
plante 1. 2. 3. C. Ré chez le 1. 2.	23 Les types de variations structurales	
plante 1. 2. 3. C. Ré chez le 1. 2. 3.	23 Les types de variations structurales	
plante: 1. 2. 3. C. Ré chez le 1. 2. 3. 4.	23 Les types de variations structurales	
plante 1. 2. 3. C. Ré chez le 1. 2. 3. 4. à pa	s	

Chapitre	1
Chapitre	2: Détermination de l'impact de deux translocations sur les ségrégations
chromos	omiques dans le cas de croisements diploïdes75
A. Co	ontexte75
B. M	latériel et Méthodes81
1.	Matériel végétal et génotypage par séquençage81
2.	Appel de variants et sélection des marqueurs SNP81
3.	Liaison entre marqueurs85
4.	Analyse des ratios alléliques parentaux, de la ploïdie et des points de
reco	ombinaison85
C. Re	ésultats
1.	Liaison entre les marqueurs89
2.	Recombinaisons91
3.	Aneuploïdie91
4.	Biais de ségrégation chromosomique95
D.	Discussion
Discussio	on et perspectives
A. Va	ariations structurales chez le bananier : avancée des connaissances109
B. In	npact des translocations sur les ségrégations chromosomiques113
C. Le	es variations structurales et la structuration de la diversité
D. Lie	ens avec les programmes d'amélioration et les recherches de QTLs119
Bibliogra	phie
Annexes	

LISTE DES FIGURES

Figure 1 - Morphologie du bananier10
Figure 2 - Répartition géographique des principales espèces et sous-espèces de bananier
sauvages12
Figure 3 - Exemple de variations structurales de grande taille
Figure 4 - Approches de détection des variations structurales adapté de Guan et Sung
2016
Figure 5 - Configuration des lectures pairées pour différents types de variations
structurales
Figure 6 - Représentation schématique des appariements chromosomiques et des
gamètes possibles chez un hybride hétérozygote de structure pour une translocation
réciproque32
Figure 7 - Représentation schématique du modèle d'incompatibilité par perte de fonction
de Dodzhansky-Muller
Figure 8 - Modèle d'évolution d'un gène dupliqué chez deux espèces divergentes et
impact sur la fitness de l'hybride36
Figure 9 - Représentation des groupes caryotypiques et nombre de translocations entre
les différents groupes identifiés sur la base des figures d'appariement chromosomique à
la méiose des Musa acuminata et Musa balbisiana
Figure 10 - Structure chromosomique de 'PT-BA-00267', portant la translocation 1/448
Figure 11 - Structure chromosomique de 'CRBP 39', portant la translocation 1/352
Figure 12 - Origine, et représentation chromosomiques des chromosomes 1,2, 8 et 9 chez
les parents de la population BCM80
Figure 13 - Représentation des distorsions de ségrégation dans la population pour les
marqueurs GBS dans la population BCM82
Figure 14 - Représentation schématique d'un "chromosome painting" et des génotypes
correspondants
Figure 15 - Exemple de "chromosome painting" obtenu sur un individu de la descendance
analysée Erreur ! Signet non défini.

Figure 16 - Représentation schématique de la figure permettant d'analyser la ploïdie le
long des chromosomes Erreur ! Signet non défini.
Figure 17 - Représentation de l'intensité de l'association génétique des marqueurs
hétérozygotes du F1 dans la population BCM88
Figure 18 - Répartition des points de recombinaisons observés dans la descendance du F1
le long des chromosomes impliqués dans les réarrangements chromosomiques90
Figure 19 - Chromosome painting du chromosome 8 de 16 individus portant une
aneuploïdie segmentale90
Figure 20 - Chromosome painting de trois situations de ploïdie différentes92
Figure 21 - Chromosome painting d'un individu aneuploïde sur le segment T8 et
présentant une recombinaison dans cette région94
Figure 22 - Distorsion de ségrégation dans la population BCM94
Figure 23 - Tableau de contingence de la structure chromosomique aux points de cassure
de la translocation 1/9 des 73 individus de la population BCM96
Figure 24 - Tableau de contingence de la structure chromosomique aux points de cassure
de la translocation 2/8 des 73 individus de la population BCM96
Figure 25 - Comparaison du chromosome painting obtenu à partir de données RAD-seq et
GBS98
Figure 26 - Représentation des groupes caryotypiques et des translocations caractérisées
à l'heure actuelle
Figure 28 - Représentation des événements chromosomiques dans la diversité
géographiques des bananiers118

LISTE DES TABLEAUX

Tableau 1 - Groupes caryotypiques et nombre de translocations entre les différents
groupes
Tableau 2 - Synthèse des cartes génétiques publiées sur l'espèce Musa acuminata40
Tableau 3 - Résumé des métriques des différents génomes et assemblages disponibles de
Musa44

RESUME

Les bananiers cultivés sont issus d'hybridations entre espèces et sous-espèce du genre Musa. L'espèce Musa acuminata (génome A, 2n=2x=22, 1C=500-600 Mbp) est impliquée dans l'ensemble des cultivars à quelques rares exceptions. L'observation des appariements chromosomiques lors de la méiose chez des bananiers sauvages (séminifères) et des hybrides cultivés (stériles ou à très faible fertilité) a montré de fréquentes irrégularités chez ces derniers avec la présence de multivalents et d'univalents. Ces irrégularités suggèrent une hétérogénéité des structures chromosomiques chez les bananiers sauvages impliqués dans les hybrides. Durant mes travaux de thèse, deux translocations réciproques ont été caractérisées chez l'accession 'Calcutta 4' (M. a. ssp. burmannicoïdes) une impliquant les chromosomes 2 et 8 générant les structure 2T8 et 8T2 et une impliquant les chromosomes 1 et 9 générant les structure 1T9 et 9T1. Ces deux variations structurales ont été repérées dans un premier temps par l'étude de la liaison génétique entre marqueurs SNP dans une population issue d'une autofécondation de l'accession 'Calcutta 4'. Elles ont ensuite été caractérisées par l'analyse de lectures pairées de 'Calcutta 4' alignées sur le génome de référence de Musa acuminata (sous-espèce malaccensis). Enfin, les bornes des réarrangements ont été précisées à la base près par comparaison entre le génome de référence et un assemblage préliminaire de 'Calcutta 4'. Une méthode de détection in silico des jonctions de segments génomiques spécifiques des structures chromosomiques recherchées à partir de données de séquençage paired-end a été développée pour étudier la distribution des translocations dans la diversité Musa. Ces deux translocations ont été détectées chez les accessions sauvages uniquement dans le groupe génétique burmannica (M. a. ssp. burmannica, burmannicoïdes et siamea). Ceci permet de proposer que ces deux réarrangements aient émergé au sein de ce groupe. Seule la translocation 2/8 a été retrouvée à l'état hétérozygote chez deux accessions cultivées. Ce résultat confirme la faible implication du groupe burmannica dans le génome des bananiers cultivés. L'impact de ces deux variations structurales sur les ségrégations chromosomiques dans une population diploïde issue de parents hétérozygotes de structure a ensuite été étudié. Des distorsions de ségrégations dans les régions impliquées dans ces deux variations structurales ont été observées dans cette population. Dans le cas de la translocation 1/9, les deux chromosomes transloqués (1T9 et 9T1) ou standard (1 et 9) sont systématiquement trouvés ensemble dans les descendants. De plus, les chromosomes transloqués sont préférentiellement transmis à la descendance par rapport aux structures standards. Dans le cas de la translocation 2/8, on observe également que les chromosomes transloqués (2T8 et 8T2) sont préférentiellement transmis ensemble à la descendance par rapport aux structures standards. Aucun descendant avec les chromosomes 2 et 8T2 n'a été observé, par contre on observe des descendants avec les chromosomes 8 et 2T8. Pour ces individus, il manque donc une copie d'un segment du chromosome standard 2 : le segment de 240 kb impliqué dans la translocation 2/8. La taille réduite du segment impliqué (240 kb) pourrait expliquer que son absence dans un gamète ou sa présence en une seule copie dans un descendant soit non délétère contrairement au cas des autres fragments impliqués dans ces deux translocations (tous de l'ordre de 10Mb). Ces résultats améliorent notre connaissance de l'évolution des génomes des bananiers. Ils apportent d'autre part des informations importantes sur l'impact des translocations sur les ségrégations chromosomiques qui pourront être prises en compte pour exploiter ce groupe burmannica riche en résistances biotiques dans les programmes d'amélioration génétique.

Mots-clefs : Variations structurales, *Musa acuminata*, génomique, bioinformatique, translocation, distorsions de ségrégations.

ABSTRACT

Cultivated bananas are derived from hybridization between species and subspecies of the Musa genus. The *Musa acuminata* species (genome A, 2n = 2x = 22, 1C = 500-600 Mbp) is involved in all cultivars with a few rare exceptions. The observation of chromosome pairings during meiosis in wild banana (seedy bananas) and cultivated hybrids (sterile or very low fertility) showed frequent irregularities in the latter with the presence of multivalents and univalents. These irregularities suggest a heterogeneity of chromosome structures in wild bananas involved in hybrids. During my thesis work, two reciprocal translocations were characterized in the accession 'Calcutta 4' (M. a. ssp. burmannicoïdes) one involving chromosomes 2 and 8 generating the structure 2T8 and 8T2 and one involving chromosomes 1 and 9 generating the 1T9 and 9T1 structures. These two structural variations were first identified by studying the genetic linkage between SNP markers in a self-fertilizing population of 'Calcutta 4' accession. They were then characterized by the analysis of paired reads of 'Calcutta 4' aligned with the reference genome of Musa acuminata (subspecies malaccensis). Finally, the rearrangement breakpoints were specified at the nucleotide level by comparison between the reference genome and a preliminary assembly of 'Calcutta 4'. An in silico detection approach of genomic segment junctions specific for the desired chromosomal structures using paired-end sequencing data was developed to study the distribution of translocations in the Musa diversity. Both translocations were detected in wild-type accessions only in the *burmannica* genetic group (*M. a.* ssp. burmannica, burmannicoïdes and siamea). This suggests that these two rearrangements would have emerged within this group. Only the 2/8 translocation was found heterozygous in two cultivated accessions. This result confirms the low involvement of the burmannica group in the genome of cultivated bananas. The impact of these two structural variations on chromosomal segregations in a diploid population with structurally heterozygous parents was then studied. Segregation distortions in the regions involved in these two structural variations were observed in this population. In the case of the 1/9 translocation, both translocated (1T9 and 9T1) or reference (1 and 9) chromosomes are systematically found together in the offspring. In addition, the translocated chromosomes are preferentially transmitted to the offspring with respect to the reference structures. In the case of translocation 2/8, we also observe that the translocated chromosomes (2T8 and 8T9) are preferentially transmitted together to the offspring in relation to the reference structures. No offspring with chromosomes 2 and 8T2 were observed, but descendants with chromosomes 8 and 2T8 were observed. For these individuals, therefore, a copy of a segment of reference chromosome 2 is missing, the 240 kb segment involved in translocation 2/8. The reduced size of the segment involved (240 kb) could explain that its absence in a gamete or its presence in a single copy in a descendant is not deleterious contrary to the case of the other fragments involved in these two translocations (all of the order of 10Mb). These results improve our knowledge of the evolution of the genomes of provide important information on the impact of translocations on chromosomal segregations that could be taken into account to exploit this burmannica group rich in biotic resistance in breeding programs.

Keywords: Structural variations, *Musa acuminata*, genomics, bioinformatics, translocation, segregation distortions.







(A) Représentation schématique des organes d'un bananier d'après Champion 1963. (B) Photo de fruits de bananier séminifère (gauche) et parthénocarpique (droite) (A. D'Hont)

INTRODUCTION

A. Le bananier

1. MORPHOLOGIE DES BANANIERS (FIGURE 1A)

Les bananiers sont des herbes géantes appartenant au genre Musa. Comme d'autres représentants des monocotylédones (les palmiers, bambous...), ces plantes sont parfois à tort assimilées à des arbres en raison de leur pseudo-tronc, de leur port arborescent et de leur grande taille, mais ne produisent cependant pas de bois (xylème secondaire). En fonction des espèces, les bananiers peuvent atteindre entre 1,5 et 9 m de haut. Le rhizome constitue la partie pérenne du bananier. Il se développe sous terre et contribue à la production de racines, des parties aériennes et des rejets clonaux. Les feuilles sont formées à partir du méristème central et émergent sous la forme d'une structure appelée « cigare ». Les gaines des feuilles s'imbriquent fermement les unes dans les autres en formant un compact pseudo-tronc enveloppant le méristème central. Lors de la floraison, le méristème cesse de produire des feuilles et commence à émettre une tige aérienne, la hampe florale ou rachis. Dans un premier temps, les fleurs femelles apparaissent. Ces fleurs forment des groupes séparés par les bractées et organisés en deux rangées. Ces groupes de fleurs sont appelés des mains, et leurs ovaires se développeront en fruits contenant des graines chez les cultivars fertiles. Les fleurs mâles apparaissent à l'extrémité de l'inflorescence. On peut parfois observer des fleurs hermaphrodites à l'interface des fleurs mâles et femelles (source proMusa, http://www.proMusa.org/), qui ne produiront généralement ni fruit ni pollen. L'ensemble des mains le long de l'inflorescence constitue un régime. Chez les cultivars comestibles, les fruits se développent sans fécondation (parthénocarpie) et ne contiennent pas de graines (Figure 1B).



Figure 2 - Répartition géographique des principales espèces et sous-espèces de bananier sauvages d'après Perrier *et al*, 2011.

2. SYSTEMATIQUE ET ORIGINE GEOGRAPHIQUE DES BANANIERS

Les bananiers appartiennent à l'ordre des Zingibérales et à la famille des Musaceae. Cette dernière est divisée en trois genres : Ensete, Musella et Musa. Les bananiers du genre Ensete sont présents dans une vaste zone géographique allant de l'Afrique à l'Asie du sudest et sont cultivés pour leurs rhizomes. Les bananiers du genre Musella sont eux originaires du sud-ouest de la Chine et sont cultivés à des fins ornementales et religieuses. Les bananiers cultivés pour la consommation de leurs fruits appartiennent au genre Musa. Anciennement divisée en cinq sections - IngentiMusa (n=7), AustraliMusa (n=10), CalliMusa (n=10), EuMusa (n=11) et Rhodochlamys (n=11) -, cette classification a été réduite à deux sections sur la bases d'analyses phylogénétiques : Musa (n=11) et CalliMusa (n=7/10) (Wong et al. 2002; Li et al. 2010; Christelová et al. 2011; Häkkinen 2013). La section *Musa* comprend de nombreuses espèces dont deux sont à l'origine de la grande majorité des bananiers cultivés pour leurs fruits (Simmonds and Shepherd 1955) : Musa acuminata (génome A, 2n = 2x = 22, 1C = 500-600 Mb) et Musa balbisiana (génome B, 2n = 2x = 22, 1C = 550 Mb). *Musa balbisiana* est une espèce génétiquement relativement peu diversifiée (Carreel et al. 1994), retrouvée au sud-est de la région continentale de l'Asie (du nord de l'Inde au sud de la Chine). Musa acuminata est divisée en six à neuf sous-espèces (selon les auteurs) plus ou moins inter-fertiles, sur la base de caractères morphologiques et de leur distribution géographique (Simmonds 1962; Perrier 2009). Perrier et al. (2011) propose une classification en 7 sous-espèces dont les aires de répartition sont clairement identifiées (Figure 2). La distinction en sous-espèces proposée initialement sur la base de caractères morphologiques majeurs est globalement confirmée par marqueurs moléculaires (Carreel et al. 1994, 2002; Hippolyte et al. 2012). Toutefois les sous-espèces burmannica/burmannicoïdes/siamea n'ont pas pu être différenciées sur la base des marqueurs nucléaires et cytoplasmiques et ont été regroupées au sein d'un même groupe recouvrant une large zone continentale de l'Asie du Sud-Est par Perrier et al. (2011). La diversification en sous-espèces peut s'expliquer avec le contexte géographique insulaire de leur aire de répartition : elles auraient divergé durant le Pliocène (il y a 2 à 5 millions d'années) lors de l'isolement des différentes régions

d'Asie du Sud Est par montée du niveau de la mer (Daniells *et al.* 2001; Perrier 2009; Janssens *et al.* 2016; Rouard *et al.* 2018).

3. DOMESTICATION ET ORIGINE DES BANANIERS CULTIVES

La mise en perspective d'analyses de données de marquage moléculaire (Carreel *et al.* 1994, 2002; Boonruangrod *et al.* 2009; Risterucci *et al.* 2009; Hippolyte *et al.* 2012; Sardos *et al.* 2016; Christelová *et al.* 2017), archéologiques et linguistiques ont permis d'affiner une hypothèse quant à l'origine des bananiers cultivés et l'histoire de la domestication de cette plante (Perrier 2009; Perrier *et al.* 2011). Cette hypothèse suggère qu'une première étape d'hybridation entre des bananiers sauvages de la sous-espèce *Musa acuminata* (appelés AAw dans ce document) s'est produite lors des migrations humaines qui ont eu lieu en Asie du sud-est lors de l'Holocène, il y a 10 milliers d'années. Les populations humaines auraient transporté certains génotypes fertiles d'une région à l'autre en raison de leur exploitation parfois millénaire en tant que nourriture, fourrage, plante médicinale, ou encore comme matériau de construction. En transportant ces plantes, les populations humaines migrantes auraient alors remis en contact différentes sous-espèces isolées depuis quelques millions d'années, générant ainsi les premiers hybrides diploïdes intersubspécifiques (notés AAcv). Dans un second temps, ces hybrides ont pu se croiser entre eux ou à nouveau s'hybrider avec les génotypes sauvages.

La présence de réarrangements chromosomiques au sein des sous-espèces a été suggérée sur la base de l'observation d'irrégularités à la méiose chez les hybrides intersubspécifiques (Shepherd 1999). Ces irrégularités à la méiose ont pu ensuite perturber la formation de gamètes haploïdes viables, et favoriser l'apparition de gamètes diploïdes qui pourraient être à l'origine des génotypes triploïdes (Dodds and Simmonds 1948). Les caractères tels que la fertilité réduite (peu de graines, chair dense) ou la parthénocarpie ont ensuite été sélectionnés par l'homme au sein de ces hybrides intersubspécifiques diploïdes et triploïdes. De Langhe *et al.* (2010) suggère que ce processus a pu impliquer plusieurs évènements de croisements et/ou rétrocroisements pour aboutir aux cultivars actuels. Ces cultivars ont ensuite été multipliés végétativement

et propagés sur de grands territoires pendant des siècles voire des millénaires, aboutissant à des groupes de cultivars comprenant une certaine diversité phénotypique résultant de variations somaclonales.

4. LE BANANIER, STRATEGIE DE CULTURE ACTUELLE ET ENJEUX AGRONOMIQUES

D'après la FAO, 148 millions de tonnes de bananes et plantains ont été produites sur la planète en 2016, représentant un cumul de 5.6 millions d'hectares de terre dédiées à cette culture. Entre 15 et 20% de cette production mondiale est destinée au marché international. Le fruit le plus communément retrouvé sur nos étals est obtenu à partir des cultivars du sous-groupe 'Cavendish', appartenant au groupe des triploïdes AAA et produisant les bananes de types « dessert ». Ce sous-groupe de cultivars (constitué de mutants somaclonaux d'un seul individu) représente à lui seul plus de 50% de la production mondiale de bananes et 97% des exportations. Ces cultivars ont émergé après l'abandon des cultivars du sous-groupe 'Gros-Michel' (AAA) lors de la crise sanitaire de la maladie de Panama (Ploetz 1994) dans les années 1960. La Chine, l'Inde, l'Indonésie et le Brésil sont les principaux producteurs de bananes 'Cavendish', destinée en très grande majorité au marché intérieur. Les principaux exportateurs de bananes 'Cavendish' sont l'Equateur, le Costa Rica et le Guatemala.

Les bananes Plantains (AAB) représentent elles 15% de la production mondiale de bananes (Lescot 2014), et sont cultivées principalement en Afrique de l'Ouest et Centrale, dans les Caraïbes et en Amérique Centrale où elles sont un aliment de base. Enfin, les Lujugira-Mutika ou « Highland bananas » (AAA) sont cultivés en Afrique de l'Est sous forme de cultures vivrières.

Jusqu'à présent, l'exploitation du bananier est basée sur la monoculture, facilitant l'adaptation des pathogènes tels que les champignons (la fusariose *-fusarium oxysporum* f.sp. *cubense-*, à l'origine de la maladie de Panama par exemple), les virus, les bactéries, les insectes ou les nématodes. Lutter contre ces pathogènes requiert l'usage important de produits phytosanitaires (Cretin and Triquenot 2018). Exploiter la diversité génétique

des bananiers pour proposer de nouveaux cultivars plus résistant sans délaisser les caractères d'intérêt agronomique qui ont fait le succès des cultivars actuels est urgent face au rôle important que joue ce fruit dans la sécurité alimentaire mondiale et à la fragilité relative de son mode de culture.

5. STRATEGIES D'AMELIORATION VARIETALE

Les programmes d'amélioration du bananier ont pour objectif d'obtenir des cultivars résistants aux maladies et prédateurs, avec de bons rendements et produisant des fruits de bonne qualité. Les critères de qualité varient selon les usages (dessert, à cuire, ...) et les goûts des consommateurs des différentes régions. D'autre part, pour les cultivars exportés s'y ajoutent des critères très stricts de conservation et d'adaptabilité aux conditions de transport établis par la filière d'exportation. Les bananiers triploïdes semblent être les plus propices à apporter un équilibre entre une fertilité très réduite (qui va de pair avec la parthénocarpie et la comestibilité des fruits) et de bonnes performances agronomiques. La principale contrainte des programmes d'amélioration du bananier est que les cultivars sont stériles ou très peu fertiles et souvent triploïdes. Deux principales stratégies sont utilisées dans les programmes d'amélioration pour produire ces bananiers triploïdes (Bakry *et al.* 2009) :

- la première est basée sur l'exploitation de la fertilité résiduelle des cultivars grâce au croisement d'un cultivar triploïde à améliorer avec une accession diploïde possédant des caractéristiques recherchées (par exemple la résistance à des maladies). Cette stratégie est difficile à mettre en place car le nombre de descendants obtenus dans ce type de croisement est souvent très faible. D'autre part ces descendances sont composées d'individus avec différents niveaux de ploïdie dont de nombreux aneuploïdes. Dans ces descendances les individus tétraploïdes sont ensuite sélectionnés pour être utilisés comme géniteurs dans des croisements avec une accession diploïde pour obtenir une descendance triploïde parmi lesquels des cultivars améliorés seront sélectionnés. Dans certains cas les accessions tétraploïdes peuvent aussi être directement sélectionnées pour être valorisées en plantation.

 La deuxième stratégie dite de "reconstruction" consiste au doublement chromosomique (à la colchicine) d'un cultivar diploïde, devenant donc tétraploïde, et qui est ensuite croisé avec une accession diploïde afin d'obtenir une descendance triploïde. Cette stratégie est utilisée dans le programme d'amélioration génétique du CIRAD en Guadeloupe.

Historiquement, des accessions sauvages ou des cultivars locaux issus de prospections étaient utilisés comme géniteurs dans ces stratégies d'amélioration afin d'apporter différents caractères de résistance et d'assurer la fertilité. Cependant les descendances hybrides résultant de ces croisements héritaient souvent de caractères rédhibitoires pour la qualité du fruit, encourageant certains programmes à entreprendre au préalable des phases d'amélioration des géniteurs diploïdes.

D'autre part les analyses de diversité au moyen de marqueurs moléculaires ont permis d'identifier des accessions diploïdes comme probables contributrices du cultivar 'Cavendish' (Raboin *et al.* 2005; Hippolyte *et al.* 2012). Depuis, ces accessions sont utilisées comme parents dans les programmes d'amélioration.



Figure 3 - Exemple de variations structurales de grande taille.

B. Les grandes variations structurales et leur impact sur les ségrégations chez les plantes

1. LES TYPES DE VARIATIONS STRUCTURALES

Les variations structurales correspondent à des variations de la séquence génomique. Elles peuvent être de différentes ampleurs, du réarrangement chromosomique impliquant des régions de plusieurs centaines de milliers de bases aux variations d'un seul nucléotide (Single Nucleotide Polymorphism, SNP). Les grandes variations structurales que sont les réarrangements chromosomiques comportent les duplications, les délétions, les inversions et les translocations (Figure 3, Feuk et al. 2006). Les duplications et délétions correspondent à une variation du nombre de copies d'une séquence entre deux génomes. Cette copie est surnuméraire en cas de duplication et sous-numéraire en cas de délétion. Lorsque le nombre de copies impliquées est élevé, on parle alors de variation du nombre de copies (Copy Number Variation, CNV). Les inversions correspondent à des régions génomiques retrouvées à l'envers dans un génome par rapport à un autre. Enfin les translocations correspondent à des séquences retrouvées à une position différente entre deux génomes comparés. Cette translocation est dite réciproque si deux positions échangent mutuellement un fragment. Le cas particulier de la translocation réciproque entre deux chromosomes acrocentriques au niveau de leurs centromères, aboutissant à un chromosome avec deux bras longs et un chromosome avec deux bras courts (souvent perdu au cours des divisions cellulaires), est robertsonienne (Robertson appelé translocation 1916). Les réarrangements chromosomiques représentent une question de recherche majeure depuis Dobzhansky et Sturtevant (1938) et sont notamment des mécanismes importants impliqués dans la spéciation et l'évolution (White 1962; Rieseberg 2001).



Figure 4 – Approches de détection des variations structurales adapté de Guan et Sung 2016.

2. METHODES BIOINFORMATIQUES DE DETECTION A PARTIR DE DONNEES GENOMIQUES

La disponibilité toujours croissante de séquences de référence pour de nombreux organismes et les nouvelles facilités de séquençage ont permis le développement de nombreux outils bioinformatiques pour la recherche de variations structurales. Quatre principales approches sont utilisées pour la découverte de ces variations à partir de données de re-séquençage et d'un génome de référence (revues par Guan and Sung 2016, Figure 4) :

- Le clustering de lectures pairées, consistant à regrouper les lectures discordantes (lectures présentant des configurations d'alignement sur la séquence de référence non attendues) dans une même région pour identifier les régions soumises à une variation structurale ;
- L'alignement de lectures découpées (split-reads), consistant à aligner sur le génome de référence les fragments non mappés des lectures découpées pour déterminer les points de cassure ;
- L'assemblage des contigs, consistant à l'assemblage local des lectures dans la région de la variation de structure supposée afin de comparer le nouveau consensus à la séquence de référence ;
- Les tests statistiques appliqués aux variations de couverture, permettant de détecter les CNV par exemple.

L'approche de clustering des lectures discordantes consiste à utiliser des lectures pairées (paires de séquences d'ADN obtenues par séquençage des extrémités d'un fragment d'ADN de taille connue) alignées sur un génome de référence. Les lectures pairées dont la configuration d'alignement est différente de la configuration attendue sont identifiées comme paires discordantes (Figure 5). Lorsqu'elles sont nombreuses dans une même région, elles sont regroupées, apportant de la robustesse dans la détection d'une variation structurale. Cette approche permet de détecter les insertions, les délétions, les inversions et les translocations inter et intrachromosomiques. Plusieurs programmes se basant sur cette approche existent parmi lesquels SVdetect (Zeitouni *et al.* 2010), BreakDancer (Chen *et al.* 2009) ou encore Scaffremodler (Martin *et al.* 2016).



Figure 5 - Configuration des lectures pairées pour différents types de variations structurales (extrait de Martin 2014)

Les lectures sont obtenues par re-séquençage d'une accession puis alignées sur la séquence de référence. L'orientation attendue des paires est « reverse » (flèche vers la gauche) – « forward » (flèche vers la droite). (A) Délétion chez le re-séquencé : les lectures pairées (flèches rouges) chevauchent la zone manquante chez le re-séquencé mais présente chez la séquence de référence. La discordance se traduit par des lectures dont la taille d'insert est beaucoup plus importante qu'attendue. (B) Insertion chez le reséquencé : les lectures pairées (flèches jaunes) chevauchent la région inexistante chez la référence. La discordance se traduit par des lectures dont la taille d'insert est plus petite qu'attendue. (C) Inversion chez le re-séquencé : les lectures pairées (flèches violettes et vertes) chevauchent la zone de rupture de synténie liée à l'inversion. La discordance se traduit par un changement de l'orientation d'une des lectures de la paire (reverse-forward vers forward-forward, flèches vertes ou reverse-reverse, flèches violettes) et peut être cumulée à une variation de la taille de l'insert par rapport à la taille attendue. (D) Duplication en tandem chez le re-séquencé : les lectures pairées (flèches bleues) à cheval sur deux motifs répétés chez le re-séquencé se retrouvent alignées sur la référence avec une orientation opposée (forward-reverse) à l'orientation attendue et peut être cumulée à une variation de la taille de l'insert. (E) Translocation : les lectures pairées (flèches de couleur) montrent alors plusieurs types de discordances. Les lectures chevauchant une extrémité de la zone transloquée chez le re-séquencé montrent une discordance de type délétion (bonne orientation mais trop grand insert, flèches marrons) alors que les lectures de l'autre extrémité (forward-reverse, flèches bleues) montrent une discordance de type duplication en tandem. Les lectures chevauchant la zone qui contenait le segment transloqué chez la référence se retrouvent avec une discordance de type délétion (bonne orientation mais trop grand insert, flèches rouges).

La seconde approche exploite les lectures découpées, c'est-à-dire partiellement alignées sur le génome de référence, en tentant d'aligner la fraction de ces lectures qui n'a pas pu l'être lors de l'étape de mapping. Si ces lectures sont partiellement alignées en raison de la présence d'une variation structurale, deux cas pourront être observés après la nouvelle étape d'alignement. Si l'intégralité de la lecture s'aligne sur le génome, mais de façon fractionnée à plusieurs bases de distance, alors une délétion dans l'accession reséquencée est détectée. Si les extrémités 5' et 3' de la lecture s'alignent consécutivement mais qu'une fraction centrale ne s'aligne pas, alors cette fraction centrale non alignée correspond à une insertion dans le génome re-séquencé. Utilisée seule, cette approche permet de détecter des insertions et des délétions dont la taille est inférieure à la taille de la lecture partiellement alignée.

L'approche d'assemblage local prend le parti d'utiliser les lectures pairées pour construire des contigs, qui, une fois suffisamment longs seront comparés à la séquence de référence pour détecter une variation de structure. Cet assemblage peut se faire *de novo* ou à l'aide d'une séquence de référence comme par exemple l'outil Socrates (Schröder *et al.* 2014). Cette méthode est soumise aux challenges classiques de l'assemblage, comme la présence de régions faiblement couvertes et de séquences répétées. Elle permet de détecter des évènements complexes, comme des intrications d'insertions et de délétions, ou des insertions de séquences nouvelles. Les réarrangements de grande ampleur et interchromosomiques (translocations) peuvent nécessiter l'assemblage d'une très grande région du génome, ce qui peut freiner leur caractérisation avec cette approche.

L'approche d'analyse des variations de couverture consiste à rechercher des régions génomiques où la couverture en lectures est significativement différente de la couverture attendue. Celle-ci peut être estimée en utilisant la distribution moyenne des couvertures sur une fenêtre donnée ou en utilisant un génome contrôle. Cette approche se base sur le postulat que l'échantillonnage des différentes zones d'un génome est aléatoire et indépendant, ce qui n'est pas systématiquement le cas. Le pourcentage en GC (Dohm *et al.* 2008) et la densité de la chromatine, qui impacte l'accessibilité de certaines régions

d'un génome, sont deux des facteurs qui peuvent avoir des conséquences sur l'échantillonnage et l'alignement des lectures. L'analyse des couvertures permet de détecter les indels et les variations du nombre de copies d'une région (CNV) mais ne permet pas de détecter les translocations ou les inversions.

Une combinaison des différentes approches peut être utilisée pour affiner la détection des points de cassure des variations structurales et exploiter au mieux les données de séquençage et d'alignement. C'est le cas par exemple de Socrates (Schröder *et al.* 2014), qui combine l'approche d'alignement des lectures découpées et d'assemblage local, de DELLY (Rausch *et al.* 2012) et Meerkat (Yang *et al.* 2013) qui combinent l'approche de clustering des lectures pairées discordantes et d'alignement des lectures découpées, ou encore de Scalpel (Narzisi *et al.* 2014) qui combine les approches de clustering des lectures pairées discordantes et d'alignement des approches de clustering des lectures pairées discordantes découpées et d'assemblage local *de novo*. Les outils développés pour la recherche de variations structurales sont majoritairement développés dans le but d'analyser le génome humain et l'adaptation à l'analyse des génomes végétaux, parfois très riches en éléments transposables (Lee and Kim 2014) peut être complexe.

3. CARTOGRAPHIE GENETIQUE ET DISTORSIONS DE SEGREGATIONS

La cartographie génétique consiste à organiser et ordonner en groupes de liaison des marqueurs génétiques. Les groupes de liaison sont construits sur la base des distances génétiques établies entre les marqueurs, elles-mêmes calculées à partir des taux de recombinaison entre marqueurs. Le taux de recombinaison mesuré entre deux marqueurs très proches est très faible car ils sont physiquement liés et ont de grandes chances d'être transmis ensemble sans qu'un crossing-over ne les sépare. Le brassage intrachromosomique (crossing-over ou recombinaison) coupe la liaison génétique entre les marqueurs, et le taux de recombinaison augmente au fur et à mesure que les marqueurs s'éloignent et que le nombre de crossing-over entre eux augmente. Plus le nombre de crossing-overs entre deux marqueurs situés sur un même chromosome est important plus la distance génétique entre ces marqueurs est grande (Haldane 1919;

Kosambi 1943). Les marqueurs présents sur des chromosomes différents ségrégent eux de façon indépendante et ont donc un taux de recombinaison entre marqueurs proches de 50%, correspondant au brassage interchromosomique.

D'autre part, les co-ségrégations entre marqueurs présents sur des chromosomes différents affectent la création des cartes génétiques en générant des pseudo-liaisons entre marqueurs pourtant non liés physiquement. Différents mécanismes discutés juste après peuvent expliquer ces co-ségrégations. Ces marqueurs peuvent alors être rassemblés dans un même groupe de liaison génétique alors que physiquement ils ne sont pas associés, comme cela a pu être remarqué chez le caféier (Lashermes *et al.* 2001), le colza (Cloutier *et al.* 1997), la luzerne (Kaló *et al.* 2000) ou encore le bananier (Hippolyte *et al.* 2010). Les conséquences sur les cartes génétiques s'étendent ensuite à l'amélioration végétale. En effet, une cartographie précise des caractères quantitatifs (QTL) est très utile pour la mise en place de programmes d'amélioration assistés par marqueurs et pour la sélection de géniteurs. Tout groupe de liaison artéfactuel dû à d'importantes distorsions de ségrégation gent des marqueurs ou à un ordre erroné des marqueurs le long d'un groupe de liaison peut diminuer la précision de la localisation des QTL et le développement de marqueurs efficaces dans le cadre de la sélection assistée par marqueurs par exemple.

Lorsque les ratios observés dans une population ségrégeante s'éloignent significativement du ratio attendu (ratio mendélien), on est en présence d'une distorsion de ségrégation. La significativité de cette distorsion est évaluée statistiquement, en général à l'aide un test de χ^2 de Pearson. Les distorsions de ségrégations sont très fréquemment observées chez les plantes. Elles ont été décrites pour la première fois chez le maïs (Mangelsdorf and Jones 1926) et ont pu être observées entre autre chez le riz (Nakagahara 1972), la tomate (Paterson *et al.* 1988), l'orge (Graner *et al.* 1991), le sorgho (Pereira *et al.* 1994), la luzerne (Echt *et al.* 1994), les agrumes (Luro *et al.* 1995), le café (Ky *et al.* 2000) et le chêne (Bodénès *et al.* 2016).

Les distorsions sont le résultat des biais dans la transmission des allèles à la descendance, qui peuvent être de deux types (revus par Maheshwari et Barbash 2011, Hassedine Diouf et Mergeai 2012):



Figure 6 - Représentation schématique des appariements chromosomiques et des gamètes possibles chez un hybride hétérozygote de structure pour une translocation réciproque (d'après Martin *et al.* 2017).

Les chromosomes (gris et noirs) sont représentés ainsi que leurs homologues avec des fragments transloqués colorés. Sur les types de gamète possibles issus ou non de bivalents, la moitié sont déficients pour l'un ou l'autre des fragments transloqué (unbalanced).

 une sélection gamétique. La distorsion est alors la conséquence d'une méiose produisant des gamètes moins viables ou moins fertiles.

 une sélection zygotique. La distorsion est alors la conséquence d'une différence de viabilité entre les descendants d'une même population hybride.

Les variations de structures chromosomiques génèrent de l'hétérozygotie structurale chez les hybrides. Lors de la méiose chez l'hybride, l'appariement des chromosomes est perturbé par ces variations de la structure des chromosomes. Ces perturbations de l'appariement peuvent favoriser la formation de multivalents et/ou d'univalents ou laisser des zones chromosomiques non appariées. Ensuite, à l'anaphase et en fonction de comment les chromosomes s'apparient, leur répartition dans les cellules filles peut provoquer l'absence ou la duplication de régions chromosomiques dans les gamètes (Figure 6). On peut supposer que les gamètes dits déséquilibrés qui en découlent sont moins viables et donc contre sélectionnés. Ainsi, une sélection gamétique peut découler de ces hétérozygoties structurales générant des co-ségrégations et des distorsions de ségrégation. Fauré et al. (1993) suppose que la fertilité des gamètes issus de ces résolutions de multivalents et univalents dépend de la taille et de la nature des régions chromosomiques impliquées. Ces méioses perturbées aboutissent à des ségrégations distordues de régions chromosomiques entières. Ce lien entre réarrangements chromosomiques et distorsions de ségrégations a été établi chez le prunier (Jáuregui et al. 2001), le tournesol (Quillet et al. 1995), les lentilles (Tadmor et al. 1987) et plus récemment chez le bananier (Martin et al. 2017; Baurens et al. 2019, détaillés au paragraphe C. 4).

D'autres facteurs que les variations structurales peuvent être à l'origine de sélections gamétiques ou zygotiques aboutissant à des distorsions de ségrégation :

 Les gènes "gamétocides" : ce sont des éléments dits "égoïstes" dont un allèle dominant est préférentiellement transmis à la descendance au dépend des gamètes portant l'allèle récessif (qui sont alors avortés), sans pour autant apporter d'avantage phénotypique à la plante. Ce phénomène a été répertorié chez le tabac (Cameron and Moav 1957), le riz (Sano 1990) ou encore l'égilope (Niranjana 2017).


Figure 7 - **Représentation schématique du modèle d'incompatibilité par perte de fonction de Dodzhansky-Muller** (adapté de Maheshwari et Barbash 2011).

Un couple de protéines (pièces de puzzle) interagissant l'une avec l'autre (zone orange) évoluent de façon différentes dans les espèces proches et ne peuvent plus interagir correctement dans un hybride interspécifique.

- La sélection préférentielle comme conséquence des mécanismes des divisions méiotiques (dit "meiotic drive" au sens strict, Sandler and Novitski 1957; Lyttle 1991; Lindholm *et al.* 2016) : ce mécanisme a été précisément décrit chez le maïs (Buckler *et al.* 1999). Chez cette espèce, un chromosome 10 anormal (Ab10) présentant une activité néocentromérique à une de ses extrémités a été détecté. Ce néocentromère favorise le chromosome Ab10 lors de l'anaphase II, qui migre alors préférentiellement vers la position de la mégaspore à l'origine de l'oosphère. Ce chromosome et ses allèles associés sont alors retrouvés plus fréquemment dans la descendance, entraînant des distorsions de ségrégations.
- Les gènes "Dobzhansky-Muller" (Dobzhansky 1937; Muller 1942; revue par Maheshwari and Barbash 2011) : ces gènes causent une incompatibilité chez les hybrides en raison des interactions différentielles qu'ils peuvent générer à l'état hétérozygote (chez les hybrides) par rapport à l'état homozygote (dans les lignées pures). Ce modèle suppose la fixation d'une mutation dans une population, mutation dont l'allèle porteur est ensuite incompatible avec d'autres allèles présents dans d'autres génotypes (Figure 7). Chez les plantes, ce phénomène a été décrit par exemple entre différentes souches d'Arabidopsis thaliana ou entre les sous-espèces indica et japonica du riz (revue par Sweigart and Willis 2012).
- Pertes de gènes (Lynch et Force 2000) : que ce soit au travers des duplications ancestrales de génomes complets, de duplications en tandem ou de la polyploïdie, l'information génétique à un locus est redondante. Cependant, plusieurs mécanismes peuvent aboutir à la suppression d'un gène fonctionnel sur un chromosome, provoquant ainsi des déséquilibres de dosage ou des absences du gène lors de l'hybridation avec un autre génotype (Figure 8) :
 - La pseudogénisation, c'est à dire la perte de fonctionnalité d'un gène paralogue à la suite de plusieurs événements de mutation. L'implication de ce mécanisme dans la sélection zygotique a été démontrée chez *Arabidopsis thaliana* par exemple (Bikard *et al.* 2009). Pour le gène étudié, une lignée en porte une copie non exprimée sur le chromosome 1 et une copie exprimée sur le chromosome 5, et l'autre lignée ne possède ce gène



Figure 8 - Modèle d'évolution d'un gène dupliqué chez deux espèces divergentes et impact sur la fitness de l'hybride (d'après Martin, 2014).

Chez une population ayant deux chromosomes (bleu et noir), un gène (point rouge) est dupliqué entre le chromosome bleu et le chromosome noir. Suite à un mécanisme d'isolation, la population s'est scindée en deux populations qui ont divergées, avec en particulier la perte d'une copie du gène dupliqué chez les deux espèces mais chacune à partir d'un chromosome différent. La formation d'un hybride entre ces deux espèces et son autofécondation entraîne la formation de zygotes dont une partie est déficiente pour le gène anciennement dupliqué. L'absence de ce gène peut impacter la survie de ce type de zygote (case rouge). Par ailleurs, si ce gène a un rôle important dans la formation des gamètes, un certain nombre de zygotes (cases oranges et rouge) seront observés avec des fréquences inférieures à celle attendues. que sur le chromosome 1. L'autofécondation des hybrides de ces deux lignées aboutis à la mortalité de certains embryons n'ayant aucune copie exprimée.

Réarrangements chromosomiques : ces derniers peuvent provoquer la perte de régions chromosomiques.



Musa balbisiana

Figure 9 - Représentation des groupes caryotypiques et nombre de translocations entre les différents groupes identifiés sur la base des figures d'appariement chromosomique à la méiose des Musa acuminata et Musa balbisiana (adapté d'après Shepherd, 1999).

Chaque trait noir représente une translocation suggérée par Shepherd entre les groupes considérés.

Tableau 1 - Groupes caryotypiques et nombre de translocations entre les différents groupes identifiés sur la base des figures d'appariements chromosomiques à la méiose des Musa acuminata et Musa balbisiana (d'après Shepherd, 1999, extrait de Martin, 2014).

	group	Standard (ban, mic, mal) ^a	Northern Malayan (mal) ^a	Malayan highland (tru) ^a	Northern 1 (bur2, sia) ^a	Northern 2 (bur1, sia) ^a	Javanese (zeb) ^a	East Africa (ind) ^a
0	Northern Malayan	1						
Musa acuminat	Malayan highland	3	4					
	Northern 1	2	3	*				
	Northern 2	3	4	*	1			
	Javanese	1	2	?	3	4		
	East Africa	1	2	?	3	4	2	
	Musa balbisiana	1	?	?	2	3	?	?

a : ban, banksii ; mic, microcarpa ; mal, malaccensis ; tru, truncata ; bur1, burmannica ; bur2, burmannicoides ; sia, siamea ; zeb, zebrina et ind, ssp indéterminés.

* différence structurale supposée mais non identifiée

? Pas de données

C. Réarrangements chromosomiques et impact sur les ségrégations chromosomiques chez le bananier

1. APPARIEMENTS CHROMOSOMIQUES

De nombreuses analyses cytogénétiques ont permis d'étudier les appariements chromosomiques à la méiose chez les bananiers sauvages, cultivés et chez les hybrides inter(sub)spécifiques (Dodds 1943; Dodds and Simmonds 1948; Dessauw 1987; Fauré *et al.* 1992; Shepherd 1999; Vilarinhos 2004; Hippolyte *et al.* 2010; Jeridi *et al.* 2011). Au sein des sous-espèces de *Musa acuminata*, les appariements se font généralement sous la forme attendue de onze bivalents. En revanche, chez les accessions issues d'hybridations interspécifiques ou intersubspécifiques, des irrégularités sont fréquemment observées avec la présence d'univalents et de multivalents.

Ces irrégularités d'appariements ont été interprétées comme résultants de réarrangements chromosomiques au sein des sous-espèces générant de l'hétérozygotie de structure chez les hybrides (Dodds and Simmonds 1948; Shepherd 1999).

Shepherd a réalisé en 1999 une synthèse des travaux de cytogénétiques sur le bananier à partir de laquelle il propose l'existence de 9 translocations réparties dans sept groupes caryotypiques structurant la diversité sauvage des *Musa acuminata* : les groupes Standard, Northern Malayan, Northern 1, Northern 2, Malayan Highland, Javanese et East African (Figure 9, Tableau 1). Cette classification en groupes caryotypiques se superpose partiellement avec la classification en sous-espèces. Le groupe Standard comprend les sous-espèces *microcarpa, banksii* et la plupart des bananiers de la sous-espèce *malaccensis.* Le groupe Northern Malayan comprend quant à lui les autres accessions de la sous-espèce *burmannicoïdes* et *siamea* et le groupe Northern 2 des bananiers des sous-espèces *burmannicoïdes* et *siamea* et le groupe Northern 2 des bananiers de la sous-espèce *sous-espèce truncata*. Le groupe Javanese est définit par des accessions de la sous-espèce *zebrina*. Enfin, le dernier groupe East African a été déterminé à partir d'une accession originaire de Tanzanie qui n'a pas pu être formellement identifiée. Les

		Sous-espèce de				Nombre de		
		Musa	Taille de la	Nombre de		groupes de	Groupes de liaisons	
Croisement	Carte	acuminata	population	marqueurs	Marqueurs distordus (%)	liaisons	les plus distordus	Références
auto-fécondation	SFB5	hybride ¹	92	77	36% ^ª	15	SFB5-LG12	Fauré <i>et al</i> . 1993
auto-fécondation	M53	hybride ²	89	185	26% ^b	18	M53-LG9, M53-LG7	Noyer <i>et al</i> . 1997
auto-fécondation	CAM	hybride ³	154	120	59% ^ª	14	CAM-LG1, CAM-LG2, CAM-LG5	Vilarinhos 2004
TMB _{2x} 6142-1 x TMB _{2x} 8075-7	TMB _{2x} 6142-1	hybride4	81	231	14.5% ^b	15	A11, A12, A14	Mbanjo <i>et al.</i> 2012
6142-1-S x TMB _{2x} 8075-7	6142-1-S	burmannica	58	152	8.2% ^b	16	В2	Mbanjo <i>et al.</i> 2012
multi parentale	TMB _{2x} 8075-7	hybride⁵	139	316	40.6% ^b	15	C1, C6, C7, C9	Mbanjo <i>et al.</i> 2012
Bornéo x Pisang lilin	Bornéo	microcarpa	180	261	12% ^b	11	chromosomes 6, 8	Hippolyte <i>et al</i> . 2010
Bornéo x Pisang lilin	Pisang lilin	malaccensis	180	359	24% ^b	11	chromosomes 1, 2, 3, 4, 10	Hippolyte <i>et al.</i> 2010
auto-fécondation	Pahang	malaccensis	180	652	17% ^b	11	chromosomes 1, 4	D'Hont <i>et al.</i> 2012

Tableau 2 - Synthèse des cartes génétiques publiées sur l'espèce Musa acuminata (adapté de Martin, 2014)

^a: déviation significative par rapport à une ségrégation mendélienne $\chi 2 p < 0.05$, ^b: déviation significative par rapport à une ségrégation mendélienne $\chi 2 p < 0.05$, ¹: génotype hybride comprenant une composante banksii, ²: génotype hybride comprenant une composante malaccensis et banksii, ³: génotype hybride comprenant une composante burmannicoides et banksii, ⁴: génotype hybride comprenant une composante burmannica, ⁵: génotype hybride comprenant une composante burmannicoides

bananiers au sein de chaque groupe caryotypique sont structurellement homozygotes. Chaque groupe diffère d'un autre par une à quatre translocations. Deux d'entre elles seraient partagés par deux groupes (Northern 1 et Northern 2). D'une manière générale, les correspondances entre les noms des accessions utilisés à l'époque des études de Shepherd et leurs versions actuelles en collection doivent être interprétées avec précaution.

2. CARTOGRAPHIE GENETIQUE DU BANANIER

A l'heure actuelle, neuf cartes génétiques de l'espèce *Musa acuminata* et deux cartes de l'espèce *Musa balbisiana* sont disponibles. Le résumé des caractéristiques des neuf cartes génétiques de l'espèce *Musa acuminata* est présenté dans le Tableau 2.

La production de carte génétique chez le bananier se heurte à plusieurs obstacles. La taille des descendances est souvent faible, les accessions parentales ont souvent des fertilités réduites et l'obtention de ces descendances nécessite de réaliser du sauvetage d'embryons par culture in-vitro (Bakry 2008), d'autre part un taux de mortalité élevé est souvent observé lors du maintien des populations en champs.

Par ailleurs, lors de la construction des cartes génétiques dans l'espèce M. *acuminata*, de fortes distorsions de ségrégations ont toujours été observées. L'existence d'hétérozygotie structurale a régulièrement été avancée comme hypothèse pour expliquer ces distorsions de ségrégations. Ces hypothèses ont plus particulièrement été creusées dans les quatre études ci-dessous.

Lors de la réalisation des cartes génétiques de 'Borneo' (*M. a.* ssp. *microcarpa*) et 'Pisang Lilin' (AAcv apparenté à la sous-espèce *malaccensis*), de fortes distorsions ont été observées. L'étude des méioses du parent 'Borneo' par cytologie a permis de mettre en évidence dans 30% des observations des appariements chromosomiques perturbés (univalents, trivalents, tétravalents, pentavalents et hexavalents), aboutissant à la conclusion d'une probable hétérozygotie structurale pour au moins deux réarrangements impliquant trois groupes de liaison (Hippolyte *et al.* 2010). Ces observations étaient en contradiction avec les études précédentes qui présentaient l'accession 'Borneo' comme

homozygote de structure sur la base de sa très bonne fertilité. D'autre part, l'étude des méioses du parent 'Pisang Lilin' a mis en évidence la présence d'appariements perturbés dans 60% des observations (Hippolyte *et al.* 2010). Sur la base de l'interprétation des figures d'appariement observées en cytologie, de représentations graphiques arborées des distances génétiques entre marqueurs et d'une approche par simulation des distorsions de ségrégations produites par la présence de réarrangements chromosomiques, il a été proposé l'existence d'une duplication entre les groupes de liaison 1 et 4 et une inversion au sein du groupe de liaison 10 à l'état hétérozygote.

Des représentations graphiques arborées des distances génétiques entre marqueurs ont également été réalisée lors de la production des cartes génétiques obtenues à partir des croisements impliquant une accession triploïde du groupe Lujugira-Mutika ('Nyamwihogora'), un cultivar diploïde AA ('SH-3362'), un cultivar diploïde proche de l'accession 'Long Tavoy' (*M. a.* ssp. *burmannica*) et deux accessions sauvages ('Calcutta 4', *M. a.* ssp. *burmannicoïdes* et 'Long Tavoy'). L'analyse de ces représentations a permis à Mbanjo *et al.* (2012) de proposer l'existence d'une translocation à l'état hétérozygote chez le parent mâle ainsi que d'autres variations de structures indéterminées impliquant un groupe de liaison chez chaque parent femelle.

La dernière carte génétique réalisée concernant l'espèce *Musa acuminata* est celle utilisée pour ancrer l'assemblage de la séquence de référence du génome A du bananier (D'Hont *et al.* 2012). La population sur laquelle est construite cette carte est issue de l'autofécondation de l'accession PT-BA-00267 (AAw, *M. a.* ssp. *malaccensis*). Les groupes de liaison 1 et 4, correspondant aux chromosomes 1 et 4, regroupent 72% des marqueurs distordus. L'analyse approfondie des ratios alléliques et d'arbres des distances entre marqueurs a suggéré une hétérozygotie structurale au niveau de ces chromosomes qui a été ensuite démontrée et caractérisée par Martin *et al.* (2017).

Les ségrégations chez un hybride impliquant les génomes A et B, l'accession 'CRBP39' (AAAB), ont été analysées à partir d'un croisement entre 'CRBP39' et l'accession 'Pahang' (AAw, *M. a.* ssp. *malaccensis*). Dans cette population, la moitié des descendants présentent des anomalies dans la transmission de certains allèles. La majorité des marqueurs des chromosomes 1 et 3 sont distordus. Les auteurs rapportent la transmission de trois allèles ou d'un seul par le parent CRBP39 tétraploïde au lieu de deux et des liaisons

	<i>M. a. ssp malaccensis</i> 'Pahang HD' Version 1	<i>M. a. ssp malaccensis</i> 'Pahang HD' Version 2	Musa balbisiana	Musa schizocarpa	<i>M. a. ssp burmannica</i> 'Calcutta 4'	M. a. ssp. zebrina 'Maia Oa'	<i>M. a. ssp banksii</i> 'Banksii'
Référence	D'Hont <i>et al.,</i> 2012	Martin <i>et al.,</i> 2016	Wang et al., 2019	Belser <i>et al.</i> , 2018	Rouard <i>et al.,</i> 2018	Rouard <i>et al.,</i> 2018	Rouard <i>et al.,</i> 2018
Technologie	454 + Sanger + Illumina	Illumina+ BioNano + DArT & SSR	Illumina + PacBio + Hi-C	Illumina + Nanopore + BioNano	Illumina	Illumina	Illumina + PacBio
Taille estimée	523 Mb	523 Mb	520 Mb	587 Mb	623 Mb	n/a	593 Mb
Scaffolds	7 513	1 532	Na	227*	47 753	56 481	9 467
N50	1.3Mb	3.0 Mb	5.0 Mb	36.8 Mb	22.2 kb	37.7 kb	435.8 kb
L50	65	26	Na	6	6 497	3 764	175
Taille de l'assemblage	472 Mb	451 Mb	492.8 Mb	525 Mb	526 Mb	638 Mb	464 Mb
% de la taille estimée	90%	86%	94.8%	89%	84%	n/a	78%
% de N	17.3%*	10.2 %	0.3%	1.37%	21.9%	52.38%	1.60%
% ancré sur les chromosomes	70%	89.5%	87.3%	94.6%	n/a	n/a	n/a
N gènes	36 542	35 276	35148	32 809	45 069	44 702	32 692

Tableau 3 - Résumé des métriques des différents génomes et assemblages disponibles de Musa

entre marqueurs situés sur des chromosomes différents. L'analyse de ces distorsions a permis de proposer l'hypothèse d'une translocation entre les chromosomes 1 et 3 dans le génome B par rapport au génome A (Noumbissié *et al.* 2016).

3. SEQUENÇAGE DU GENOME DU BANANIER

A ce jour, plusieurs assemblages de génomes de bananiers appartenant à différentes espèces et sous-espèces de *Musa* sont disponibles. Ces séquences ont permis de mettre en place de nouvelles approches pour caractériser les variations structurales qui touchent les bananiers. Les différentes métriques des séquençages décrits sont résumées dans le Tableau 3.

La première séquence de bananier a été réalisée à partir d'un haploïde doublé de la sousespèce M. acuminata ssp. malaccensis (D'Hont et al. 2012). 24 425 contigs ont été obtenus à partir de lectures 454 et Sanger corrigées par des données Illumina. Ces contigs ont été assemblés en 7 513 scaffolds avec un N50 (taille minimum du scaffold au-dessus de laquelle 50% de l'assemblage peut être trouvé) de 1.3Mb et comprenant 17.3% de N. 90% de la taille du génome estimée peut être retrouvée dans les 7 513 scaffolds de l'assemblage. 70% de l'assemblage a pu être ancré sur les 11 chromosomes du bananier. Cet ancrage a été réalisé en utilisant une carte génétique construite à partir de 589 marqueurs microsatellites et 63 marqueurs DArT utilisant une autofécondation d'un apparenté de l'haploïde doublé (l'accession 'PT-BA-00267'). Cette première séquence, qui constitue la séquence de référence de Musa acuminata, a été améliorée par Martin et al. (2016) à l'aide de séquençage Illumina et d'une carte optique BioNano. Ces méthodes ont permis de drastiquement réduire le nombre de scaffolds (de 7 513 à 1 532) tout en augmentant leur N50 (de 1.3 Mb à 3 Mb). Une plus grande proportion de l'assemblage a pu être ancrée le long des 11 chromosomes du bananier (89.5 %). La proportion de N a été réduite de 17.3% à 10.2%.

La séquence du génome de *Musa balbisiana* a été produite à partir de lectures courtes Illumina, de lectures longues PacBio et de données de capture de chromosome (Hi-C) (Wang *et al.* 2019). Cet assemblage a un N50 de 5.0 Mb, et représenterait 94.8% de la

taille estimée du génome avec une très faible proportion de N (0.3%). 87,3% de l'assemblage a pu être ancré sur la carte de contact chromosomique réalisée à partir des données Hi-C.

La séquence du génome de *Musa schizocarpa* a été réalisée à partir de lectures courtes Illumina, lectures longues Nanopore et d'une carte optique BioNano (Belser *et al.* 2018). Cette méthode a permis d'aboutir à un nombre très réduit de scaffolds (227) avec un N50 très élevé (36.8 Mb). Cet assemblage représente 89% de la taille du génome estimée, avec un faible taux de N (1.37%). 94.6% de l'assemblage a pu être ancré sur la carte optique réalisée à partir des données BioNano. Cette méthode réalisée par le Genoscope est actuellement utilisée pour produire une troisième version du génome de référence *M. acuminata* ssp. *malaccensis*.

Enfin, des assemblages de trois accessions des sous-espèces *M. a. burmannicoïdes* ('Calcutta 4'), *zebrina* ('Maia Oa') et *banksii* ('Banksii') ont été réalisés à partir de lectures Illumina, complétées par un séquençage PacBio pour l'accession de la sous-espèces *banksii* (Rouard *et al.* 2018).

Au sein de l'équipe, le projet France Génomique "Dynamo" est en cours. Son objectif est de caractériser les mosaïques en terme d'origines ancestrales dans la diversité sauvage et cultivée des bananiers. Pour cela, 159 accessions représentatives des pôles de la diversité sauvage de *Musa acuminata*, des cultivars diploïdes et triploïdes et d'espèces autres qu'*acuminata* ont été séquencées avec des courtes lectures pairées (Illumina). Ce projet s'accompagne du séquençage et de l'assemblage de génome complet de 'Calcutta 4' (*M. a.* ssp. *burmannicoïdes*), 'Maia Oa' (*M. a.* ssp. *zebrina*), 'Banksii' (*M. a.* ssp. *banksii*) pour améliorer la qualité des assemblages décrits juste avant en utilisant la stratégie de séquençage déployée par le Genoscope pour la production de la séquence de *Musa schizocarpa* (Belser *et al.* 2018). Par ailleurs ce projet est également en train de produire une troisième version du génome de référence 'Pahang HD' (*M. a.* ssp. *malaccensis*) et une première version du génome de *Musa textilis* (une espèce de bananier utilisée pour ses fibres).





'Pahang HD' (M. a. ssp. malaccensis)

'PT-BA-00267' (M. a. ssp. malaccensis)

Figure 10 – Structure chromosomique de 'PT-BA-00267', portant la translocation 1/4.

A gauche, la structure chromosomique du génome de référence *M. a.* ssp. *malaccensis* 'Pahang HD' pour les chromosomes 1 et 4. A droite, la structure chromosomique des chromosomes 1, 1T4, 4 et 4T1 de l'accession hétérozygote de structure 'PT-BA-00267' (*M. a.* ssp. *malaccensis*).

L'amélioration des techniques de séquençage, mais surtout la combinaison de différentes méthodes a été très bénéfique au séquençage du bananier. Si les lectures courtes Illumina permettent un séquençage très haut débit et faible en erreurs, l'utilisation de longues lectures (PacBio ou Nanopore) permet de dépasser certains problèmes d'assemblage en améliorant l'assemblage des éléments répétés par exemple. Enfin, les cartes optiques et le Hi-C permettent d'obtenir des scaffolds de très grande tailles qui facilitent grandement l'étape d'ancrage sur les chromosomes.

4. CARACTERISATION DE VARIATIONS STRUCTURALES ET DE LEUR IMPACT SUR LES SEGREGATIONS A PARTIR DE DONNEES DE SEQUENÇAGE

La disponibilité d'une séquence de référence pour l'espèce *Musa acuminata* (D'Hont *et al.* 2012) a permis de mettre au point de nouvelles approches pour caractériser les grandes variations structurales chez le bananier et leur impact sur les ségrégations.

Martin *et al.* (2017) met au point une méthode basée sur le re-séquençage mate-pair et l'alignement des lectures pairées sur le génome de référence. En analysant les paires de lectures alignées de façon discordante (distance entre les deux lectures plus petite ou plus grande que la taille de l'insert, lectures alignées dans le mauvais sens, ou alignées sur des chromosomes différents) à l'aide d'un algorithme bioinformatique, il caractérise une translocation réciproque entre un fragment distal du chromosome 1 (T1 = 2.9Mb) et une région péricentromérique du chromosome 4 (T4 = 10Mb) à l'état hétérozygote chez l'accession 'PT-BA-00267' (*M. a.* ssp. *malaccensis*) (Figure 10). Cette méthode est appliquée à 14 autres accessions re-séquencées à l'aide d'une approche mate-pair pour rechercher la translocation caractérisée dans une partie de la diversité des bananiers. La translocation 1/4 a ainsi été détectée à l'état homozygote chez 'Malaccensis Nain' et à l'état hétérozygote chez 'Paka', 'IDN 110', 'Akondro Mainty', 'Grande Naine' et 'Pisang Lilin'. La détection de la translocation 1/4 chez 'Pisang Lilin' permet de réinterpréter les données obtenues par Hippolyte *et al.* (2010) à partir du croisement 'Borneo' x 'Pisang Lilin' et de mettre à jour l'hypothèse de la présence d'une duplication en la remplaçant

par la translocation 1/4. Les huit autres accessions testées sont homozygotes pour la structure de référence aux points de cassure de la translocation 1/4.

Finalement, 169 accessions de M. *acuminata* ont été testées par PCR pour la présence des chromosomes 1, 4 et 1T4. Ainsi, sur les 48 accessions diploïdes sauvages testées, deux ('Pa Songkhla' et 'Malaccensis Nain') n'ont amplifié que le chromosome 1T4 et ont donc été caractérisées comme homozygotes pour la structure transloquée. Toutes les autres accessions ont amplifié les fragments des chromosomes 1 et 4, et ont donc été caractérisées comme homozygotes pour la structure de référence. Concernant les accessions triploïdes et tétraploïdes cultivées, 31 ont été testées. La moitié ont été caractérisées comme hétérozygotes pour la structure de référence. Enfin, les neuf accessions *Musa* appartenant à d'autres espèces qu'*acuminata* ont été caractérisées comme homozygotes pour la structure de référence.

Ces résultats ont permis de proposer que la translocation 1/4 correspond au réarrangement chromosomique qui différencie le groupe Northern Malayan du groupe Standard proposé par Shepherd (1999) sur la base d'observations cytogénétique et que la structure portée par le groupe Standard correspond à la structure du génome de référence séquencé (DH-Pahang) et à la structure ancestrale chez *M. acuminata*.

L'analyse des distorsions de ségrégation dans la population issue de l'autofécondation de l'accession structurellement hétérozygote 'PT-BA-00267' met en évidence la transmission préférentielle des chromosomes transloqués dans la descendance. L'étude des recombinaisons le long des chromosomes a permis de montrer l'absence de recombinaison au niveau du fragment T1 du chromosome 1 et un taux de recombinaison très faible au niveau des points de translocation.





A gauche la structure chromosomique du génome de référence diploïde AA *M. a.* ssp. *malaccensis* 'Pahang HD'. A droite, la structure chromosomique de l'accession diploïde BB 'PKW' (*M. balbisiana*). L'inversion sur le chromosome 5 est représentée par une flèche noire.

En 2019, Baurens et al. construit une carte génétique de l'espèce Musa balbisiana à partir de l'autofécondation de l'accession 'Pisang Klutuk Wulung' (BB, Musa balbisiana). Les marqueurs SNPs issus du génotypage DArT-seq de la population ont été alignés le long de la séquence de référence de M. acuminata afin de comparer les génomes A et B. Cette analyse a permis de détecter deux réarrangements chromosomiques propres au génome B : une translocation réciproque entre le fragment distal de 0.65Mb du chromosome 1 et un fragment de 8.5 Mb du chromosome 3, ainsi qu'une inversion d'un segment de 9Mb au sein du chromosome 5 (Figure 11). La translocation 1/3 génère de fortes distorsions de ségrégation dans les hybrides A/B. L'analyse des ratios en allèles d'origine parentale A ou B par chromosome painting et de la couverture en marqueurs des 185 individus issus du croisement CRBP39 (AAAB) x Pahang (AA) met en évidence l'absence de recombinaison le long du segment inversé du chromosome 5 et une baisse de recombinaison dans les régions de cassure de la translocation entre les chromosomes 1 et 3. De plus, 61% des individus sont aneuploïdes dans cette population, c'est à dire qu'ils ne possèdent pas exactement trois chromosomes homéologues complets pour les 11 sets de chromosomes. Cette aneuploïdie touche majoritairement les chromosomes concernés par un réarrangement chromosomique (1, 3 ou 5). Enfin, on constate que dans les individus euploïdes (triploïdes) de cette descendance, la structure transloquée des chromosomes 1 et 3 est préférentiellement transmise, avec 57% des individus de génotype AAB et 43% AAA.

En conclusion, les deux translocations réciproques caractérisées jusqu'alors chez le bananier concernent le chromosome acrocentrique 1. Le fragment distal transloqué est cependant de taille différente (3 Mb pour la translocation 1/4 et 0.65 Mb pour la translocation 1/3). Ces deux translocations génèrent de fortes distorsions de ségrégation dans une descendance issue de parents structurellement hétérozygotes et sont préférentiellement transmises à la descendance euploïde. Des taux de recombinaisons plus réduits sont observés aux bornes des réarrangements. Contrairement à la translocation 1/3, la translocation 1/4 provoque une suppression totale des recombinaisons le long du segment T1. La translocation entre les chromosomes 1 et 4 ne semble pas produire d'aneuploïdie dans le contexte diploïde étudié. La translocation 1/3,

produit de nombreux aneuploïdes (aneuploïdie totale ou segmentale) dans le contexte hétérozygote tétraploïde étudié.

D. Présentation du sujet de thèse

Les bananiers cultivés sont issus d'hybridations interspécifiques et intersubspécifiques impliquant essentiellement les sous-espèces de M. *acuminata* (génome A, 2n=2x=22) et pour certains l'espèce M. *balbisiana* (génome B, 2n=2x=22).

Sur la base de l'analyse d'appariements chromosomiques chez des hybrides inter(sub)spécifiques de *Musa* la présence de variations structurales a été suspectée et Shepherd (1999) a proposé l'existence de 6 groupes caryotypiques au sein de *Musa acuminata*.

De plus, les études de cartographie génétique ont révélé d'importantes distorsions de ségrégations, affectant parfois des groupes de liaison entiers. Leur origine a souvent été attribuée à la présence de larges réarrangements chromosomiques. Deux translocations réciproques (entre les chromosomes 1 et 4 dans un sous-groupe de la sous-espèce *M. a. malaccensis* d'une part et entre les chromosomes 1 et 3 dans l'espèce *Musa balbisiana* d'autre part) ont été récemment caractérisées et leur impact sur les ségrégations étudiées.

En générant des distorsions de ségrégation ou en impactant la recombinaison, les variations structurales peuvent être un frein à l'analyse du déterminisme génétique des caractères agronomiques d'intérêt et au bon déroulement des stratégies de brassage allélique dans les programmes d'amélioration. Mieux appréhender les différences de structures au sein des *Musa acuminata*, leur distribution dans la diversité génétique sauvage et cultivée, et leur impact sur les ségrégations chromosomiques est crucial pour accompagner la recherche en génétique quantitative sur le bananier et la création variétale.

Dans ce contexte, l'objectif de ma thèse a été de continuer à caractériser les variations structurales au sein de la diversité sauvage de *Musa acuminata*, d'en appréhender la

distribution dans la diversité sauvage et cultivée et enfin d'en cerner l'impact sur les appariements et sur les ségrégations chromosomiques.

Pour cela je me suis attachée à caractériser les variations structurales suspectées au sein de la sous-espèce *M. a.* ssp. *burmannicoïdes*. Je me suis basée en particulier sur l'accession 'Calcutta 4', supposée porter deux variations structurales d'après les travaux de Shepherd (1999). Pour caractériser ces variations structurales j'ai exploité des données de cartographie génétique impliquant cette accession, des données de séquençage de grand insert, des outils bioinformatiques mis au point par Martin *et al.* (2016) et des données d'assemblage préliminaire du génome de cette accession.

Une fois ces variations structurales précisément caractérisées, j'ai développé une approche bioinformatique exploitant des données de séquençage pairé pour étudier leur distribution dans la diversité *Musa*. Cette approche a permis de détecter les signatures des translocations dans 123 accessions de la diversité sauvage et cultivé.

Ce travail de caractérisation et d'analyse de la distribution dans la diversité fait l'objet du chapitre 1 de la thèse rédigé sous forme d'un article publié dans la revue Annals of Botany.

J'ai ensuite analysé l'impact de ces translocations sur les ségrégations chromosomiques dans un contexte diploïde hétérozygote de structure en utilisant les données de génotypage par séquençage (GBS) de la population "BCM" issue du croisement ('Banksii' x 'Calcutta 4') x 'Malaccensis Nain'. Ce travail fait l'objet du chapitre 2 de la thèse.

CHAPITRE 1

La détection, caractérisation puis analyse de la distribution de deux translocations réciproques dans la diversité sauvage et cultivée s'est déroulée en trois étapes :

Dans un premier temps, une population de 67 individus issue de l'autofécondation de l'accession 'Calcutta 4' (*Musa acuminata* ssp. *burmannicoïdes*) a été génotypée par séquençage (GBS). Ces données GBS ont permis de construire une carte de liaison à partir des marqueurs SNP ségrégeant dans la population. Ces marqueurs SNP ont été ordonnés le long du génome de référence de *Musa acuminata*. Ceci a mis en évidence quatre ruptures de liaison génétique entre des marqueurs au sein des chromosomes 1, 2, 8 et 9 ainsi que deux groupes de marqueurs présentant de très fortes liaisons génétiques entre des régions des chromosomes 2 et 8 d'une part et 1 et 9 d'autre part. A partir de ces analyses, j'ai donc conclu à la présence de deux translocations réciproques présentes à l'état homozygote chez l'accession 'Calcutta 4' impliquant les chromosomes 2 et 8 d'une part et 1 et 9 d'autre part.

J'ai ensuite complété la caractérisation de ces translocations en identifiant précisément les points de translocation et les fragments chromosomiques impliqués. Pour cela, les données de séquençage mate-pair de 'Calcutta 4' ont été alignées sur le génome de référence de *Musa acuminata* ('Pahang HD') puis fouillées à la recherche de lectures discordantes, c'est-à-dire des lectures présentant des configurations d'alignement non attendues (lectures 1 et 2 inversées, trop proches ou trop éloignées par rapport à la taille d'insert attendue ou alignées sur différents chromosomes). L'outil Scaffremodler (Martin *et al.* 2016) a été utilisé et a permis de détecter et interpréter automatiquement la translocation réciproque entre un fragment distal de 240 kb du chromosome 2 et un fragment distal de 7.16 Mb du chromosome 8 chez 'Calcutta 4'. Dans le cas de la translocation entre les chromosomes 1 et 9, l'exploration des clusters de lectures discordantes a dû être faite manuellement. J'ai ainsi pu découvrir une configuration complexe comportant des clusters de lectures discordantes intrachromosomiques et interchromosomiques. L'interprétation que nous proposons implique plusieurs évènements de délétion et de duplication aux bornes des cassures de la translocation

réciproque entre un fragment distal de 8,24 Mb du chromosome 1 et un fragment distal de 11.56 Mb du chromosome 9. Les bornes précises des translocations ont ensuite été déterminées en comparant un assemblage préliminaire de l'accession 'Calcutta 4' avec le génome de référence de *Musa acuminata* ('Pahang HD').

Pour analyser la distribution des deux translocations dans la diversité sauvage et cultivée des bananiers, j'ai recherché la présence de signatures de ces translocations et de la structure chromosomique de référence, sous la forme de jonctions de segments chromosomiques (Signature Segment Junction ou SSJ dans l'article) dans les données de séquençage Illumina pairé en petit insert (150 pb) disponible pour 123 accessions issues du projet Dynamo.

Concernant les chromosomes 2 et 8, j'ai ainsi pu détecter la translocation 2/8 sans détecter la structure de référence dans toutes les accessions appartenant au pôle génétique *burmannica/burmannicoïdes/siamea* ('Calcutta 4', 'Pisang Prentel', 'Long Tavoy', 'Khae Phrae' et 'Pa Rayong'). Ces accessions ont été considérées comme homozygotes pour la translocation 2/8. Deux accessions ont montré des signaux pour les structures de référence et transloquées : 'Manang' et 'Hom', et ont été considérées comme hétérozygotes pour la translocation 2/8. Les résultats obtenus sur toutes les autres accessions ont été interprétés comme résultant d'un état homozygote pour la structure de référence aux bornes de la translocation 2/8, à l'exception de 'Pisang Serun 400' et 'Sinwobogi' dont la structure reste indéterminée.

Concernant la translocation 1/9, les SSJ ont été retrouvés chez les accessions appartenant au pôle génétique *burmannica/burmannicoïdes/siamea* ('Calcutta 4', 'Pisang Prentel', 'Long Tavoy', 'Khae Phrae' et 'Pa Rayong'). Cependant, la configuration complexe des régions dans lesquelles ont lieu les cassures a rendu impossible la détection de SSJ propre à la structure de référence dans ces régions. Il n'a donc pas été possible à partir de ces données de statuer quant à la présence de la structure de référence aux bornes de la translocation 1/9.

L'article qui suit, intitulé "Two large reciprocal translocations characterized in the disease resistance-rich *burmannica* genetic group of *Musa acuminata*" décrit en détail les méthodes utilisées pour détecter et analyser la distribution de ces deux translocations

dans la diversité, ainsi que les résultats qui en découlent. Il a été soumis en février 2019 dans le journal Annals of Botany et accepté en juin 2019. Les figures et tableaux supplémentaires sont consultables en annexe 1 à 8.



Two large reciprocal translocations characterized in the disease resistance-rich burmannica genetic group of Musa acuminata

Marion Dupouy^{1,2}, Franc-Christophe Baurens^{1,2}, Paco Derouault^{1,2}, Catherine Hervouet^{1,2}, Céline Cardi^{1,2}, Corinne Cruaud³, Benjamin Istace³, Karine Labadie³, Chantal Guiougou⁴, Lyonel Toubi⁵, Frederic Salmon⁴, Pierre Mournet^{1,2}, Mathieu Rouard⁶, Nabila Yahiaoui^{1,2}, Arnaud Lemainque³, Guillaume Martin^{1,2,†} and Angélique D'Hont^{1,2,†,*}

¹CIRAD, UMR AGAP, F-34398 Montpellier, France, ²AGAP, Université Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France, ³Genoscope, Institut de biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, Evry, France, ⁴CIRAD, UMR AGAP, F-97130 Capesterre Belle Eau, Guadeloupe, France, ⁵CIRAD, UMR AGAP, F-97170 Petit-Bourg, Guadeloupe, France and ⁶Bioversity International, Montpellier, France

†These authors contributed equally to this work.

* For correspondence. E-mail dhont@cirad.fr

Received: 18 February 2019 Returned for revision: 9 April 2019 Editorial decision: 1 May 2019 Accepted: 9 June 2019

• Background and Aims Banana cultivars are derived from hybridizations involving *Musa acuminata* subspecies. The latter diverged following geographical isolation in distinct South-east Asian continental regions and islands. Observation of chromosome pairing irregularities in meiosis of hybrids between these subspecies suggested the presence of large chromosomal structural variations. The aim of this study was to characterize such rearrangements.

• **Methods** Marker (single nucleotide polymorphism) segregation in a self-progeny of the 'Calcutta 4' accession and mate-pair sequencing were used to search for chromosomal rearrangements in comparison with the *M. acuminata* ssp. *malaccensis* genome reference sequence. Signature segment junctions of the revealed chromosome structures were identified and searched in whole-genome sequencing data from 123 wild and cultivated *Musa* accessions.

• **Key Results** Two large reciprocal translocations were characterized in the seedy banana *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' accession. One consisted of an exchange of a 240 kb distal region of chromosome 2 with a 7.2 Mb distal region of chromosome 8. The other involved an exchange of a 20.8 Mb distal region of chromosome 1 with a 11.6 Mb distal region of chromosome 9. Both translocations were found only in wild accessions belonging to the *burmannicoides/burmannicalsiamea* subspecies. Only two of the 87 cultivars analysed displayed the 2/8 translocation, while none displayed the 1/9 translocation.

• Conclusion Two large reciprocal translocations were identified that probably originated in the *burmannica* genetic group. Accurate characterization of these translocations should enhance the use of this disease resistance-rich *burmannica* group in breeding programmes.

Key words: *Musa acuminata*, banana, chromosomal rearrangement, reciprocal translocation, paired-end sequencing, genotyping by sequencing.

INTRODUCTION

Bananas (*Musa* genus) provide staple food for millions of people in Asia, Africa, Latin America and Oceania, and represent a major cash crop around the world. Most current cultivars are derived from *Musa acuminata* (A genome, 2n = 2x = 22), sometimes combined with *Musa balbisiana* (B genome, 2n = 2x = 22). *Musa acuminata* diverged following the Pliocene geographical isolation in distinct South-east Asian continental regions and islands (Daniells et al., 2001; Perrier et al., 2009; Janssens et al., 2016; Rouard et al., 2018). It has been divided into 6–9 partially interfertile subspecies (*M. a.* ssp. *burmannica, burmannicoides, siamea, malaccensis, truncata, errans, microcarpa, zebrina* and *banksii*) based on morphology and geographic distribution (Simmonds, 1962; Perrier et al., 2009). The current domestication scenario suggests that plants were transported during human migrations, leading to hybridization between *Musa* species and subspecies (Perrier *et al.*, 2011). This resulted in the emergence of inter(sub)specific hybrids with reduced fertility (Dodds and Simmonds, 1948; Shepherd, 1999). Early farmers would then have selected parthenocarpic diploid and triploid hybrids producing edible fruits with high flesh and low seed content. These hybrids were then propagated clonally. The hundreds of current cultivars are thus products of centuries – in some cases millennia – of vegetative propagation. Only a few of them are cultivated at a large scale and one clonal cultivar group, i.e. Cavendish, accounts for >50 % of world banana production. These large-scale monocultures have favoured the emergence of diseases and extensive pesticide use (Lescot *et al.*, 2008).

Observations of chromosomal pairing during meiosis showed irregularities in the M. acuminata and M. balbisiana interspecific hybrid and in M. acuminata intersubspecific hybrids, with the presence of univalents and multivalents (Shepherd, 1999; Jeridi et al., 2011). Within subspecies, meioses were generally found to be regular with bivalents. In addition, genetic mapping studies involving M. acuminata hybrids revealed substantial segregation distortion involving a few linkage groups (Fauré et al., 1993; Hippolyte et al., 2010; Mbanjo et al., 2012; Noumbissié et al., 2016). The presence of large chromosomal structural variations between species and subspecies was suggested to be responsible for these irregularities. Based on the study of pairing configurations in a limited number of M. acuminata intersubspecific hybrids, Shepherd (1999) proposed that six translocation groups exist, differing from a Standard' (ST) group by 1-3 translocations. These groups are only partly concordant with subspecies delimitation, with the 'Standard' group being the largest one and comprised banksii, microcarpa and malaccensis ssp. accessions.

The availability of a M. acuminata reference genome sequence (D'Hont et al., 2012; Martin et al., 2016) allowed us to implement next-generation sequencing (NGS) technologies and comparative genomics to investigate large structural variations within Musa. A first large reciprocal translocation between chromosomes 1 and 4 was identified as compared with the M. acuminata reference genome sequence obtained from the 'DH-Pahang' M. acuminata ssp. malaccensis accession. Among wild accessions, this 1/4 translocation was found only in some of the ssp. malaccensis accessions, suggesting that it emerged within this subspecies (Martin et al., 2017). The 1/4 translocation was suggested to correspond to the 'Northern Malayan' karyotypic group proposed by Shepherd (1999), and the M. acuminata reference genome sequence was suggested to correspond to the ST group. Strong segregation distortion and reduced recombination were observed around translocation breakpoints in progeny from structurally heterozygous accessions. Interestingly, the 'Northern Malayan' structure was found to be preferentially transmitted to the progeny (Martin et al., 2017).

Another large reciprocal translocation between chromosomes 1 and 3 and a large inversion on chromosome 5 were characterized in *M. balbisiana* as compared with the *M. acuminata* reference genome sequence (Baurens *et al.*, 2019). A high proportion of aneuploids was observed, involving chromosomes 1, 3 and 5, in the progeny of a tetraploid accession bearing this *M. balbisiana* chromosome structure in one copy and three copies of the standard *M. acuminata* structure. Moreover, high segregation distortion and reduced recombination were observed for these chromosomes (Baurens *et al.*, 2019).

These studies showed that large structural variations cause segregation distortion and recombination reduction or suppression in *Musa* structural hybrids, as observed in several other plants such as *Lens*, *Helianthus* and peach (Tadmor *et al.*, 1987; Quillet *et al.*, 1995; Jáuregui *et al.*, 2001). Characterizing such rearrangements is thus important to facilitate the use of genetic resources in breeding programmes, in particular to improve disease resistance.

In this study, we analysed the segregation of a self-progeny of a wild *M. acuminata* ssp. *burmannicoides* accession and detected two large reciprocal translocations compared with the *M. acuminata* ssp. *malaccensis* reference sequence. We then accurately characterized breakpoints of these structural variations through discordant mate-pair read analysis. We developed an *in silico* approach based on signature segment junctions to detect translocation breakpoints in paired-end sequences from 123 accessions representative of *M. acuminata* germplasm.

MATERIALS AND METHODS

Plant and sequencing material

A self-progeny of 76 individuals from the 'Calcutta 4' accession (PT-BA-00051, CRB Plantes Tropicale, Guadeloupe) was produced at the CIRAD research station in Guadeloupe. DNA was extracted, and sequencing libraries were constructed using a genotyping by sequencing (GBS) approach and sequenced on the GeT-PlaGe platform (https://get.genotoul.fr/) using the Illumina Hiseq3000 sequencer.

Other banana plant material was obtained either from CRB Plantes Tropicales (Guadeloupe, French West Indies, http:// crb-tropicaux.com/Portail) or from the Bioversity International *Musa* Germplasm Transit Centre (Ruas *et al.*, 2017) (Supplementary data Table S1).

A total of 123 accessions were sequenced to obtain standard Illumina 2×150 paired-end reads using the HiSeq4000 platform at Genoscope (http://www.genoscope.cns.fr).

For 11 of these accessions, 5 kb insert mate-pair libraries were constructed and sequenced using the Illumina Hiseq2500 platform at Genoscope (http://www.genoscope.cns.fr). This data set was supplemented by Illumina 5 kb mate-pair sequencing data generated by Rouard *et al.* (2018) for three accessions ('Banksii', 'Maia'Oa' and 'Calcutta 4') and deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under run IDs SRR7013759, SRR7013765 and SRR7013755. The assembly of the *M. acuminata* ssp. *banksii* accession 'Banksii' published by these authors, as well as PacBio reads related to this project (SRA ID SRR7013757), were also used.

Finally, 'Calcutta 4' and 'Maia'Oa' accessions were sequenced with a MinION nanopore sequencer from Oxford Nanopore Technologies (Jain *et al.*, 2016). As described by Belser *et al.* (2018), a 30× sub-set of the longest nanopore reads was extracted from the total read set. These reads were then assembled using the Smartdenovo assembler (Jue Ruan, https://github.com/ruanjue/smartdenovo, git-commit 3d9c22e) with -k 17, as advised by the developer for larger genomes, and -c 1 to generate a consensus sequence.

SNP marker segregation analysis

Raw GBS sequencing data from the self-progeny were demultiplexed using GBSX version 1.2 (Herten *et al.*, 2015) and adaptors were removed with cutadapt (Martin, 2011). Variant calling was processed as described by Garsmeur *et al.* (2018), using the VcfHunter package (https://github.com/ SouthGreenPlatform/VcfHunter). Mapping was performed on the *M. acuminata* reference sequence version 2 (Martin *et al.*, 2016). Sites with read depths of <10 and >1000 were excluded. Sites with a minor allele frequency of <5 % were considered homozygous. Sites with an alternate allele frequency of >15 % supported by at least three reads were considered heterozygous. Other sites were converted into missing data. single nucleotide polymorphism (SNP) markers were then filtered out based on their segregation ratio in the mapping population: a *P*-value threshold of 4e-06 was applied with a χ^2 test to compare the observed marker segregation with the expected marker segregation (1:2:1 with a self-population). SNP markers with >2.5 % missing data and individuals with >1% missing data were removed from the SNP data set. Finally, 67 individuals were kept and 5281 SNP markers were selected to build the genetic map.

Five kilobase discordant mate-pair read analysis

To refine the location of the rearrangement breakpoints and to characterize the chromosome structures of 14 accessions, large insert mate-pair reads from the 14 mate-pair sequenced accessions were aligned to the *M. acuminata* 'DH-Pahang' reference genome sequence (version 2) available on the Banana Genome Hub (Droc *et al.*, 2013). They were analysed to identify discordant read clusters according to the methodology described in Martin *et al.* (2017), using Scaffremodler tools (Martin *et al.*, 2016) available on the South Green platform (Southgreen Collaborators, 2016). Discordant read clusters were viewed using CIRCOS software (Krzywinski *et al.*, 2009).

Alignment of translocation breakpoints between 'DH-Pahang' and 'Calcutta 4'

To precisely compare the rearrangement breakpoints between 'DH-Pahang' and 'Calcutta, the four genomic regions of the 'DH-Pahang' reference genome (version 2) corresponding to translocation breakpoint regions (chromosome 1 between 8.23 and 8.27 Mb, chromosome 2 between 29.26 and 29.28 Mb, chromosome 8 between 37.715 and 37.735 Mb and chromosome 9 between 11.57 and 12.61 Mb) were aligned to the 'Calcutta 4' preliminary assembly using blastn (BLAST+ 2.6.0; Camacho *et al.*, 2009) with an e-value threshold of 1e-20 and without masking low-complexity regions. Then comparisons between sequences were performed using Gepard (Krumsiek *et al.*, 2007) (Supplementary data Table S2).

Signature segment junction detection

In order to characterize the chromosome structure of the 123 studied *Musa* accessions, paired reads from 123 accessions were aligned to the complete *M. acuminata* 'DH-Pahang' reference genome, the 'Calcutta 4' assembly, the 'Maia'Oa' assembly and the 'Banksii' corrected assembly using bowtie2 in very sensitive mode (see next section). Pairs properly mapped with reads on either side of a signature segment junction (SSJ; i.e. the junction between two genomic segments that is specific to one of the chromosome structures; Supplementary data Fig. S1) in a flanking window of 1000 bp around the SSJ were counted.

Local reassembly of the chromosome 8 translocation breakpoint in 'Maia'Oa' and 'Banksii' accessions

To precisely characterize the positions of the SSJs relative to the breakpoint region of chromosome 8 in two accessions representative of zebrina and banksii subspecies, long reads from 'Maia Oa' and 'Banksii' were compared with the M. acuminata 'DH-Pahang' reference chromosome 8 breakpoint region. Raw long reads from 'Maia'Oa' and 'Banksii' accessions were corrected using the correct option of Canu 1.6 (Koren et al., 2017) with default parameters. Corrected reads were aligned to the M. acuminata 'DH-Pahang' reference genome version 2 using Graphmap (Sović et al., 2016) with default parameters. For each accession, the longest read overlapping the two SSJs of M. acuminata 'DH-Pahang' reference chromosome 8 was kept. Selected long reads were corrected by aligning Illumina short reads from their respective accessions with bwa mem (Li and Durbin, 2009) with default parameters. Consensus between short reads was calculated with bcftools (Li, 2011).

The resulting consensus sequences were compared with the 37.718-37.732 Mb region of M. acuminata 'DH-Pahang' reference chromosome 8 using Gepard. Then we compared the 'Maia'Oa' consensus sequences of the chromosome 8 breakpoint with the 'Maia'Oa' scaffolds and the 'Banksii' consensus sequences of the chromosome 8 breakpoint with the 'Banksii' scaffolds using blastn, with an e-value threshold of 1e-20, while not masking low-complexity regions to confirm that the region has been correctly assembled. An inconsistency between the 'Banksii' corrected long read and its assembly was noticed, probably due to a short read assembly error. The genomic region in the original assembly was thus replaced by the new consensus sequence. Finally, the corresponding regions of the 'Banksii' and 'Maia'Oa' assembly were aligned and compared with the 1000 bp flanking window around the two M. acuminata 'DH-Pahang' reference chromosome 8 SSJs.

Factorial analysis using SNP data

To investigate the distribution of the translocated chromosomes in M. acuminata, a factorial analysis was performed. Paired-end reads (2×150) Illumina sequencing data from the 83 M. acuminata diploid accession studied were used to produce a genotype call for each accession using vcfHunter tools (Garsmeur et al., 2018; https://github.com/SouthGreenPlatform/ VcfHunter). The process_reseq_1.0.py and VcfPreFilter.1.0.py programs were used to (1) align Illumina reads along the M. acuminata 'DH-Pahang' reference genome sequence (version 2) (Martin et al., 2016) using Burrows-Wheeler Alignment version 0.7.15; (2) remove duplicates using the MarkDuplicates option of Picard Tools (version 2.7) (http://broadinstitute. github.io/picard/); (3) locally realign reads around indels using the IndelRealigner tool of the GATK software package (version 3.6) (McKenna et al., 2010); (4) count all mapping bases that had a mapping quality ≥ 10 with the bam-readcount program (https://github.com/genome/bam-readcount); and (5) call a genotype based on the maximum likelihood of the genotype according to a binomial distribution assuming a sequencing error rate of 0.005. The variant calling file was formatted in VCF format. The VCF file was then filtered using the vcfFilter. 1.0.pv program with the following options: a minimal coverage by

accession of 10 (-MinCov 10), a maximal coverage by accession of 119 (-MaxCov 119), a minimal relative allele frequency by accession of 0.1 (-MinFreq 0.1), a minimal absolute allele frequency by accession of 3 (-MinAl 3), no missing genotypes in a line allowed (-nMiss 0), filtering out indels (-RmType INDELS) and non-diallelic sites (-RmAlAlt 1:3:4:5). A sub-set of 3092 polymorphous SNPs dispersed along the 11 chromosomes of the reference sequence was used to calculate a dissimilarity matrix using vcf2dis.1.0.py script. The dissimilarity index between two accessions was calculated from the VCF file as the proportion of unmatching alleles between two accessions. The dissimilarity matrix was used to perform a factorial analysis using R version 3.2.4 software (http://www.r-project. org). Since the cultivated accessions originated from the wild banana genepool, the factorial analysis was performed with the 32 wild accessions. The 51 cultivated accessions were then projected along the synthetic axes.

RESULTS

Genetic evidence of two large reciprocal translocations in the 'Calcutta 4' accession

A self-progeny from the 'Calcutta 4' accession was genotyped by sequencing. A total of 5281 markers segregating in 67 individuals with a 1:2:1 ratio were selected. Pairwise associations between 'Calcutta 4' markers projected on the 11 M. acuminata reference chromosome sequence showed, for most markers physically close on the reference sequence, strong genetic linkage, as expected (Supplementary data Fig. S2A, B). However, a complete break in genetic linkage was observed within a few chromosomes. Linkage breaks were observed on chromosome 2 at position 29.2 Mb and on chromosome 8 at position 37.7 Mb (Fig. 1A). In addition, a small cluster of markers (four markers) from the distal region of reference chromosome 2 was found in 'Calcutta 4' genetically linked with chromosome 8 markers, whereas a distal part of reference chromosome 8 was found to be linked with chromosome 2 in 'Calcutta 4'.

A similar situation was observed for chromosomes 1 and 9. Linkage breaks were observed on chromosomes 1 and 9 at 8.2 and 11.4 Mb, respectively (Fig. 1B). Markers



FIG. 1. Genetic linkage between SNP markers in the 'Calcutta 4' self-progeny projected on *M. acuminata* reference chromosomes 2 and 8 (A) and 1 and 9 (B). Each dot represents linkage between two markers. Linkage intensity is represented by a warm-cool colour gradient from dark red (strong) to blue (weak).

surrounding these breaks linked, in 'Calcutta 4', a distal region of reference chromosome 1 to reference chromosome 9 and the rest of chromosome 1 to a distal part of reference chromosome 9.

These observations suggested the presence of distinct chromosome structures in 'Calcutta 4' as compared with the *M. acuminata* reference sequence. These structures resulted from two reciprocal translocations, one involving chromosomes 1 and 9 and the other involving chromosomes 2 and 8. The observed linkage breaks were complete, indicating that 'Calcutta 4' is structurally homozygous for these translocations.

Characterization of the translocation breakpoints

Translocation breakpoints were further characterized using 5 kb mate-pair reads from 'Calcutta 4' aligned to the 'DH-Pahang' reference sequence to search for discordant read alignments.

For chromosomes 2 and 8, two clusters of discordant reads were detected in the region of chromosomes 2 and 8 and found to be genetically linked in the self-progeny of 'Calcutta 4' (Fig. 2A). Using Scaffremodler, these discordant read clusters were identified as resulting from the presence of a reciprocal translocation in 'Calcutta 4' as compared with the reference sequence (Fig. 2B). Cluster 2 linked a distal region of 240 kb (c) of chromosome 2 to 45 Mb of chromosome 8 (d), resulting in a distinct chromosome structure hereafter referred to as 8T2 (Fig. 2B). Cluster 1 linked a distal region of 7.2 Mb from chromosome 8 (f) to 29.3 Mb of chromosome 2 (a), resulting in a distinct chromosome structure hereafter referred to as 2T8 (Fig. 2B). On reference chromosomes 2 and 8, the two discordant read clusters are separated by small chromosome segments of 3 kb (e) and 1.2 kb (b), respectively, that were not covered by any read sequences in 'Calcutta 4', thus suggesting their absence in this accession (Fig. 2A). No concordant reads overlapped the breakpoint regions, thus indicating that 'Calcutta 4' is homozygous for this translocation.

For chromosomes 1 and 9, the Scaffremodler program could not automatically interpret the discordant read cluster pattern. Therefore, we manually explored discordant read clusters linking those chromosomes and found a complex pattern of interchromosomal and intrachromosomal discordant read clusters (Fig. 2C). Discordant read cluster 3 and concordant reads along the k2 segment indicated the deletion of this segment in one chromosome. Discordant read cluster 1 linked the $k_1 - k_3$ segment to the h segment, resulting in k_1-k_3-h and $k_1-k_2-k_3-l$ chromosome fragments. Finally, cluster 2 linking the h fragment to the k, fragment reduced possible combinations to reconstruct the chromosome regions, without missing any segment, to a g-h-k₁-k₂-k₃-l structure for one chromosome and a j-k₁-k₃-h-i structure for the other. These chromosome structures corresponded to a reciprocal translocation that occurred in two distinct chromosomes. The first one, hereafter referred to as 1T9, linked a 8.24 Mb distal region of chromosome 1 (g) with 29.7 Mb of chromosome 9 (1). The second one, hereafter referred to as 9T1, linked an 11.6 Mb distal region of chromosome 9 (j), with 20.8 Mb of chromosome 1 (i) (Fig. 2D).



FIG. 2. Paired-read mapping evidence for two reciprocal translocations in 'Calcutta 4' and schematic chromosomal diagram representations. (A and C) Circos representation of significant discordant read clusters from 'Calcutta 4' compared with the *M. acuminata* reference sequence assembly, with a focus on paired read clusters detected in the targeted regions of chromosomes 2 and 8, and chromosomes 1 and 9. Grey internal lines correspond to 'Calcutta 4' paired reads with correct orientation and insert size, while red and blue lines correspond to discordant pairs with higher insert size and reverse mapping orientation, respectively. Note: cluster 3 on segment c indicates a small deletion in 'Calcutta 4' outside the translocation event. (B and D) Schematic chromosome structures for 'Calcutta 4' chromosomes based on the paired-read mapping interpretation. Black segments are accession specific. Signature segment junctions (SSJs) are represented with diamonds.

Refining the translocation breakpoints in 'Calcutta 4'

A 'Calcutta 4' preliminary assembly was compared with the 'DH-Pahang' *M. acuminata* reference sequence to identify 'Calcutta 4' scaffolds overlapping the translocation breakpoints. Two large scaffolds were identified for the 2/8 translocation. One scaffold (2.54 Mb long) contained the 2T8 breakpoint with a segment that aligned to reference chromosome 2 (a) and a segment that aligned to reference chromosome 8 (f) separated by a 600 bp segment (x) with no similarity to the corresponding region in reference chromosomes 2 (b) or 8 (e) (Fig. 3A). The second scaffold (3.15 Mb long) contained the 8T2 breakpoint, with a segment that aligned to reference chromosome 8 (d) and another that aligned to reference chromosome 2 (c) separated by a 1.2 kb segment (y), with no similarity to the corresponding region in reference chromosome 2 (b) or 8 (e) (Fig. 3A).

For the 1/9 translocation, two scaffolds were also identified. One scaffold (6.9 Mb long) contained the 1T9 breakpoint, with segments that aligned to reference chromosome 1 (g and h) and segments that aligned to reference chromosome 9 (k_1 , k_2 , k_3 and 1) (Fig. 3B). The second scaffold (5.7 Mb long) contained the 9T1 breakpoint, with segments that aligned to reference chromosome 9 (j, k_1 and k_3) and segments that aligned to chromosome 1 (h and i).

These data validated the proposed chromosome structures based on discordant read analysis and revealed the presence, at the breakpoints, of chromosomes 2T8 and 8T2 of segments (x and y) not present in the reference sequence.

Distribution of the two translocations in Musa germplasm

Five kilobase mate-pair discordant read analysis. To analyse the distribution of the two translocations in *Musa* germplasm, we first used 5 kb mate-pair reads available from 13 *M. acuminata* accessions (five wild seedy accessions and eight hybrid cultivated accessions; Supplementary data Table S1). Paired reads



FIG. 5. Dot-plot representation of alignments between 'Calcutta 4' scatfolds and the *M. acuminata* reference chromosome sequence at translocation breakpoints. (A) Translocation 2/8: 'Calcutta 4' for breakpoint regions of chromosomes 2T8 (scaffold utg154) and 8T2 (scaffold utg170) are represented on the horizontal axis. The corresponding regions in the *M. acuminata* reference genome are represented on the vertical axis. Each compared region is 6 kb long. A small deletion in segment c of chromosome 2 can be noticed, which corresponds to part of the 5 kb deletion observed in Fig. 2A. (B) Translocation 1/9: 'Calcutta 4' for breakpoint regions of chromosomes 1T9 (scaffold utg195) and 9T1 (scaffold utg94) are represented on the horizontal axis. The corresponding regions in the *M. acuminata* reference genome are represented on the vertical axis. Each compared region is 70 kb long. The colour coding is the same as in Fig. 2.

were aligned to the 'DH-Pahang' reference sequence, and discordant clusters identified in 'Calcutta 4' as corresponding to the 2T8, 8T2, 1T9 and 9T1 chromosome structures were searched (Supplementary data Fig. S3A, B).

Among the 13 *M. acuminata* accessions, 2T8 and 8T2 chromosome structures were detected in one accession, i.e. the 'Manang' diploid cultivar. In this accession, concordant mate-pair reads spanning the breakpoints were also observed, revealing that this accession is structurally heterozygous, with one set of reference chromosomes 2 and 8 and one set of chromosome structures 2T8 and 8T2. All other accessions were found to be homozygous with two sets of reference chromosomes 2 and 8 (Supplementary data Fig. S3A).

No discordant clusters corresponding to chromosomes 1T9 and 9T1 were observed in the 13 accessions surveyed, and all accessions showed concordant mate-pair reads corresponding to reference chromosomes 1 and 9. These accessions were thus homozygous for reference chromosomes 1 and 9 (Supplementary data Fig. S3B).

Signature segment junctions searched in whole-genome sequencing (WGS) data. To further characterize the distribution of 2T8, 8T2, 1T9 and 9T1 chromosome structures in *Musa* diversity, we used the Illumina paired-end sequence generated on 123 accessions. These accessions included 97 *M. acuminata* accessions with 32 diploid wild seedy accessions and 65 cultivated intersubspecific hybrids (51 diploids and 14 triploids), four diploid accessions from other species (two *M. balbisiana* and two *M. schizocarpa*) and 22 interspecific cultivated hybrids involving *M. acuminata* (four diploids and 18 triploids) (Supplementary data Table S1). In the distinct chromosomal structures, distinct chromosomal segments were juxtaposed, so a junction between chromosome segments could be specific to one chromosomal structure and constitute a signature of this structure (hereafter referred to as an SSJ). Such SSJs were searched at the 2/8 and 1/9 translocation breakpoints in 'Calcutta 4' and in the corresponding region of the 'DH-Pahang' reference genome sequence.

Eight SSJs could be identified for translocation 2/8, with two per chromosome structure: a–b and b–c segment junctions for chromosome 2, d–e and e–f segment junctions for chromosome 8, a–x and x–f segment junctions for chromosome 2T8, and d–y and y–c segment junctions for chromosome 8T2 (Fig. 2; Supplementary data Table S3).

In a first round of analysis, for many accessions we detected SSJs of reference chromosome 2 but did not detect the signatures of reference chromosome 8. We thus compared the d-e-f breakpoint region of 'DH-Pahang' with the corresponding region of two additional accessions, for which long read sequences and sequence scaffolds were available (accession 'Banksii' from ssp. banksii and accession 'Maia'Oa' from ssp. zebrina). The comparison revealed small structural variations in 'Maia'Oa' and 'Banksii' compared with 'DH-Pahang' (Fig. 4). These structural variations, when present in the tested accession, resulted in paired-end reads that could not map on the sequence used as reference for the alignment (here 'DH-Pahang'), which may explain why SSJs of reference chromosome 8 could not be detected in some accessions. To increase the chance of detecting SSJs corresponding to chromosome 8, we accurately identified their positions on the 'Maia'Oa' preliminary assembly and the 'Banksii' corrected assembly, and aligned paired-end data from the 123 accessions using these assemblies
as additional references (Supplementary data Table S1) (see the Materials and Method for 'Banksii' assembly correction).

The expected SSJs were detected for the four accessions used as sequence template to align the paired reads ('DH-Pahang', 'Maia'Oa', 'Banksii' and 'Calcutta 4'). SSJs of chromosome 2T8 and of chromosome 8T2 were detected as expected in 'Calcutta 4'. SSJs of chromosome 2 were detected in 'Pahang', 'Maia'Oa' and 'Banksii'. However, SSJs of reference chromosome 8 were detected in paired-end reads from 'Pahang', 'Maia'Oa' or 'Banksii' only when alignment was performed on scaffolds from the same accession. This confirmed that small structural variations within accessions can hamper proper detection of SSJs and thus that only a positive result for the presence of an SSJ can be directly interpreted with confidence. However, since 'DH-Pahang', 'Banksii' and 'Maia'Oa' share the same macrostructure for chromosome 8 (Fig. 4A), the detection of SSJs in any of these three accessions was hereafter interpreted as indicating the presence of the chromosome 8 structure.

Among the 123 accessions tested, only seven displayed signatures of chromosome 2T8 and chromosome 8T2. Five wild *M.* accessions ('Long Tavoy', 'Pisang Prentel', 'Calcutta 4', 'Khae Phrae' and 'Pa Rayong') displayed the four signatures of chromosome 2T8 and chromosome 8T2, but no SSJs of reference chromosomes 2 and 8. Moreover, two *M. acuminata* cultivated accessions displayed some signatures of chromosomes 2T8 and 8T2. Accession 'Hom' had four SSJs of reference



FIG. 4. Dot-plot representation of alignments between the *M. acuminata* reference chromosome 8 and 'Maia'Oa' and 'Banksii' sequences in the region corresponding to the 2/8 translocation breakpoints. (A) Macrostructure: *M. acuminata* reference sequence for chromosome 8 (from 37 718 Mb to 37 732 Mb) is represented on the vertical axes. Homologous 'Banksii' PacBio long read and 'Maia'Oa' and opore long reads are represented on the horizontal axis. (B) Microstructure: 500 bp regions of the *M. acuminata* reference sequence – for encount of the horizontal axis. (B) Microstructure: 500 bp regions of the *M. acuminata* reference sequence of segment junction is represented on the horizontal axis. Homologous regions in 'Banksii' (corrected contig94) and 'Maia'Oa' (scaffold utg412) are represented on the vertical axes. The colour coding is the same as in Fig. 2.

chromosomes 2 and 8 and three signatures of chromosomes 2T8 and 8T2, and was thus classified as structurally heterozygous with chromosomes 2, 8, 2T8 and 8T2 structures. The 'Manang' accession had three SSJs for reference chromosomes 2 and 8 and two SSJs for chromosome 2T8, thus supporting its structurally heterozygous classification based on the 5 kb matepair discordant read analysis.

Among the remaining 90 *M. acuminata* accessions, 82 displayed the four SSJs for reference chromosomes 2 and 8 and eight accessions had at least one signature for reference chromosome 2 and one SSJ for reference chromosome 8 (Supplementary data Table S1). Two other *M. acuminata* accessions ('Pa Patthalong' and 'Pisang Serun 400') had SSJs only for reference chromosome 2.

Among the four accessions representative of other *Musa* species, one *M. schizocarpa* accession (ITC0926) had the four signatures of chromosomes 2 and 8. The other *M. schizocarpa* accession (ITC0599) and both *M. balbisiana* accessions only displayed the two SSJs of reference chromosome 2.

Finally, the 22 interspecific hybrid accessions showed 3–4 SSJs of chromosomes 2 and 8, and were thus identified as having at least one copy of reference chromosomes 2 and 8.

Regarding the 1/9 translocation, the intricate duplications at translocation breakpoints complicated the identification of SSJs specific to the four chromosome structures. We identified two SSJs for chromosome 1T9 that corresponded to k_1-k_3 and k_3 -h segment junctions, and one signature position for chromosome 9T1 that corresponded to the $h-k_1$ segment junction (Supplementary data Table S4). No segment junctions were specific to reference chromosomes 1 and 9, so no SSJs could be used to reveal these chromosome structures.

Within the 123 accessions tested, five accessions displayed one SSJ for chromosome 1T9 (SSJ k_1-k_3) ('Calcutta 4', 'Pisang Prentel', 'Long Tavoy', 'Khae Phrae' and 'Pa Rayong'). In addition, 'Calcutta 4' showed the SSJ of chromosome 9T1 (Supplementary data Table S1). Among them, 'Calcutta 4' proved to be homozygous for chromosome structures 1T9 and 9T1 based on the segregation analysis findings.

Factorial analysis with Musa acuminata diploid accessions.

A factorial analysis was performed on 3092 SNPs identified from Illumina sequences of the 83 diploid wild and cultivated *M. acuminata* accessions (Fig. 5). The first axis separated ssp. *burmannica, burmannicoides* and *siamea* accessions from the others, and explained 27 % of the SNP variance. The second axis separated ssp. *zebrina* and *banksii*, and explained 21% of the data variance. The third axis explained an additional 16 % of the data variance and separated ssp. *malaccensis* from the other accessions (Supplementary data Fig.S4).

The five wild accessions displaying only chromosomes 2T8 and 8T2 were confined in one genetic group, i.e. the *burmannica* group, which includes accessions for ssp. *burmannica*, *burmannicoides* and *siamea*. Those accessions were also those in which the translocation between chromosomes 1 and 9 was detected. Heterozygous accessions for chromosomes 2, 8, 2T8 and 8T2 ('Hom' and 'Manang') corresponded to cultivars and were located between the *burmannica*.

group and the rest of the accessions. 'Hom' was equidistant and isolated between both groups, while 'Manang' was closer to the non-*burmannica* group.

DISCUSSION

We identified two reciprocal translocations in the *Musa* acuminata ssp. burmannicoides accession 'Calcutta 4' compared with the reference genome sequence, i.e. *M. acuminata* ssp. malaccensis 'DH-Pahang'. One reciprocal translocation involved chromosomes 2 and 8, resulting in chromosomes 2 T8 and 8T2. The other one involved chromosomes 1 and 9, resulting in chromosomes 1T9 and 9T1. 'Calcutta 4' was found to be homozygous for both translocations.

These translocated chromosome structures were first searched in 13 accessions using 5 kb mate-pair discordant read analysis. In addition, SSJs specific to the distinct chromosome structures were searched in 123 accessions representative of *Musa* diversity involved in cultivated banana for which paired-end sequence data were available.

We found that these methods, particularly SSJ detection, were sensitive to the presence of polymorphism (SNPs and INDELs) and repeated sequences. Thus, only a positive result for the presence of an SSJ could be directly interpreted with confidence. However, in diploid accessions, we expected that an accession in which we detected chromosomes 1T9, 2T8, 2 or 1 also had the complementary chromosome structures 9T1, 8T2, 8 or 9, respectively. Otherwise, this accession would miss a chromosome segment, i.e. a configuration that would most probably be lethal (particularly during the haploid gametic stage). This assumption was confirmed in all cases where it could be verified. This was the case for 'Pisang Madu', 'Manang' and 'Malaccensis Nain' accessions for which chromosome 8 was detected by discordant 5 kb mate-pair read analysis, while not or partially being detected with SSJs.

Among the 36 wild accessions analysed, only five accessions were identified as having the 2T8 and 8T2 structure, most probably in the homozygous state. The same five accessions were the only ones identified as having at least one copy of the 1T9 and 9T1 chromosome structure. Chromosome 9T1 could not be detected in four of them. However, considering that the absence of the complementary structure would probably be lethal, these accessions were assumed to bear the 9T1 chromosomal structure. One of these accessions, 'Calcutta 4', proved to be homozygous for both translocations based on segregation analysis and on 5 kb mate-pair discordant read analysis. These five accessions corresponded to all wild accessions tested from ssp. burmannicoides ('Calcutta 4'), siamea ('Pa Rayong' and 'Khae Phrae') and burmannica ('Long Tavoy' and 'Pisang Prentel'). Twenty-six wild M. acuminata accessions were identified as having the 2 and 8 chromosomal structures, most probably in the homozygous state. Among them, four accessions ('Malaccensis Nain', 'Pahang', 'Borneo' and 'Maia'Oa') were shown to be homozygous for these structures based on 5 kb mate-pair discordant read analysis. One last wild M. acuminata accession ('Pisang Serun 400') could not be structurally characterized in our study.

Musa balbisiana and M. schizocarpa were only partially characterized using SSJ analysis, but they were previously



FIG. 5. Factorial analysis performed on 32 wild *M. acuminata* accessions with projections of 51 cultivated accessions. The dissimilarity matrix was based on SNPs from WGS data. Pink dots indicate accessions homozygous for chromosomes 2T8 and 8T2 and carrying at least one version of chromosomes 1T9 and 9T1. Black dots indicate accessions homozygous for chromosomes 2 and 8. Purple dots represent heterozygous accessions for the 2/8 translocation. Grey dots represent heterozygous accessions for the 2/8 translocation. Grey dots represent heterozygous accessions for the 2/8 translocation. Wild accession and signed according to their subspecies: yellow, *M. acuminata* ssp. *burmannica, burmannicoides* and *siamea*; green, *M. acuminata* subspecies. Eigenvalues: axis 1 = 27.27; axis 2 = 21.11.

reported as being homozygous for reference chromosomes 2 and 8 at the 2/8 breakpoint and for chromosomes 1 and 9 at the 1/9 breakpoint (Belser *et al.*, 2018; Baurens *et al.*, 2019).

Among the 87 cultivated accessions tested, only two accessions ('Manang' and 'Hom') were identified as having the 2/8 translocated structure, both in the heterozygous state. The 1/9 translocated structure was not detected in any cultivars. A total of 84 cultivated accessions were identified as having the 2 and 8 chromosome structure, most probably in the homozygous state. Among them, seven accessions ('Akondro Mainty', 'Pisang Madu', 'Galeo', 'Pisang Lilin', 'Paka', 'Grande Naine' and 'IDN 110') proved to be homozygous for these structures based on 5 kb mate-pair discordant read analysis. No structural characterization could be obtained for the diploid cultivated accession ('Sinwobogi').

Based on chromosomal pairing irregularities at meiosis in intersubspecific hybrids, Shepherd (1999) suggested the existence of seven *M. acuminata* translocation groups in which accessions were structurally homogeneous. The 'Standard' (ST) group is the largest one, consisting of ssp. *banksii, microcarpa* and *malaccensis* accessions. The other groups were named according to the geographic origins of their representatives. The 'Northern Malayan' (NM) group includes

some malaccensis accessions, the 'Northern 1' group includes some burmannicoides and siamea accessions, the 'Northern 2' group includes other burmannica and siamea accessions, the 'Malayan Highland' group is based on one truncata accession and the Javanese group is based on two zebrina accessions, whereas the 'East African' group is based on one unclassified accession. Martin et al. (2017) identified a reciprocal translocation involving chromosomes 1 and 4, compared with the reference genome sequence, in several malaccensis accessions, and suggested that it corresponds to the NM group, while the reference genome sequence corresponds to the ST group. The 'Calcutta 4' accession considered in this study was used by Shepherd (1999) as a representative of the 'Northern 1' group. The 'Northern 1' group was suggested to differ from the ST group by two translocations. Our results showed that the two reciprocal translocations between chromosomes 2 and 8 and between chromosomes 1 and 9 corresponded to this 'Northern group defined by Shepherd (1999).

Burmannica, siamea and burmannicoides subspecies were originally distinguished based on morphological characters and the geographical distribution (Simmonds, 1962; Perrier et al., 2009). Recent diversity studies based on diversity array technology (DArT) and simple sequence repeat (SSR) markers did not support this separation in three distinct subspecies (Sardos *et al.*, 2016; Christelová *et al.*, 2017). The presence of 1/9 and 2/8 translocations in the *burmannica*, *siamea* and *burmannicoides* accessions we analysed supported the idea of a unified *burmannica* genetic group. However, the homogeneity of this genetic group should be reassessed once the hypothesis of a third translocation specific to the 'Northern 2' group, which is assumed to include *burmannica* and *siamea* accessions (Shepherd, 1999), is confirmed. Based on our sampling, the results obtained in this study suggested that translocations 1/9 and 2/8 emerged in the *burmannica* group and supported the hypothesis that the *Musa* reference genome obtained with the *malaccensis* accession 'DH-Pahang' represented the ST group of Shepherd (1999) and the ancestral structure.

The *burmannica* genetic group has been shown to constitute an interesting genetic reservoir of disease resistance: 'Calcutta 4' shows resistance to various pests and diseases such as nematodes or black leaf streak disease, while 'Long Tavoy' and 'Khae Phrae' show resistance to *Fusarium oxysporum* f.sp. *cubense* race 4 and nematodes, respectively (Vuylsteke *et al.*, 1992; Jones, 2000; Quénéhervé *et al.*, 2009).

Our study detected only two cultivated accessions ('Manang' and 'Hom') among the 87 tested as being heterozygous for the 2/8 translocated structure, while none was detected for the 1/9 translocated structure. This supports the hypothesis that the *burmannica* group has not been an important contributor to present banana cultivars (Carreel *et al.*, 1994).

Large chromosomal translocations have been shown to generate segregation distortion and reduced recombination in progeny from structurally heterozygous parents in several plants (Tadmor *et al.*, 1987; Quillet *et al.*, 1995; Jáuregui *et al.*, 2001) including banana (Martin *et al.*, 2017; Baurens *et al.*, 2019). Their impact on chromosome segregation is expected to vary depending on the structure of the chromosomes involved (metacentric, acrocentric, etc) and the characteristic of the translocated segments (size, position, gene content, etc.). Characterizing the impact of the two translocations we described on chromosome segregation will be essential to propose breeding strategies to enhance the use of the disease resistance-rich *burmannica* genetic group in breeding programmes.

Availability of supporting data

Raw sequence reads for the 'Calcutta 4' selfing population were deposited in the SRA of the NCBI (BioProject: PRJNA532826). The 'Calcutta 4' selfing population genotyping matrix, the breakpoint regions of the 'Calcutta 4' assembly, the 'Maia'Oa' and 'Banksii' long reads including the breakpoint regions of chromosome 8, the Illumina 5 kb reads mapping in the breakpoint regions of chromosomes 1, 2, 8 and 9 of 123 Musa accessions, and the VCF file comprising 3092 SNPs for 83 diploid wild and cultivated Musa acuminata accessions are available in the download section of the Banana Genome Hub (http://banana-genome-hub.southgreen.fr/download) under AF-C4_genotypingmatrix.tab, Calcutta4_breakpoints.fasta, corrected_long_reads_chr08breakpoint_MaiaOaBanksii.fasta, breakpoints 123Musa.vcf and 2-8 1-9 translocs.vcf names, respectively.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic. oup.com/aob and consist of the following. Figure S1: principle of signature segment junction detection. Figure S2: genetic marker statistics along the 11 *Musa acuminata* chromosomes. Figure S3: paired read mapping of 14 *Musa* accessions focused on the two reciprocal translocations detected in 'Calcutta 4'. Figure S4: factorial analysis performed on 32 wild *M. acuminata* accessions with projection of 51 cultivated accessions. Table S1: accessions tested for their structure on chromosomes 2, 8, 1 and 9. Table S2: genomic co-ordinates of compared sequences. Table S3: genomic position of the SSJs for the translocation 1/8. Table S4: genomic position of the

FUNDING

This work was supported by the Agropolis Fondation 'GenomeHarvest project' (ID 1504-006) through the 'Investissements d'Avenir' programme (Labex Agro:ANR-10-LABX-0001-01), the French Government's 'Investissements d'Avenir' FRANCE GENOMIQUE programme (ANR-10-INBS-09) and the CGIAR Research Programme on Roots, Tubers and Bananas (RTB).

ACKNOWLEDGEMENTS

We are grateful to the CRB Plantes Tropicales Antilles CIRAD-INRA (Guadeloupe) and the Bioversity International *Musa* Transit Centre (ITC) for providing the plant material used in this study. The authors also thank Remy Habas, Jean-Marie Delos, Franck Marius and Jean-Claude Efile for helping with plant management, and the South Green Platform (http://www. southgreen.fr) for providing computational resources (Galaxy and UMR AGAP HPC Data Center).

LITERATURE CITED

- Baurens F-C, Martin G, Hervouet C, et al. 2019. Recombination and large structural variations shape interspecific edible bananas genomes. *Molecular Biology and Evolution* 36: 97–111.
- Belser C, Istace B, Denis E, et al. 2018. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants* 4: 879–887.
- Carreel F, Fauré S, De Leon DG, et al. 1994. Evaluation de la diversité génétique chez les bananiers diploïdes (Musa sp). Genetics Selection Evolution 26: S125.
- Camacho C, Coulouris G, Avagyan V, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10: 421. doi: 10.1186/1471-2105-10-421
- Christelová P, De Langhe E, Hřibová E, 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: 801–824.
- Daniells J, Jenny C, Karamura D, Tomekpe K. 2001. Musalogue: a catalogue of Musa gemplasm. Diversity in the genus Musa. Montpellier: INIBAP.
- D'Hont A, Denoeud F, Aury J-M, et al. 2012. The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. Nature 488: 213–217.
- Dodds KS, Simmonds NW. 1948. Sterility and parthenocarpy in diploid hybrids of *musa*. *Heredity* 2: 101–117.
- Droc G, Larivière D, Guignon V, et al. 2013. The Banana Genome Hub. Database 2013: bat035. doi: 10.1093/database/bat035.

- Fauré S, Noyer JL, Horry JP, et al. 1993. A molecular marker-based linkage M map of diploid bananas (Musa acuminata). Theoretical and Applied
- Genetics 87: 517–526. Garsmeur O, Droc G, Antonise R, et al. 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nature
- Commications 9: 12638. doi: 10.1038/s41467-018-05051-5.
 Herten K, Hestand MS, Vermeesch JR, Van Houdt JK. 2015. GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinformatics* 16: 73. doi: 10.1186/s12859-015-0514-3.
- Hippolyte I, Bakry F, Seguin M, et al. 2010. A saturated SSR/DArT linkage map of Musa acuminata addressing genome rearrangements among bananas. BMC Plant Biology 10: 65. doi: 10.1186/1471-2229-10-65.
- Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* 17: 239. doi: 10.1186/s13059-016-1103-0
- Janssens SB, Vandelook F, Langhe ED, 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. New Phytologist 210: 1453–1465.
- Jáuregui B, De Vicente MC, Messeguer R, et al. 2001. A reciprocal translocation between 'Garfi' almond and 'Nemared' peach. Theoretical and Applied Genetics 102: 1169–1176.
- Jeridi M, Bakry F, Escoute J, et al. 2011. Homoeologous chromosome pairing between the A and B genomes of *Musa* spp. revealed by genomic in situ hybridization. *Annals of Botany* 108: 975–981.
- Jones DR. 2000. Diseases of banana, abaca and enset. Wallingford, UK: CABI Publishing. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 27: 722–736.
 Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for
- Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23: 1026–1028.
- Krzywinski MI, Schein JE, Birol I, et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research* 19: 1639–1645.
- Lescot M, Piffanelli P, Ciampi AY, et al. 2008. Insights into the Musa genome: syntenic relationships to rice and between Musa species. BMC Genomics 9: 58. doi: 10.1186/1471-2164-9-58.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10–12.
- Martin G, Baurens F-C, Droc G, et al. 2016. Improvement of the banana 'Musa acuminata' reference sequence using NGS data and semi-automated bioinformatics methods. BMC Genomics 17: 243. doi: 10.1186/s12864-016-2579-4.

- Martin G, Carreel F, Coriton O, et al. 2017. Evolution of the banana genome (Musa acuminata) is impacted by large chromosomal translocations. Molecular Biology and Evolution 34: 2140–2152.
- Mbanjo EGN, Tchoumbougnang F, Mouelle AS, et al. 2012. Molecular marker-based genetic linkage map of a diploid banana population (*Musa* acuminata Colla). Euphytica 188: 369–386.
 McKenna A, Hanna M, Banks E, et al. 2010. The Genome Analysis Toolkit:
- McKenna A, Hanna M, Banks E, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20: 1297–1303.
- Noumbissié GB, Chabannes M, Bakry F, et al. 2016. Chromosome segregation in an allotetraploid banana hybrid (AAAB) suggests a translocation between the A and B genomes and results in eBSV-free offsprings. *Molecular Breeding* 36: 1–14.
- Perrier X, Bakry F, Carreel F, 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, et al. 2011. Multidisciplinary perspectives on banana (Musa spp.) domestication. Proceedings of the National Academy of Sciences, USA 108: 11311–11318.
- Quénéhervé P, Valette C, Topart P, Tezenas du Montcel H, Salmon F. 2009. Nematode resistance in bananas: screening results on some wild and cultivated accessions of *Musa* spp. *Euphytica* 165: 123–136.
- Quillet MC, Madjidian N, Griveau Y, et al. 1995. Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. Helianthus. *Theoretical and Applied Genetics* 91: 1195–1202.
 Rouard M, Droc G, Martin G, et al. 2018. Three new genome assemblies
- Rouard M, Droc G, Martin G, et al. 2018. Three new genome assemblies support a rapid radiation in *Musa acuminata* (wild banana). *Genome Biology and Evolution* 10: 3129–3140.
- Ruas M, Guignon V, Sempere G, et al. 2017. MGIS: managing banana (Musa spp.) genetic resources information and high-throughput genotyping data. Database 2017: doi: 10.1093/database/bax046.
 Sardos J, Perrier X, Doležel J, et al. 2016. DArT whole genome profiling
- Sardos J, Perrier X, Doležel J, 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.
 Shepherd K. 1999. Translocations and insertions in diploid Musa acuminata.
- Shepherd K. 1999. Translocations and insertions in diploid Musa acuminata In: Cytogenetics of the genus Musa. Montpellier: INIBAP, 15–38.
- Simmonds NW. 1962. The evolution of the bananas. London: Longmans. Southgreen Collaborators. 2016. The South Green portal: a comprehensive resource for tropical and Mediterranean crop genomics. Current Plant
- resource for tropical and Mediterranean crop genomics. *Current Plant Biology* **7–8**: 6–9. Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. 2016. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nature*
- sensitive mapping of nanopore sequencing reads with GraphMap. Nature Communications 7: 11307. doi: 10.1038/ncomms11307.
 Tadmor Y, Zamir D, Ladizinsky G. 1987. Genetic mapping of an ancient trans-
- location in the genus Lens. *Theoretical & Applied Genetics* **73**: 883–892. **Vuylsteke DR, Swennen RL, Ortiz R. 1992.** Development and performance
- of black sigatoka-resistant tetraploid hybrids of plantain (*Musa* spp., AAB group). *Euphytica* **65**: 33–42.

CHAPITRE 2 : DETERMINATION DE L'IMPACT DE DEUX TRANSLOCATIONS SUR LES SEGREGATIONS CHROMOSOMIQUES DANS LE CAS DE CROISEMENTS DIPLOÏDES

A. Contexte

Les quelques études de cartographie génétique sur le bananier réalisées jusque-là ont montré d'importantes distorsions de ségrégation (Fauré *et al.* 1993; Noyer *et al.* 1997; Vilarinhos 2004; Hippolyte *et al.* 2010; D'Hont *et al.* 2012; Mbanjo *et al.* 2012; Noumbissié *et al.* 2016). Ces distorsions étaient en général concentrées sur peu de groupes de liaisons. Jusque récemment, le lien entre ces distorsions de ségrégation et les variations structurales n'était qu'hypothétique. Un tel lien a pu être mis en évidence par Martin *et al.* (2017) sur un croisement diploïde impliquant une hétérozygotie de structure entre les chromosomes 1 et 4 et par Baurens *et al.* (2019) dans une descendance triploïde issue du croisement entre un individu tétraploïde ('CRBP39' - AAAB) présentant une hétérozygotie structurale entre les chromosomes 1, 3 et 5 (translocation 1/3 et inversion sur le chromosome 5) et un individu diploïde ('Pahang' - AA). Pour ces deux translocations, il a été observé une réduction de la recombinaison au niveau des régions entourant les points de cassure et d'importantes distorsions de ségrégation résultant d'un biais de transmission des chromosomes en faveur des chromosomes ayant la structure transloquée par rapport à la structure standard ancestrale.

Les deux translocations citées impliquent un fragment distal du chromosome 1, de 0.65Mb pour la translocation 1/3 et de 2.9 Mb pour la translocation 1/4 (Figure 11). Dans le cas de la translocation 1/4 aucune recombinaison n'a été observée le long du fragment transloqué de 2.9 Mb du chromosome 1. Par contre, dans le cas de la translocation impliquant les chromosomes 1 et 3 des recombinaisons sont observées sur le fragment transloqué de 0.65 Mb du chromosome 1. Par ailleurs, la translocation 1/3 a généré un très fort taux d'aneuploïdes (61%) dans la population triploïde dans laquelle elle a été étudiée alors qu'aucun aneuploïde n'a été observé dans la population diploïde impliquant

la translocation 1/4. On peut supposer que le niveau polyploïde (tétraploïde) est moins sensible à l'aneuploïdie, c'est-à-dire que l'aneuploïdie dans un gamète diploïde (issu d'un individu polyploïde) n'est pas forcément létale car elle est compensée par l'autre copie du chromosome alors que chez un gamète haploïde (issu d'un individu diploïde) tout chromosome ou région chromosomique manquant n'est pas compensé par un second chromosome.

Enfin, dans la population triploïde issue du croisement CRBP 39 (4X - AAAB) x 'Pahang' (2x -AA) l'inversion de 9Mb présente sur le chromosome 5 du génome B engendre une absence de recombinaison dans la région concernée et un taux accru d'aneuploïdie sur ce chromosome (Baurens et al., 2019). Un tel blocage de la recombinaison peut être expliqué par les boucles d'inversions qui se forment lors de la méiose (Dobzhansky and Sturtevant 1938) et qui produisent in fine soit des chromosomes sans centromère, soit des chromosomes avec deux centromères lorsqu'une recombinaison a lieu dans la région inversée. Bien qu'elles soient étudiées dans des contexte différents (diploïde contre tétraploïde), ces différentes variations structurales ont donc un certain nombre de conséquences communes sur la ségrégation et la recombinaison des chromosomes.

L'amélioration variétale du bananier repose sur la ségrégation des chromosomes et la possibilité de recombinaison pour combiner des caractères d'intérêt agronomique. De plus, l'identification des régions génomiques impliquées dans ces caractères dépend elle aussi de la ségrégation des chromosomes et des recombinaisons. Dans ce contexte, comprendre l'impact des différentes structures chromosomique du bananier sur ces deux composantes est crucial.

Le chapitre précédent décrit deux nouvelles translocations réciproques caractérisées chez l'accession 'Calcutta 4'. L'une concerne l'échange d'un fragment distal de 240 kb du chromosome 2 (nommé T2) et d'un fragment distal de 7.16 Mb du chromosome 8 (nommé T8) et l'autre concerne l'échange d'un fragment distal de 8,24 Mb du chromosome 1 (nommé T1) et un fragment distal de 11.56 Mb du chromosome 9 (nommé T9). Déterminer l'impact de ces deux nouvelles variations structurales sur les ségrégations chromosomiques a constitué le deuxième volet de mon travail de thèse.

Pour cela j'ai analysé la descendance du croisement ('Banksii' x 'Calcutta 4') x 'Malaccensis Nain'. L'accession 'Calcutta 4' appartient à la sous-espèce burmannicoïdes et au groupe caryotypique Northern 1 de Shepherd. Elle est porteuse des translocations 2/8 et 1/9 à l'état homozygote comme démontré dans le chapitre 1. L'accession 'Banksii' appartient à la sous-espèce banksii et est supposée appartenir au groupe caryotypique Standard de Shepherd. De fait elle est considérée homozygote pour la structure ancestrale de référence. Dans tous les cas, ni les translocations 1/9, 1/4 et 2/8 n'ont été détectées dans cette accession (Dupouy et al. 2019, Martin et al. 2017). Ainsi en croisant ces deux accessions diploïdes, on crée une descendance de F1 hétérozygotes de structure pour les translocations 1/9 et 2/8. Les deux accessions parentales appartenant à deux sousespèces différentes de Musa acuminata, elles présentent donc une divergence relativement importante. Cela devait permettre d'obtenir un grand nombre de SNPs et donc de sites ségrégeants dans les descendants étudiés. Afin de ne pas perdre de marqueurs spécifiques à ces deux sous-espèces, le F1 retenu pour étudier sa ségrégation a été croisé avec une troisième accession diploïde appartenant à une troisième sousespèce différente de Musa acuminata: l'accession 'Malaccensis Nain'. Elle appartient à la sous-espèce malaccensis et a été caractérisée comme porteuse de la translocation 1/4 à l'état homozygote (Martin et al. 2017). C'est donc la descendance issue de ce deuxième croisement qui a été analysée pour étudier l'impact des translocation 2/8 et 1/9 à l'état hétérozygote sur la ségrégation des chromosomes.



Figure 12 - Origine, et représentation chromosomiques des chromosomes 1,2, 8 et 9 chez les parents de la population BCM.

Les zones colorées sur un chromosome indiquent les fragments transloqués notés Tx (x indiquant le numéro du chromosome d'origine dans le groupe standard).

B. Matériel et Méthodes

1. MATERIEL VEGETAL ET GENOTYPAGE PAR SEQUENÇAGE

Une population (appelé BCM) de 77 individus a été obtenue à partir du croisement de l'accession 'Banksii' (*M. a.* ssp. *banksii*) avec l'accession 'Calcutta 4' (*M. a.* ssp. *burmannicoïdes*), suivi du croisement d'un des hybrides F1 obtenu avec l'accession 'Malaccensis Nain' (*M. a.* ssp. *malaccensis*) (Figure 12).

La population a été produite sur la station de recherche du CIRAD en Guadeloupe. L'ADN a été extrait au robot en plaque de 96 échantillons selon le protocole MATAB à partir des feuilles des plantules reçues sous forme de vitroplants. Les ADN ont été préparés en vue d'un génotypage par séquençage (GBS) selon le protocole établi par Poland *et al.* (2012), en utilisant les enzymes de restriction PstI et MseI. La banque ainsi construite a été séquencée au GENOSCOPE sur séquenceur Illumina HiSeq4000.

2. APPEL DE VARIANTS ET SELECTION DES MARQUEURS SNP

Les données de génotypage par séquençage de la population ont été démultiplexées en utilisant GBSX 1.2 (Herten *et al.* 2015) et les adaptateurs ont été enlevés avec l'outil cutadapt (Martin 2011). L'appel de variants a été réalisé avec un jeu de données regroupant les données GBS obtenues à partir de la population BCM et un ensemble de données de séquence (WGS -Séquençage d'ADN total- et GBS) disponibles pour les accessions parentales. Pour le parent 'Malaccensis nain' nous disposons de données de séquences provenant de 5 séquençages en GBS, un séquençage de génome complet Illumina mate-pair et un séquençage de génome complet Illumina pairé standard, pour le grand-parent 'Banksii' d'un séquençage de génome complet Illumina pairé standard et pour le grand parent été réalisé en utilisant le package VcfHunter (Garsmeur *et al.* 2018). Le mapping a été réalisé sur la séquence de référence *Musa acuminata* version 2



Figure 13 - Représentation des distorsions de ségrégation dans la population pour les marqueurs GBS dans la population BCM.

Distribution des valeurs de probabilités du χ^2 de Pearson (-Log₁₀ p-value).

(Martin *et al.* 2016). Les sites avec une profondeur inférieure à 10 ou supérieure à 1000 ont été exclus. Les sites avec un des deux allèles ayant une fréquence inférieure à 5% sont considérés homozygotes. Réciproquement, les sites avec un des deux allèles ayant une fréquence supérieure à 15% et supportée par au moins 3 lectures sont considérés hétérozygotes. Les autres sites sont convertis en données manquantes. Les variants obtenus à l'issue de cette étape vont être utilisés pour l'analyse des ratios alléliques par chromosome painting. Pour les analyses de ségrégation des marqueurs, plusieurs étapes de filtrations additionnelles vont être réalisées.

Dans cette analyse nous voulions accéder à la ségrégation des marqueurs du parent hétérozygote de structure, c'est à dire la ségrégation du parent F1 issu du croisement 'Banksii' X 'Calcutta 4'. Nous recherchons donc les sites polymorphes qui sont hétérozygotes dans le F1 et homozygote chez 'Malaccensis Nain'. Pour de tels marqueurs (un parent hétérozygote et l'autre parent homozygote) et selon les lois de Mendel, la ségrégation attendue dans la descendance est de 50% d'individus homozygotes et 50% d'individus hétérozygotes. Les sites polymorphes ont donc été sélectionnés sur la base de cette ségrégation avec un test de Khi2. Le seuil minimal de p-valueur a été fixé à 1e-15 après analyse de leur distribution (Figure 13) afin d'exclure la queue de la distribution correspondant à une grande quantité de marqueurs non ségrégeants (probablement issus d'erreur de génotypage ou de génotypage de régions multilocis), tout en évitant d'éliminer des marqueurs dont la ségrégation pourrait diverger de celle attendue (vraies distorsions de ségrégation). Durant cette étape de sélection des marqueurs, seuls les marqueurs avec moins de 2.5% de données manquantes au sein de la population ont été utilisés. Un total de 23 276 marqueurs homozygotes chez le parent 'Malaccensis Nain' et hétérozygotes chez le parent F1 ont été sélectionnés. Les 73 individus avec moins de 25 % de données manquantes ont été conservés.

Pour réaliser la suite des analyses, un marqueur sur sept le long des chromosomes a été retenu. Ceci nous a permis de réduire le jeu de données final à 3326 marqueurs afin de réduire le volume des données traitées et ainsi diminuer les temps de calcul.



Figure 14 - Représentation schématique d'un "chromosome painting" et des génotypes correspondants.

Le génotype de l'hybride étudié est dessiné sous forme de barre colorées horizontales. Pour chacun des allèles de l'hybride, une couleur spécifique est assignée en fonction de l'origine parentale. Le ratio allélique est représenté le long des chromosomes du génome de référence de Musa acuminata 'Pahang HD' (abscisse). Chaque point de couleur représente donc les allèles d'un parent, les ratios alléliques attendus chez les individus diploïdes (a et b) sont de 50%. Des ratios alléliques de 33% (c et d) et / ou 66% (d) peuvent être observés si les individus sont triploïdes (c et d).

3. LIAISON ENTRE MARQUEURS

Pour chacun des marqueurs SNP du jeu de données final, le taux de recombinaison par paire de marqueurs a été calculé avec le script RecombCalculatorDDose.py du package VcfHunter. Cette information est ensuite représentée sous forme de dotplot en utilisant l'outils Draw_dot_plot.py du même package.

4. ANALYSE DES RATIOS ALLELIQUES PARENTAUX, DE LA PLOÏDIE ET DES POINTS DE RECOMBINAISON

Pour analyser les différentes contributions des parents dans la descendance, la ploïdie de ces individus et les recombinaisons ayant eu lieu, le programme vcf2allPropAndCov.py de la suite d'outils VcfHunter (Baurens *et al.* 2018) a été utilisé sur les 73 individus de la population.

Pour identifier les contributions parentales, le programme fonctionne en deux étapes. La première étape permet d'identifier l'origine parentale des allèles sur la base de leur présence dans l'un des parents et de leur absence dans l'autre parent. Les allèles sont attribués à une origine parentale lorsqu'ils sont retrouvés exclusivement dans le pool d'allèles parentaux de cette origine. La deuxième étape calcule le ratio allélique (nombre de reads soutenant cet allèle divisé par le nombre de reads couvrant la zone) pour chaque individu de la descendance et pour chacun des allèles attribués à un parent. Ce ratio allélique est ensuite représenté le long des chromosomes de la séquence de référence (Figure 14) avec une couleur spécifique à l'origine parentale (bleu pour 'Malaccensis Nain', orange pour 'Calcutta 4' et vert pour 'Banksii'). Ce type de représentation est produit pour chaque individu de la population BCM. Dans la suite du document, lorsque nous ferons référence à ce type d'analyse nous parleront de "chromosome painting". Ces représentations sont ensuite utilisées pour identifier pour chaque individu de la descendance l'origine parentale et "grand-parentale" des différents segments chromosomiques qui la composent. Ainsi la population étant issue du croisement ('Banksii' x 'Calcutta 4') x 'Malaccensis Nain', les descendants sont supposés avoir hérité



Figure 15 - Exemple de "chromosome painting" obtenu sur un individu de la descendance analysée.

Pour chacun des sites polymorphes analysés, les allèles sont attribués à l'un des individus parentaux ou grand-parentaux. Et la proportion de reads soutenant chacun des allèles pour l'individu étudié est calculé. Cette valeur est ensuite représentée en ordonnée avec un point coloré correspondant à l'origine du parent/grand-parent: bleu, vert, jaune respectivement pour 'Malaccensis nain', 'Banksii' et 'Calcutta 4'. Etant donné le croisement, une origine bleu est attendue pour tous les chromosomes (points bleus autours de 0.5) et une alternance blocs de points vert et jaunes correspondant à une alternance d'origine 'Banksii' et 'Calcutta 4' est attendue. Le passage le long d'un chromosome d'un bloc vert à un bloc jaune et inversement sera interprété comme un point de recombinaisons. Les traits verticaux représentent les zones de réarrangements chromosomiques entre les structures "Standard" et "Northern 1". L'analyse de l'origine de l'individu étudié.

pour chaque chromosome d'une copie (bleue) provenant du parent '*Malaccensis* Nain' et d'une copie résultant de recombinaisons entre les chromosomes des grands parents 'Banksii' et 'Calcutta 4' (vert et orange). Grâce à cette information, nous pouvons donc accéder pour chaque individu aux événements de recombinaison ayant eu lieu dans le F1. Ces événements sont repérables par l'alternance entre les régions composées de marqueurs colorés en orange et les régions composées de marqueurs colorés en vert. On a également accès à la structure chromosomique des individus. En effet, une origine 'Banksii' dans les régions autour des points de réarrangement indique une structure de chromosomes Standard alors qu'une origine 'Calcutta 4' indique une structure Northern 1 (Figure 15).

D'autre part, l'outil vcf2allPropAndCov.py permet une analyse de l'aneuploïdie des individus le long de leur génome. Pour cela, pour chaque individu, le nombre de reads couvrant chacun des marqueurs est calculé puis divisé par la couverture moyenne de ce même individu. La figure 16 schématise le type de représentation généré par le programme.

Pour réaliser cette analyse, j'ai utilisé le fichier VCF de 671 223 marqueurs SNP généré avant les étapes de filtration sur les ségrégations et l'homozygotie du parent F1 afin de maximiser la densité en points de donnée.



r osition le long du chromosom

Figure 16 - Représentation schématique de la figure permettant d'analyser la ploïdie le long des chromosomes.

La couverture en lectures (ordonnée) de chacun des marqueurs est représentée selon sa position (abscisse) le long du chromosome de référence. La couverture moyenne de l'accession est représentée par la droite noire avec une zone rouge représentant plus ou moins 0.25 fois la valeur moyenne.



Figure 17 - Représentation de l'intensité de l'association génétique des marqueurs hétérozygotes du F1 dans la population BCM.

Chaque marqueur est ordonné le long des chromosomes de la séquence de référence (de 1 à 11). L'intensité de l'association génétique est représentée par un point de couleur chaude (forte liaison) à froide (faible liaison). (A) Les marqueurs génétiques sont représentés à la suite les uns des autres avec une distance arbitraire. (B) Les marqueurs génétiques sont représentés suivant leurs coordonnées sur les scaffolds de la séquence de référence (blocs gris).

C. Résultats

1. LIAISON ENTRE LES MARQUEURS

J'ai calculé le taux de recombinaison pour chaque paire de marqueurs dans le sous ensemble de 3326 marqueurs sélectionnés. Les taux de recombinaison ont ensuite été projetés sur la carte physique du génome de référence *Musa acuminata* ssp. *malaccensis* 'Pahang HD' (Figure 17). On observe comme attendu une forte liaison (symbolisée par des couleurs chaudes) entre les marqueurs physiquement proches et entre les marqueurs dans les zones centromériques, au niveau desquelles un taux de recombinaison faible voire nul est attendu (Talbert and Henikoff 2010). Hors des zones centromériques, les liaisons génétiques entre marqueurs s'affaiblissent avec l'augmentation de la distance physique. Les marqueurs situés sur les deux bras des chromosomes métacentriques ainsi que les marqueurs situés sur différents chromosomes apparaissent non liés.

On observe toutefois deux exceptions notables : (i) une forte liaison génétique entre des marqueurs de la fin du chromosome 2 et des marqueurs du chromosome 8 ; (ii) une forte liaison génétique entre des marqueurs du chromosome 1 et des marqueurs du chromosome 9.

Ces régions fortement liées sont celles impliquées dans les translocations 2/8 et 1/9 qui sont présentes à l'état hétérozygote chez le parent F1 de la population étudiée. Cette liaison est donc attendue puisque pour une des versions des chromosomes du F1, des marqueurs des chromosomes 1 et 9 (chromosome 1T9 et 9T1) d'une part et des marqueurs des chromosomes 2 et 8 (chromosome 2T8 et 8T2) d'autre part sont physiquement liés (sur le même chromosome) et donc apparaissent génétiquement liés.

On observe également du bruit de fond, c'est à dire des liaisons génétiques entre marqueurs non proches physiquement (souvent un des marqueurs de la paire est associé au chromosome 11). Ces liaisons impliquent des marqueurs isolés et peuvent correspondre à des erreurs d'assemblage dans le génome de référence. Il en résulte un mauvais placement du marqueur. Il s'agit plus probablement de marqueurs de mauvaise qualité qui n'ont pas été éliminés lors de l'étape de filtration sur la ségrégation.



Figure 18 - Répartition des points de recombinaisons observés dans la descendance du F1 le long des chromosomes impliqués dans les réarrangements chromosomiques.

Chaque trait vertical représente une recombinaison. Les blocs notés T1, T2, T9 et T8 représente les blocs transloqués. Les traits noirs horizontaux indiquent les zones de faibles recombinaisons.



Figure 19 – Chromosome painting du chromosome 8 de 16 individus portant une aneuploïdie segmentale.

Les points colorés en bleu, vert et orange représentent les allèles d'origine 'Malaccensis Nain', 'Banksii' et 'Calcutta 4' respectivement. Le ratio allélique est représenté le long des chromosomes du génome de référence de Musa acuminata 'Pahang HD' (abscisse). Le trait vertical représente le point de cassure du chromosome 8 pour la translocation avec le chromosome 2.

2. RECOMBINAISONS

L'approche de détection des événements de recombinaison par chromosome painting a été réalisée sur les 73 individus de la descendance à partir des 671 223 marqueurs SNP générés avant les étapes de filtration sur la ségrégation des marqueurs et leur homozygotie.

De façon à déterminer si la présence des translocations réciproques 1/9 et 2/8 à l'état hétérozygote a un impact sur les recombinaisons, les événements de recombinaison le long de ces chromosomes 1, 2, 8 et 9 ont été comptabilisés pour tous les individus et représentés sur un diagramme pour chaque chromosome (Figure 18). On a pu dénombrer 33 recombinaisons impliquant le segment (T1) de 8,24 Mb (soit environ 4 recombinaisons par Mb), 47 recombinaisons impliquant le segment (T9) de 11,56 Mb (soit 4.1 recombinaisons par Mb) et 38 recombinaisons impliquant le segment (T8) de 7,16 (soit 5.3 recombinaisons par Mb). Aucune recombinaison n'a été détecté au niveau de la zone transloqué du chromosome 2 (T2), mais ce segment étant de très petite taille (240 kb) on ne peut exclure que la résolution dans la méthode de chromosome painting n'a pas permis de détecter des recombinaisons. On observe des régions de faibles recombinaisons au voisinage direct des points de translocation des chromosomes 1, 2 et du chromosome 8.

3. ANEUPLOÏDIE

L'analyse de la couverture des marqueurs le long des chromosomes de chaque descendant a montré pour 16 descendants une augmentation de la couverture au niveau du segment T8 du chromosome 8. D'autre part, ces 16 individus présentent des ratios alléliques particuliers sur ce même segment avec la présence d'allèles des trois origines



Origine : 📰 'Banksii' 🦰 'Calcutta 4' 📩 'Malaccensis Nain'

Figure 20 – Chromosome painting de trois situations de ploïdie différentes.

La couverture en lectures de chacun des marqueurs est représentée selon sa position le long du chromosome de référence. La couverture moyenne de l'accession est représentée par la droite noire avec une zone rouge représentant plus ou moins 0.25 fois la valeur moyenne. Les points colorés en bleu, vert et orange représentent les allèles d'origine 'Malaccensis Nain', 'Banksii' et 'Calcutta 4' respectivement. Le ratio allélique est représenté le long des chromosomes du génome de référence de Musa acuminata 'Pahang HD' (abscisse). Le génotype de l'individu étudié est dessiné sous forme de barre colorées horizontales. (A) Chromosome euploïde : la couverture en lecture est en moyenne centrée autour de la couverture moyenne de l'accession. A un même locus, les ratios alléliques sont en moyenne de 0.5 pour ceux origines parentales exclusivement. L'individu étudié possède un haplotype d'origine 'Malaccensis Nain' et un haplotype d'origine F1 ('Banksii' x 'Calcutta 4'). (B) Aneuploïdie segmentale : la couverture en lecture n'est pas homogène le long du chromosome et augmente le long du fragment T8. Les ratios alléliques sont de 0.5/0.5 en moyenne sur la majeure partie du chromosome, mais chutent à 0.33/0.33/0.33 avec des allèles appartenant aux trois origines parentales le long du fragment T8. L'individu étudié est euploïde le long de la majeure partie du chromosome 8 et triploïde le long du fragment T8. L'individu est porteur d'un haplotype d'origine 'Malaccensis Nain', d'un haplotype du chromosome 8 d'origine F1 et d'un haplotype du chromosome 2T8 d'origine F1. (C) Chromosome aneuploïde complet : La couverture moyenne en lectures le long du chromosome est plus élevée que la couverture moyenne de l'accession. Tout le long du chromosome, les ratios alléliques sont de 0.33/0.33/0.33 pour les trois origines parentales. L'individu est porteur d'un haplotype du chromosome 2 d'origine 'Malaccensis nain', d'un haplotype du chromosome 2 d'origine F1 et d'un haplotype du chromosome 2T8 d'origine F1.

parentales ('Malaccensis Nain', 'Banksii' et 'Calcutta 4') systématiquement retrouvés, à un ratio de 33% chacune, alors que pour le reste du chromosome 8, on observe bien un ratio allélique de 50% 'Malaccensis Nain' (bleu) complété par 50% de 'Banksii' (vert) ou 50% de 'Calcutta 4' (orange) (Figure 19). Chez ces 16 accessions, ces observations s'expliquent par la présence du fragment T8 en trois copies au lieu de deux. Ceci résulte de la présence d'un chromosome 2 et d'un chromosome 8 venant de 'Malaccensis nain' et un chromosome 8 et d'un chromosome 2T8 provenant du F1 (Figure 20 B). Ces individus sont donc des aneuploïdes segmentaux.

Un de ces 16 descendants aneuploïdes segmental (BCM_21, Annexe 10), portant une copie supplémentaire du fragment T8 présente également une augmentation de la couverture pour le chromosome 2 et trois origines parentales en proportion équivalente (33% chacune) le long du chromosome 2 (Figure 20 C). Ceci suggère que cet individu porte une copie supplémentaire du chromosome 2T8. Aucune recombinaison n'est observée le long du chromosome 2 de cet individu.

Concernant le segment T2 du chromosome 2, il est trop petit pour pouvoir conclure s'il est concerné par les changements de couverture et donc des aneuploïdies segmentales.

On peut noter que 8 des 16 accessions aneuploïdes segmentales présentent un événement de recombinaison dans le segment T8, que l'on peut constater par la disparition de l'une des deux origines du parent F1 et le doublement de la dose de l'origine restante. Ces individus ont donc en région terminale du fragment T8 un ratio allélique de 33% 'Malaccensis Nain' (bleu) et 66% 'Banksii' (vert) ou 'Calcutta 4' (orange) (Figure 21).



Figure 21 – Chromosome painting d'un individu aneuploïde sur le segment T8 et présentant une recombinaison dans cette région.



Figure 22 - Distorsion de ségrégation dans la population BCM.

La valeur de la distorsion (- log₁₀ de la probabilité d'un test de χ^2 testant la ségrégation observée par rapport à un attendu de 50% / 50%) est représentée le long des 11 chromosomes de la séquence de référence Pahang HD. Les traits verticaux indiquent les coordonnées des points de translocation.

4. BIAIS DE SEGREGATION CHROMOSOMIQUE

a) Distorsions de ségrégation des marqueurs

L'analyse des distorsions de ségrégation (-log10(χ^2)) le long des 11 chromosomes du bananier sur les 57 individus diploïdes de la population révèle des distorsions de ségrégation sur les chromosomes 1, 2, 8 et 9 impliqués dans les réarrangements chromosomiques (Figure 22). Cependant d'autres chromosomes (3, 4 et 10) présentent des régions où la distorsion de ségrégation est similaire à celle observée sur les chromosomes réarrangés. Enfin, le chromosome 11 présente de très importantes distorsions de la ségrégation dans sa région centromérique ayant pour cause une important surreprésentation de l'haplotype 'Calcutta 4' au détriment de l'haplotype 'Banksii' dans la descendance.

b) Biais de transmission des combinaisons alléliques

J'ai ensuite analysé les combinaisons structurales transmises à la descendance au niveau des segments impliqués dans les translocations. Pour cela j'ai exploité les figures de chromosome painting pour identifier, pour chaque descendant, son contenu en chromosome 2, 8, 2T8, 8T2, 1, 9, 1T9 et 9T1. Ainsi, pour un individu, lorsqu'au point de translocation d'un chromosome le chromosome painting indique une origine 'Banksii', on peut alors conclure que cet individu porte, pour ce chromosome, la structure de référence. Réciproquement, si au point de translocation d'un chromosome le chromosome que cet individu porte, pour ce chromosome les translocations réciproques 2/8 et 1/9), on peut alors conclure que cet individu porte que cet individu porte, pour ce chromosome, la structure transloquée. C'est ainsi qu'ont été dénombrées toutes les combinaisons structurales des 73 descendants de la population, répertoriées dans la figure 23 pour les structures impliquant les chromosomes 1 et 9 et dans la figure 24 pour les structures impliquant les chromosomes 2 et 8.



Figure 23 – Tableau de contingence de la structure chromosomique aux points de cassure de la translocation 1/9 des 73 individus de la population BCM.



Figure 24 - Tableau de contingence de la structure chromosomique aux points de cassure de la translocation 2/8 des 73 individus de la population BCM.

Dans le cas de la structure 1/9, on observe aucun individu portant un mélange des deux structures (chr1 + chr9T1 ou chr9 + chr1T9). Seules les combinaisons de chromosomes 1 + 9 et chromosomes 1T9 + 9T1 sont observées. Ces deux chromosomes ne sont donc pas indépendants dans leur ségrégation. Par ailleurs, la combinaison chromosomique 1T9 + 9T1 est préférentiellement transmise à la descendance puisque plus de deux fois plus représentée (73%) que la structure 1 + 9 (27%).

Dans le cas de la structure 2/8, 22 individus (29%) présentent la structure standard avec les chromosomes 2 et 8, 35 individus (49%) présentent la combinaison de structure alternative (2T8 et 8T2) et 15 individus (20%) présente une combinaison des deux structures (2T8 + 8). La combinaison des deux structures (2 + 8T2) n'est pas retrouvée dans la descendance. D'autre part, l'individu aneuploïde BCM 21 présente 23 chromosomes au lieu des 22 chromosome attendus, ainsi qu'un segment T8 surnuméraire. Il a hérité des chromosomes 2 et 8 du parent *'Malaccensis* nain', des chromosomes 2 et 8 du grand parent *'Banksii'* et du chromosome supplémentaire 2T8 du grand parent 'Calcutta 4'.

D. Discussion

Les 73 individus d'une population issus du croisement ('Banksii' x 'Calcutta 4') x 'Malaccensis Nain' ont été génotypés par séquençage (GBS) et analysés. Dans cette partie de mon travail, j'ai analysé l'impact des deux translocations réciproques 2/8 et 1/9 (caractérisées chez l'accession 'Calcutta 4') sur les recombinaisons et les ségrégations chromosomiques dans une descendance issue d'un parent hétérozygote de structure. Ces deux translocations consistent en l'échange d'un fragment distal de 240 kb du chromosome 2 et d'un fragment distal de 7.16 Mb du chromosome 8 et l'échange d'un fragment distal de 8,24 Mb du chromosome 1 et un fragment distal de 11.56 Mb du chromosome 9.



Figure 25 – Comparaison du chromosome painting obtenu à partir de données RAD-seq (A) et GBS (B), adapté de Baurens et al., 2019. Les ratios alléliques obtenus à partir de données RAD-seq sont plus denses que les ratios alléliques obtenus à partir de données GBS.

671 223 marqueurs SNP obtenus à partir de cette population ont pu être attribués à l'une des trois origines parentales (Banksii, Calcutta 4 ou Malaccensis Nain). Cette analyse visualisée à l'aide d'un chromosome painting a permis de déterminer l'origine parentale des 22 chromosomes de chaque individu de la population, de détecter les évènements de recombinaison et de repérer des cas d'aneuploïdie et d'aneuploïdie segmentale (régions chromosomiques non diploïdes) chez certains individus.

Le calcul des ratios alléliques a été réalisé sur un large jeu de données filtré uniquement sur les données manquantes afin de maximiser la densité du chromosome painting, en dépit du bruit de fond. On observe, de fait, une dispersion des ratios alléliques dans un large nuage de points chez toutes les accessions, alors que l'on pourrait théoriquement s'attendre à une dispersion plus réduite autour du ratio de 50% le long de tous les chromosomes de chaque individu, aux régions aneuploïdes près. Les méthodes de Génotypage par Séquençage sont connues pour générer des variations de couverture des régions séquencées. On peut observer chez Baurens et al. (2019) la différence entre les ratios alléliques calculés à partir d'individus séquencés par GBS et d'autres séquencés par RAD-seq (Figure 25). Contrairement au GBS, le RAD-seq inclus en plus des étapes de restriction enzymatique une étape de shearing (technique de fragmentation mécanique du génome) et de sélection des fragments en fonction de leur taille (Baird *et al.* 2008). Le RAD-seq semble générer des résultats de meilleure qualité avec notamment de très fortes et stables couvertures, mais est aussi plus long et plus délicat à mettre en place. Dans le cadre de notre analyse, bien que dispersés, les nuages de points permettent une lecture claire du chromosome painting, c'est-à-dire l'identification du nombre de copies de chaque fragment chromosomique et de leur origine parentale. Les quelques régions où ont été détectés des ratios alléliques différents des 50%/50% attendu sont clairement détectables (Figure 18). Nous avons donc pu exploiter les données de ratio alléliques via le chromosome painting afin d'analyser l'impact des deux translocations sur les recombinaisons et les ségrégations chromosomiques.

Dans le voisinage direct des points de translocation du chromosome 1, 2 et du chromosome 8, on observe des régions où la recombinaison est réduite d'environ 2.2Mb, 2.6Mb et 1.2Mb respectivement. Les "cold spots" de recombinaison sont fréquents au sein des génomes, et il a été montré chez d'autres organismes qu'ils pouvaient être liés à la présence de variations structurales (He and Dooner 2009; Morgan et al. 2017). Concernant le chromosome 9, le point de translocation se trouve en région péricentromérique, or ces régions centromériques et péricentromériques correspondent elles-mêmes à des régions ou la recombinaison est absente ou très limitée (Beadle 1932). Ces réductions de recombinaison autour des points de translocations confirment les observations déjà réalisées dans le cas de la translocation entre les chromosomes 1 et 4 présente chez un sous-groupe de la sous-espèce malaccensis (Martin et al. 2017) et de la translocation entre les chromosomes 1 et 3 présente chez l'espèce Musa balbisiana (Baurens et al. 2019). Dans le cas de la translocation 1/4, l'intégralité d'un fragment transloqué (3Mb du chromosome 1) présentait une suppression totale de recombinaison. Deux hypothèses ont été avancées pour expliquer ce phénomène : la présence d'une inversion subterminale mais qui à l'heure actuelle n'a pas été détectée ; et la taille relativement petite du fragment couplée au caractère acrocentrique du chromosome. Concernant les segments transloqués T1, T8 et T9, ils ne semblent pas soumis à une réduction de la recombinaison sur le fragment entier. Au niveau du segment T2 aucune recombinaison n'a été détectée. Ceci peut être dû à sa taille très réduite (on retrouve en moyenne 4 recombinaisons par Mb dans la population, il y a donc une forte probabilité qu'un fragment de 240 kb n'en présente aucune) et à la méthode utilisée (chromosome painting) qui ne serait pas suffisamment résolutive.

Lorsque l'on s'intéresse à la ségrégation des chromosomes dans cette population, on observe systématiquement dans le cadre de la translocation 1/9, soit des descendants présentant les deux structures chromosomiques transloquées, soit des descendants présentant les deux structures chromosomiques non transloquées. Cette configuration résulte vraisemblablement du fait que les gamètes déséquilibrés (c'est-à-dire dans lesquels un des segments T1 (8.24 Mb) ou T9 (11.58 Mb) est absent et l'autre en excès) sont non viables. Ce type de biais a déjà été observé dans un contexte diploïde dans la population hétérozygote de structure pour la translocation 1/4 (Martin *et al.* 2017).

Dans le cas de la translocation 2/8, on observe une même tendance et aucune combinaison 2+8T2 n'a été détectée, évitant ainsi le déséquilibre gamétique dû à l'absence du fragment T8 (7.16 Mb). En revanche, 15 individus (20,5% de la population) portent la combinaison 8+2T8, une combinaison où le segment T2 est absent. Le segment T2 étant de très petite taille (240 kb incluant 38 gènes d'après l'annotation de D'Hont *et al.* 2012, https://banana-genome-hub.southgreen.fr/) son absence pourrait ne pas être délétère au niveau des gamètes contrairement à l'absence de segments plus grands tels que les T8, T1, T9. Les seize individus portant la combinaison 8+2T8 sont donc des aneuploïdes segmentaux avec 3 copies du fragment T8 (une copie provenant du chromosome 8 issu du parent *'Malaccensis* Nain', une copie provenant du chromosome 2T8 issu du parent T2 serait lui en une copie (c'est-à-dire à l'état haploïde), avec comme unique source le chromosome 2 du parent *'Malaccensis* Nain' chez quinze de ces individus.

Lorsque l'on observe maintenant la répartition de ces combinaisons de structures chromosomiques dans la population, on constate que les haplotypes transloqués (1T9, 9T1, 2T8 et 8T2 de 'Calcutta 4') sont préférentiellement transmis à la descendance. En effet, dans le cas de la translocation 1/9, la combinaison 1T9+9T1 est retrouvée dans 73% de la population. Cette transmission préférentielle et son ratio associé correspondent au comportement de la translocation 1/4 dans une population issue de l'autofécondation d'un parent hétérozygote de structure décrite par Martin *et al.* (2017). Dans le cas de la translocation 2/8, on constate que le chromosome 2 non transloqué d'origine 'Banksii' est transmis dans 30% de la population, alors que sa version transloquée 2T8 d'origine 'Calcutta 4' est transmise dans 70% de la population, corroborant ainsi les observations faites dans le cadre des translocations 1/9 et 1/4. La transmission préférentielle des haplotypes transloqués a aussi été observée dans le contexte triploïde de la translocation 1/3. En revanche, ce pattern de transmission n'est pas retrouvé pour le chromosome 8 et son alternative transloquée 8T2. En effet, la présence des individus aneuploïdes vient équilibrer ces ratios de transmission. L'haplotype 8T2 originaire du parent 'Calcutta 4' est retrouvé chez 49% des individus de la population (61% si on exclut les individus

aneuploïdes) alors que l'haplotype 8 originaire du parent 'Banksii' est retrouvé chez 51% des individus de la population (39% si on exclut les individus aneuploïdes).

Les distorsions de ségrégation des marqueurs observées le long des chromosomes 1, 2, 8 et 9 peuvent être expliquées par la présence de translocations impliquant ces chromosomes. Au niveau de la plupart des autres chromosomes, on observe également des zones avec des ségrégations distordues dans des proportions similaires (pour les chromosomes 3, 4, 6 et 10) voire dans des proportions beaucoup plus importantes (chromosome 11). Il pourrait être intéressant d'étudier les mécanismes à l'origine des distorsions observées, en explorant le contenu en gènes de ces régions par exemple. Dans le cas du chromosome 11, les très fortes distorsions de ségrégation des marqueurs sont liées à la très faible présence d'allèles attribués au parent 'Banksii' dans les régions péricentromériques et centromériques de ce chromosome. 68/73 individus de cette population ont les allèles du parent 'Calcutta 4' dans cette région du chromosome 11.

En conclusion, comme les translocations 1/3 (Baurens *et al.* 2019) et 1/4 (Martin *et al.* 2017), les translocations réciproques 1/9 et 2/8 génèrent à l'état hétérozygote une baisse de la fréquence des recombinaisons dans l'entourage de leurs points de cassure. En revanche elles ne provoquent pas de suppression totale des recombinaisons dans les fragments transloqués contrairement à ce qui a pu être observé sur le chromosome 1 dans le cadre de la translocation 1/4. Elles génèrent également des distorsions qui résultent de la ségrégation biaisée des chromosomes concernés, distorsions qui résultent de la ségrégation biaisée des chromosomes en faveur des structures chromosomiques transloquées 1T9, 9T1 et 2T8, corroborant là aussi les observations de Martin *et al.* (2017) et Baurens *et al.* (2019). La translocation 2/8 favorise quant à elle l'existence d'individus aneuploïdes (20% de la population) porteurs des structures chromosomiques 8 et 2T8. Ces individus sont donc triploïdes sur le fragment T8 et haploïdes sur le fragment T2. La présence de ces individus aneuploïdes équilibre la ségrégation des chromosomes 8 et 8T2, qui serait elle aussi en faveur de la structure transloquée sans cette aneuploïdie.

Ces informations sur l'impact des translocations réciproques 1/9 et 2/8 de l'accession 'Calcutta 4' (*M. a.* ssp. *burmannicoïdes*) sur les ségrégations chromosomiques dans une population structurellement hétérozygote pourront être utile pour réfléchir à des
stratégies d'exploitation de ce matériel génétique dans les programmes d'amélioration. A l'heure actuelle, cette accession et plus généralement le pôle génétique *'burmannica'* contribue peu à la diversité des bananiers cultivés (Carreel *et al.* 1994; Dupouy *et al.* 2019) et ce malgré sa richesse en caractères de résistance à divers parasites et maladies (Vuylsteke *et al.* 1992; Jones 2000; Quénéhervé *et al.* 2009; Arinaitwe *et al.* 2019). D'après Vilarinhos (p95, 2004), sur la base de travaux de cartographie et de recherche de QTL réalisés par F. Carreel (com. pers.), un QTL de résistance à la cercosporiose a été identifié chez 'Calcutta 4' au sein d'un groupe de liaison (LGII) qui correspondrait au fragment T8. Les analyses présentes pourront permettre d'évaluer la probabilité de transfert des locus d'intérêt agronomique dans une descendance en fonction de l'appartenance des parents à un groupe de translocation et ainsi guider les sélectionneurs dans leur choix de géniteurs.



Figure 26 - Représentation des groupes caryotypiques et des translocations caractérisées à l'heure actuelle.

DISCUSSION ET PERSPECTIVES

A. Variations structurales chez le bananier : avancée des connaissances

Avant ma thèse, l'équipe avait caractérisé, au sein de l'espèce Musa acuminata, une translocation impliquant les chromosomes 1 et 4 et avait montré que cette structure a émergé dans un sous-groupe de la sous-espèce malaccensis (Martin et al. 2017). Shepherd (1999) sur la base de l'analyse des figures d'appariement chez des hybrides entre sousespèces de M. acuminata avait identifié au sein des M. acuminata 9 translocations réparties dans 6 groupes caryotypique (Figure 26). Les travaux de l'équipe suggéraient que la translocation entre les chromosomes 1 et 4 correspondait à la translocation portée par le groupe caryotypique Northern Malayan. D'autre part, en parallèle de ma thèse, deux grandes variations structurales (une translocation entre les chromosomes 1 et 3 et une inversion sur le chromosome 5) ont été caractérisées chez l'espèce Musa balbisiana par rapport à l'espèce M. acuminata (Baurens et al. 2019). Ces deux études suggèrent que la structure Standard définie par Shepherd (1999) correspond à la structure ancestrale chez M. acuminata et à la structure du génome de M. a. ssp. malaccensis 'Pahang HD' qui a servi pour établir la séquence de référence du génome de M. acuminata. Mes travaux ont permis de caractériser au sein de l'espèce Musa acuminata deux nouvelles translocations impliquant les chromosomes 1 et 9 d'une part et 2 et 8 d'autre part. Ces deux translocations ont été retrouvées au sein du groupe génétique burmannica, qui comprends les sous-espèces burmannica, burmannicoïdes et siamea. Elles semblent être ubiquitaires au sein de ce groupe génétique, puisqu'elles ont été retrouvées chez toutes les accessions des trois sous-espèces concernées que j'ai pu tester. Ce groupe caryotypique comportant ces deux translocations correspondrait au groupe Northern 1 de Shepherd (1999). Ces résultats consolident l'hypothèse que la structure portée par la séquence de référence de M. a. ssp. malaccensis 'Pahang HD' correspondrait à la structure ancestrale et au groupe Standard défini par Shepherd (1999).

Shepherd avait suggéré la présence de 9 translocations chez *M. acuminata*, il resterait donc au moins 6 translocations à caractériser chez *M. acuminata* qui correspondraient aux groupes de translocation de Shepherd Northern 2 (identifié par Shepherd chez trois accessions d'un sous-groupe du groupe génétique *burmannica*), Javanese (identifié par Shepherd chez deux accessions de la sous-espèce *zebrina*), Malayan Highland (identifié par Shepherd dans une accession appartenant à la sous-espèce *truncata*) et East Africa (identifié par Shepherd chez une accession trouvée sur l'île de Pemba en Tanzanie). Les travaux de Shepherd ont concerné un nombre limité d'accessions, il se peut donc que d'autres grandes variations structurales existent au sein des *M. acuminata*.

On peut constater que trois des quatre translocations caractérisées à l'heure actuelle impliquent le chromosome 1, avec des points de translocation situés au niveau de différentes régions (0.65, 2.9 et 8.24 Mb pour les translocations 1/3, 1/4 et 1/9 respectivement). On peut noter également que les 4 translocations caractérisées à l'heure actuelle implique toujours un des deux chromosomes acrocentriques de M. *acuminata* (chromosome 1 et chromosome 2).

La localisation des points de réarrangements a pu être statistiquement associée à différentes caractéristiques structurales selon les génomes eucaryotes impliqués : un fort contenu en bases GC, une grande densité en gènes, l'origine de la réplication, la présence de séquences répétées, l'hypométhylation de l'ADN ou encore la mise en contact de chromatine faiblement condensée dans le noyau (Ma *et al.* 2006; Larkin *et al.* 2009; Gordon *et al.* 2009; Lemaitre *et al.* 2009; Drier *et al.* 2013; Berthelot *et al.* 2015). Ces pistes pourraient donc être investiguées chez le bananier.

Enfin, l'analyse génomique du point de translocation du chromosome 8 a permis de mettre en évidence le fort polymorphisme impliquant des séquences répétées de cette région entre les différents pôles de la diversité. Il est toutefois difficile à ce stade de déterminer si cette forte variabilité est une caractéristique fréquente au niveau des points de translocations chez le bananier ou un cas isolé.

B. Impact des translocations sur les ségrégations chromosomiques

Dans le cas de la translocation 1/4, Martin *et al.* (2017) avait montré une absence de recombinaison au niveau du fragment T1 et la transmission préférentielle des structures chromosomiques transloquées au détriment de la structure Standard dans la descendance d'un parent diploïde structurellement hétérozygote.

Concernant la recombinaison, nous avons également observé une absence de recombinaison au niveau du fragment T2 de la translocation 2/8 dans une descendance issus d'un parent structurellement hétérozygote par contre les fragment T1 et T9 des translocations 1/9 et 2/8 recombinent.

Deux hypothèses ont été avancées par Martin *et al.* (2017) pour expliquer l'absence de recombinaison au niveau des fragment T1 de la translocation 1/4 :

- La présence d'une région inversée au point de translocation, suggérée par Shepherd sur la base d'observations cytogénétiques. Cette inversion n'a pas pu être détectée à partir des données génomiques disponibles.
- Le non appariement du fragment T1, qui entraînerait de fait une absence totale de recombinaison. Cette hypothèse est à l'heure actuelle la plus probable.

Concernant l'absence de recombinaison au niveau du fragment T2 dans le cas de la translocation 2/8, une autre hypothèse, avec absence de recombinaison malgré un appariement correct, peut aussi être envisagée compte tenu de la très petite taille de segment. Le taux de recombinaison dans la population produite par Martin et al. est de 0,045 évènements / Mb. La probabilité d'avoir une recombinaison dans 1Mb suit une loi binomiale de paramètre p=0.045. D'après cette loi, la probabilité de ne retrouver "par hasard" aucune recombinaison dans une région de 3Mb (T1 de la translocation 1/4) chez 180 individus si ceux-ci sont indépendants est de 1.59x10⁻¹¹, soit extrêmement faible. Par contre, en estimant que le taux de recombinaison de la population ('Banksii' x 'Calcutta 4') x 'Malaccensis Nain' est similaire à celui de la population étudiée pour la translocation 1/4, alors la probabilité que le fragment T2 (240 kb) de la translocation 2/8 n'ait été impliqué "par hasard" dans aucune recombinaison chez les 73 individus BCM est de 0.446.

Cette hypothèse est donc solide dans le cas du fragment T2 de la translocation 2/8, mais très peu probable dans le cas du fragment T1 de la translocation 1/4.

Concernant la transmission préférentielle des structures chromosomiques transloquées au détriment de la structure Standard, l'analyse des translocations 1/9 et 2/8 renforcent ce constat dans les descendances issues d'un parent structurellement hétérozygote.

Ce phénomène a été observé pour les 3 translocations étudiées jusqu'à présent chez M. *acuminata*. La question qui se pose alors est pourquoi ces structures transloquées sont majoritairement transmises dans les descendances d'individus hétérozygotes ?

Plusieurs mécanismes pourraient induire une sélection gamétique ou zygotique en faveur des structures transloquées. Dans le cas de la translocation 1/4 du bananier, un mécanisme induisant une sélection gamétique a été proposée. Grâce aux observations de trivalents/univalents rapportées par Shepherd (1999), au caractère acrocentrique du chromosome 1 et à l'absence de recombinaisons le long de l'un des fragments transloqués (T1), Martin et al. (2017) émet l'hypothèse que dans le cas de la formation de bivalents, les chromosomes ne s'apparient pas au niveau des fragments transloqués (1T4 avec 1 et 4 avec 4T1). Dans le cas de la formation de trivalents/univalents, il n'y a qu'une configuration qui est produite (le trivalent 1T4/4/4T1) dont le seul gamète équilibré produit est celui contenant les chromosome 1T4 et 4T1. Cette hypothèse implique que l'univalent n'est pas transmis. In fine, en partant du principe que les gamètes déséquilibrés ne sont pas viables, seule la paire de chromosomes transloqués est transmise à l'issue de la résolution des trivalents. Ce mécanisme est envisageable dans le cas des translocations 1/9 et 2/8, mais il est pour l'instant compliqué d'aboutir à une hypothèse aussi précise. Les données rapportées par Shepherd (1999) sur la base de l'observation de 455 cellules en métaphase indiquent la présence de tétravalents (sous forme d'anneaux, 2% des méioses observées, et sous forme de chaînes, 18% des méioses observées) et de trivalents (36% des méioses observées). Cependant, les hybrides entre Northern 1 et Standard sont hétérozygotes pour deux translocations. Il est donc impossible à partir de ces données d'attribuer les multivalents à l'une ou l'autre des structures transloquées. Par ailleurs, la translocation entre les chromosomes 1 et 9 ne génère pas de suppression de recombinaison, ce qui suggère un mécanisme différent de celui supposé pour la translocation 1/4.

D'autres facteurs ont été rapportés chez les plantes pour expliquer la transmission préférentielle de certaines structures chromosomiques. Chez le maïs, un facteur mécanistique invoque une version anormale du chromosome 10 (Ab10) qui présente une activité néocentromérique due à la présence d'un knob plus important que celui présent chez le chromosome 10 normal. Les knobs sont des répétitions en tandem de plusieurs milliers à plusieurs millions de séquences de 180 à 350 paires de bases et représentent jusqu'à 8% du génome du maïs (Peacock *et al.* 1981). Les plus importants d'entre eux sont visibles au microscope. Dans une population hétérozygote pour le chromosome 10 (Ab10/10), Ab10 produit une distorsion de ségrégation en sa propre faveur, de l'ordre de 70% au lieu de 50% (Buckler *et al.* 1999). La présence de knobs chez le bananier n'a pas été documentée jusqu'à présent, mais c'est une piste qui pourrait être investiguée.

D'autre facteurs impliquant des différences de 'fitness', soit au niveau des gamètes soit au niveau des descendants, peuvent être impliqués. C'est le cas au sein de l'espèce *Mimulus* (Fishman *et al.* 2001), où les gamètes mâles possédant les structures chromosomiques transloquées ont un tube pollinique se développant plus rapidement que les autres gamètes. Elles sont donc plus compétitives et aboutissent à une surreprésentation des structures transloquées dans les zygotes.

Enfin, concernant les translocations 2/8 et 1/9, les hybrides étudiés issus du croisement ('Banksii' x 'Calcutta 4') x 'Malaccensis Nain' impliquent plusieurs pôles de la diversité. Cette diversité importante pourrait générer des incompatibilité d'hybrides à certains locus qui donc favoriserait certaines combinaisons chromosomiques. Des gènes gamétocides ou des gènes "Dobzhansky-Muller" ont été évoqué dans des cas d'incompatibilités d'hybrides (Maheshwari et barbash, 2011).

Il est important de noter que nos descendances ont été obtenues après une étape de sauvetage d'embryons, les proportions de combinaisons chromosomiques que l'on observe sont donc potentiellement différentes de celles qui auraient été observées dans la nature pour un tel croisement.



Figure 27 - Représentation des événements chromosomiques dans la diversité géographiques des bananiers.

Finalement, quels que soient les facteurs à l'origine de ces transmissions préférentielles, on peut imaginer que des translocations réciproques apparaissent de temps en temps mais que seules celles transmises préférentiellement aux cours des générations peuvent se répandre ou remplacer à terme la structure ancestrale présente dans la population.

C. Les variations structurales et la structuration de la diversité

Les groupes caryotypiques établis à partir des translocations se superposent en grande partie à la structuration en sous espèces et espèces (Figure 27). La translocation 1/3 a été détectée chez toutes les accessions M. balbisiana testées jusque-là et exclusivement chez cette espèce balbisiana. La translocation 1/4 est présente dans un sous-groupe de la sousespèce malaccensis. Cependant, étant préférentiellement transmise en cas d'hybridation avec une accession portant la structure Standard non transloquée (présente également dans cette sous-espèce), on peut imaginer qu'à terme, la translocation 1/4 devienne la structure majoritaire dans la sous-espèce *malaccensis*. Ce scénario d'émergence puis de "conquête" d'une translocation réciproque sur un groupe génétique peut s'appliquer aux translocations 1/9 et 2/8, que mes résultats ont montré être omniprésentes et exclusives au groupe génétique burmannica (dans la mesure des accessions testées). Selon cette hypothèse, deux translocations réciproques ont émergé au sein du groupe burmannica, ancestralement certainement porteur de la structure chromosomique Standard puis progressivement remplacé la structure ancestrale. Il semblerait donc que l'espèce Musa se soit structurée en sous-espèces notamment en raison d'un isolement géographique, et que les variations structurales aient émergées a posteriori, renforçant cette structuration par un isolement reproductif.

D. Liens avec les programmes d'amélioration et les recherches de QTLs

L'accession 'Calcutta 4' a été utilisée par les programmes d'amélioration dans le cadre de la stratégie d'amélioration d'un cultivar triploïde par croisement avec une accession

diploïde. 'Calcutta 4' possède des atouts qui en ont fait une accession diploïde potentiellement intéressante pour l'amélioration variétale : ses nombreuses résistances, sa fertilité et la morphologie de son rachis et de son régime (rachis vertical, alors que de nombreux bananiers sauvages ont un rachis horizontal, et régime peu compact et dégagé de la cime de la plante). Cette accession a été utilisée par Shepherd et Simmonds pour tenter d'améliorer 'Gros Michel' ou encore par le programme du CIRAD au Cameroun pour améliorer les cultivars de type 'Plantain' (Christophe Jenny, com. pers.). Au CIRAD, le passage à la stratégie dite de "reconstruction" a entraîné l'abandon de cette accession et de ses voisines du groupe *Burmannica* à l'avantage des contributeurs probables cultivars du groupe 'Cavendish' et du groupe Plantain. Dans l'optique de diversifier les cultivars utilisés pour la production de bananes dessert, les accessions du groupe *Burmannica* pourrait être utilisée dans le cadre du "pré-breeding", c'est à dire pour améliorer une des accessions diploïdes utilisées ensuite dans le croisement utilisant la stratégie de reconstruction.

L'impact des variations structurales sur la ségrégation et la recombinaison dans les descendances hétérozygotes de structure est une information précieuse pour la recherche en génétique quantitative et les programmes d'amélioration. Deux aspects de la détection de loci de trait quantitatif peuvent être impactés : la précision de la détection de QTL et la robustesse des tests statistiques.

Concernant la précision de la détection de QTL:

Nous avons montré qu'au niveau des points de translocation la recombinaison est réduite. D'autre part, dans certain cas on observe une réduction drastique des recombinaisons sur un des fragments impliqués dans la translocation. Les recombinaisons produisent à l'échelle d'une descendance une fragmentation fine des génotypes en fonction des allèles parentaux. Réciproquement, l'absence de recombinaisons le long d'un segment chromosomique génère à l'échelle d'une descendance une homogénéité des génotypes pour ce segment, provoquant des difficultés à ordonner les marqueurs. Lorsqu'un QTL est détecté dans une région, ce sont les recombinaisons qui permettent de réduire la taille du locus concerné et d'apporter une haute définition à l'analyse. Ainsi, pour éviter la présence de régions non recombinantes (comme la région T1 de la translocation 1/4 ou l'inversion 5) lors d'analyses pour affiner la position de QTLs, il faudra préférer l'utilisation

d'accessions homozygotes de structures au niveau de ces QTLs. Enfin, générer des populations plus grandes pourra permettre de pallier à la baisse de recombinaison observée autour des points de translocation. Cependant, la génération de grandes populations est très laborieuse chez le bananier étant donné la faible fertilité et la nécessité de recourir au sauvetage d'embryons.

D'autre part, la transmission préférentielle de certaines structures dans les populations issues de parents hétérozygotes de structure peut générer des co-ségrégations entre chromosomes ou fragments de chromosomes. C'est ce qui a été observé pour les translocations 1/3, 1/4 et 1/9. Ces co-ségrégations peuvent générer des pseudo-liaisons lors de la réalisation de cartes génétiques. Ces limitations à la génération de cartes génétiques précises auront pour conséquence de gêner la localisation fiable et précise des QTLs. Là encore, limiter l'utilisation d'accessions hétérozygotes pourra permettre de contourner ce problème.

Concernant la robustesse des tests statistiques :

La significativité des QTL est évaluée à l'aide de tests statistiques (ANOVA, scores de vraisemblance...). La robustesse de ces tests est basée notamment sur la taille des classes. Les distorsions de ségrégation peuvent impacter la taille de ces classes, en déséquilibrant les ratios alléliques. Toutes les variations structurales détectées chez le bananier génèrent des distorsions de ségrégation. La recherche de QTL dans des croisements impliquant des parents de structure chromosomique différente pourra impliquer de produire des populations plus grandes pour pallier à ce problème. Or chez le bananier la production large descendance est très laborieuse.

Les programmes d'amélioration peuvent eux aussi être impactés par les conséquences des réarrangements chromosomiques sur la recombinaison et les distorsions de ségrégation dans les populations hétérozygotes de structure. En effet, la baisse du taux de recombinaison (aux bornes des points de cassure par exemple) et a fortiori la suppression des recombinaisons (le long du segment T1 dans le cas de la translocation 1/4 ou le long des segments inversés) peut limiter la possibilité de combiner différents caractères d'intérêts entre plusieurs géniteurs, ou au contraire servir à "protéger" une combinaison favorable de caractères.

Les distorsions de ségrégation peuvent quant à elles biaiser la transmission d'allèles d'intérêt en leur faveur ou défaveur. Dans le cas où les ségrégations sont en défaveur de l'allèle d'intérêt, il faudra générer une descendance plus importante qu'habituellement pour obtenir un nombre d'individu portant l'allèle d'intérêt satisfaisant, et se confronter à nouveau à la difficulté de produire des grandes populations chez le bananier.

Les travaux de recherche sur les gènes impliqués dans les caractère d'intérêt agronomique sont encore extrêmement limités chez le bananier. Le CIRAD a entrepris le génotypage de plusieurs millier d'accessions générées dans le cadre du programme d'amélioration pour réaliser de telle analyses. La caractérisation des translocations présentes dans les populations étudiées et de leur impact sur les ségrégations devrait permettre d'anticiper les freins que peuvent représenter ces variations structurales pour l'identification et la cartographie de QTL et de proposer des solutions pour y palier. Ces travaux devraient également aider à orienter les plans de croisements pour bénéficier ou contourner la réduction des recombinaisons liés aux translocations.

D'autre part, les hétérozygoties de structure sont l'un des facteurs engendrant des faibles fertilités chez le bananier. Ces faibles fertilités sont une des principales contraintes des programmes d'amélioration du bananier. Mieux comprendre leur origine pourrait permettre de prédire les types hybrides qui seront les plus affectés et en tenir compte dans les schémas d'amélioration.

BIBLIOGRAPHIE

Arinaitwe IK, Teo CH, Kayat F, et al. 2019. Evaluation of banana germplasm and genetic analysis of an F1 population for resistance to Fusarium oxysporum f. sp. cubense race 1. *Euphytica* **215**: 175.

Baird NA, Etter PD, Atwood TS, et al. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers (JC Fay, Ed.). *PLoS ONE* **3**: e3376.

Bakry F. 2008. Zygotic embryo rescue in bananas. Fruits 63: 111–115.

Bakry F, Carreel F, Jenny C, Horry J-P. **2009**. Genetic Improvement of Banana In: Jain SM, Priyadarshan PM, eds. *Breeding Plantation Tree Crops: Tropical Species*. New York, NY: Springer, 3–50.

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

Beadle GW. **1932**. A Possible Influence of the Spindle Fibre on Crossing-Over in Drosophila. *Proceedings of the National Academy of Sciences* **18**: 160–165.

Belser C, Istace B, Denis E, et al. 2018. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants* **4**: 879.

Berthelot C, Muffato M, Abecassis J, Roest Crollius H. **2015**. The 3D Organization of Chromatin Explains Evolutionary Fragile Genomic Regions. *Cell Reports* **10**: 1913–1924.

Bikard D, Patel D, Le Metté C, et al. 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within A. thaliana. *Science* **323**: 623–626.

Bodénès C, Chancerel E, Ehrenmann F, Kremer A, Plomion C. **2016**. High-density linkage mapping and distribution of segregation distortion regions in the oak genome. *DNA Research* **23**: 115–124.

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.

Buckler ES, Phelps-Durr TL, Buckler CSK, Dawe RK, Doebley JF, Holtsford TP. **1999**. Meiotic Drive of Chromosomal Knobs Reshaped the Maize Genome. *Genetics* **153**: 415–426.

Cameron DR, Moav RM. **1957**. Inheritance in Nicotiana Tabacum Xxvii. Pollen Killer, an Alien Genetic Locus Inducing Abortion of Microspores Not Carrying It. *Genetics* **42**: 326–335.

Carreel F, Fauré S, De Leon DG, et al. 1994. Evaluation de la diversité génétique chez les bananiers diploïdes (Musa sp). *Genetics Selection Evolution* **26**: S125.

Carreel F, de Leon DG, Lagoda P, et al. 2002. Ascertaining maternal and paternal lineage within Musa by chloroplast and mitochondrial DNA RFLP analyses. *Genome* **45**: 679–692.

Champion J. 1963. Le Bananier. Maisonneuve & Larose.

Chen K, Wallis JW, McLellan MD, et al. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods* **6**: 677–681.

Christelová P, De Langhe E, Hřibová E, 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**: 801–824.

Christelová P, Valárik M, Hřibová E, et al. 2011. A platform for efficient genotyping in Musa using microsatellite markers. *AoB PLANTS* **2011**.

Cloutier S, Cappadocia M, Landry BS. **1997**. Analysis of RFLP mapping inaccuracy in Brassica napus L. *Theoretical and Applied Genetics* **95**: 83–91.

Cretin L, Triquenot A. **2018**. Apports de produits phytopharmaceutiques en arboriculture : nombre de traitements et indicateur de fréquence de traitements. Campagne agricoles 2015 et 2012.

Daniells J, Jenny C, Karamura D, Tomekpe K. **2001**. *Musalogue : a catalogue of #Musa# gemplasm. Diversity in the genus #Musa#*. Montpellier: INIBAP.

De Langhe E, Hřibová E, Carpentier S, Doležel J, Swennen R. **2010**. Did backcrossing contribute to the origin of hybrid edible bananas? *Annals of Botany* **106**: 849–857.

Dessauw D. 1987. Etude des facteurs de la stérilité du bananier (*Musa* spp.) et des relations cytotaxinomiques entre *M. acuminata* Colla et *M. balbisiana* Colla.

D'Hont A, Denoeud F, Aury J-M, *et al.* 2012. The banana *Musa acuminata* genome and the evolution of monocotyledonous plants. *Nature* **488**: 213–219.

Dobzhansky Th. **1937**. Genetic Nature of Species Differences. *The American Naturalist* **71**: 404–420.

Dobzhansky Th, Sturtevant AH. 1938. Inversions in the Chromosomes of Drosophila Pseudoobscura. *Genetics* **23**: 28–64.

Dodds KS. **1943**. Genetical and cytological studies of *Musa*. V. Certain edible diploids. *Journal of Genetics* **45**: 113–138.

Dodds KS, Simmonds NW. **1948**. Sterility and parthenocarpy in diploid hybrids of *musa*. *Heredity* **2**: 101–117.

Dohm JC, Lottaz C, Borodina T, Himmelbauer H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research* **36**: e105.

Drier Y, Lawrence MS, Carter SL, et al. 2013. Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Research* **23**: 228–235.

Dupouy M, Baurens F-C, Derouault P, et al. 2019. Two large reciprocal translocations characterized in the disease resistance-rich burmannica genetic group of Musa acuminata. *Annals of Botany*: mcz078.

Echt CS, Kidwell KK, Knapp SJ, Osborn TC, McCoy TJ. 1994. Linkage mapping in diploid alfalfa (Medicago sativa). *Genome* **37**: 61–71.

Fauré S, Bakry F, Gonzalez de Leon D. 1992. Cytogenetic studies of diploid bananas In: Ganry J, ed. *Breeding bananas and plantains : proceedings of an International Symposium on Genetic Improvement of Bananas for their Resistance to diseases and Pests.* Montpellier (France): CIRAD-FLHOR, 77–92.

Fauré S, Noyer JL, Horry JP, Bakry F, Lanaud C, Gońzalez de León D. 1993. A molecular marker-based linkage map of diploid bananas (Musa acuminata). *Theoretical and Applied Genetics* 87: 517–526.

Feuk L, Marshall CR, Wintle RF, Scherer SW. **2006**. Structural variants: changing the landscape of chromosomes and design of disease studies. *Human Molecular Genetics* **15**: R57–R66.

Fishman L, Kelly AJ, Morgan E, Willis JH. **2001**. A Genetic Map in the Mimulus guttatus Species Complex Reveals Transmission Ratio Distortion due to Heterospecific Interactions. *Genetics* **159**: 1701–1716.

Garsmeur O, Droc G, Antonise R, et al. 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications* **9**: 2638.

Gordon JL, Byrne KP, Wolfe KH. 2009. Additions, Losses, and Rearrangements on the Evolutionary Route from a Reconstructed Ancestor to the Modern Saccharomyces cerevisiae Genome. *PLOS Genetics* **5**: e1000485.

Graner A, Jahoor A, Schondelmaier J, et al. 1991. Construction of an RFLP map of barley. *Theoretical and Applied Genetics* **83**: 250–256.

Guan P, Sung W-K. **2016**. Structural variation detection using next-generation sequencing data: A comparative technical review. *Methods* **102**: 36–49.

Häkkinen M. 2013. Reappraisal of sectional taxonomy in Musa (Musaceae). TAXON 62: 809–813.

Haldane JBS. **1919**. The combination of linkage values and the calculation of distance between the loci fo linked factors. *Journal of Genetics* **8**: 299–309.

Hassedine Diouf FB, Mergeai G. 2012. Distorsions de ségrégation et amélioration génétique des plantes (synthèse bibliographique). BASE.

He L, Dooner HK. 2009. Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for Helitron and retrotransposon insertions. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 8410–8416.

Herten K, Hestand MS, Vermeesch JR, Van Houdt JK. 2015. GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinformatics* 16: 73.

Hippolyte I, Bakry F, Seguin M, et al. 2010. A saturated SSR/DArT linkage map of Musa acuminata addressing genome rearrangements among bananas. BMC Plant Biology 10: 65.

Hippolyte I, Jenny C, Gardes L, et al. 2012. Foundation characteristics of edible Musa triploids revealed from allelic distribution of SSR markers. *Annals of Botany* **109**: 937–951.

Janssens SB, Vandelook F, Langhe ED, 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. *New Phytologist* **210**: 1453–1465.

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

Jeridi M, Bakry F, Escoute J, et al. 2011. Homoeologous chromosome pairing between the A and B genomes of Musa spp. revealed by genomic in situ hybridization. *Annals of Botany*: mcr207.

Jones DR. 2000. Diseases of banana, abaca and enset. Wallingford: CABI Publishing.

Kaló P, Endre G, Zimányi L, Csanádi G, Kiss GB. 2000. Construction of an improved linkage map of diploid alfalfa (Medicago sativa). *Theoretical and Applied Genetics* **100**: 641–657.

Kosambi DD. **1943**. The Estimation of Map Distances from Recombination Values. *Annals of Eugenics* **12**: 172–175.

Ky CL, Barre P, Lorieux M, et al. 2000. Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (Coffea sp.). *Theoretical and Applied Genetics* **101**: 669–676.

Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA. 2009. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Research* **19**: 770–777.

Lashermes P, Combes MC, Prakash NS, Trouslot P, Lorieux M, Charrier A. 2001. Genetic linkage map of Coffea canephora: effect of segregation distortion and analysis of recombination rate in male and female meioses. *Genome* 44: 589–595.

Lee S-I, Kim N-S. 2014. Transposable Elements and Genome Size Variations in Plants. *Genomics & Informatics* 12: 87–97.

Lemaitre C, Zaghloul L, Sagot M-F, et al. 2009. Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. *BMC Genomics* **10**: 335.

Lescot T. 2014. Genetic diversity of the banana. *Fruitrop (English ed.)*.

Li L-F, Häkkinen 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.

Lindholm AK, Dyer KA, Firman RC, et al. 2016. The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends in Ecology & Evolution* **31**: 315–326.

Luro F, Lorieux M, Laigret F, Bové JM, Ollitrault P. **1995**. Cartographie du génome des agrumes à l'aide des marqueurs moléculaires et distorsions de ségrégation.

Lynch M, Force AG. **2000**. The Origin of Interspecific Genomic Incompatibility via Gene Duplication. *The American Naturalist* **156**: 590–605.

Lyttle TW. 1991. Segregation Distorters. Annual Review of Genetics 25: 511–581.

Ma J, Zhang L, Suh BB, et al. 2006. Reconstructing contiguous regions of an ancestral genome. *Genome Research* 16: 1557–1565.

Maheshwari S, Barbash DA. 2011. The Genetics of Hybrid Incompatibilities. *Annual Review of Genetics* 45: 331–355.

Mangelsdorf PC, Jones DF. 1926. The Expression of Mendelian Factors in the Gametophyte of Maize. *Genetics* **11**: 423–455.

Martin M. **2011**. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.

Martin GE. **2014**. Caractérisation des différences de structures chromosomiques dans l'espèce Musa acuminata par re-séquençage NGS : le cas de l'accession "Pahan."

Martin G, Baurens F-C, Droc G, et al. 2016. Improvement of the banana "Musa acuminata" reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genomics* 17: 243.

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

Mbanjo EGN, Tchoumbougnang F, Mouelle AS, et al. 2012. Molecular marker-based genetic linkage map of a diploid banana population (Musa acuminata Colla). *Euphytica* **188**: 369–386.

Morgan AP, Gatti DM, Najarian ML, *et al.* 2017. Structural Variation Shapes the Landscape of Recombination in Mouse. *Genetics* 206: 603–619.

Muller H. 1942. Isolating mechanisms, evolution, and temperature In: 71–125.

Nakagahara M. 1972. Genetic mechanism on the distorted segregation of marker genes belonging to the eleventh linkage group in cultivated rice. *Japanese Journal of Breeding* **22**: 232–238.

Narzisi G, O'Rawe JA, Iossifov I, et al. 2014. Accurate *de novo* and transmitted indel detection in exome-capture data using microassembly. *Nature Methods* 11: 1033–1036.

Niranjana M. **2017**. Gametocidal genes of *Aegilops* : segregation distorters in wheat– *Aegilops* wide hybridization (M Francki, Ed.). *Genome* **60**: 639–647.

Noumbissié GB, Chabannes M, Bakry F, et al. 2016. Chromosome segregation in an allotetraploid banana hybrid (AAAB) suggests a translocation between the A and B genomes and results in eBSV-free offsprings. *Molecular Breeding* **36**: 1–14.

Noyer J-L, Dambier D, Lanaud C, Lagoda P. 1997. The saturated map of diploid banana (Musa acuminata). https://agritrop.cirad.fr/389836/. 31 Oct. 2019.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD. **1988**. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**: 721–726.

Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ. 1981. Highly repeated DNA sequence limited to knob heterochromatin in maize. *Proceedings of the National Academy of Sciences* **78**: 4490–4494.

Pereira MG, Lee M, Bramel-Cox P, Woodman W, Doebley J, Whitkus R. 1994. Construction of an RFLP map in sorghum and comparative mapping in maize. *Genome* **37**: 236–243.

Perrier X. **2009**. Combining Biological Approaches to Shed Light on the Evolution of Edible Bananas. *Ethnobotany Research and Applications* **7**: 199.

Perrier X, De Langhe E, Donohue M, et al. 2011. Multidisciplinary perspectives on banana (Musa spp.) domestication. *Proceedings of the National Academy of Sciences* **108**: 11311–11318.

Ploetz RC. **1994**. Panama disease : return of the first banana menace. *International Journal of Pest Management* **40**: 326–336.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L. 2012. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE* **7**: e32253.

Quénéhervé P, Valette C, Topart P, Tezenas du Montcel H, Salmon F. 2009. Nematode resistance in bananas: screening results on some wild and cultivated accessions of Musa spp. *Euphytica* **165**: 123–136.

Quillet MC, Madjidian N, Griveau Y, et al. 1995. Mapping genetic factors controlling pollen viability in an interspecific cross in Helianthus sect. Helianthus. *Theoretical and Applied Genetics* **91**: 1195–1202.

Raboin L-M, Carreel F, Noyer J-L, et al. 2005. Diploid Ancestors of Triploid Export Banana Cultivars: Molecular Identification of 2n Restitution Gamete Donors and n Gamete Donors. *Molecular Breeding* **16**: 333–341.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28: i333–i339.

Rieseberg LH. **2001**. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution* **16**: 351–358.

Risterucci A-M, Hippolyte I, Perrier X, et al. 2009. Development and assessment of Diversity Arrays Technology for high-throughput DNA analyses in Musa. *Theoretical and Applied Genetics* **119**: 1093–1103.

Robertson WRB. **1916**. Chromosome studies. I. Taxonomic relationships shown in the chromosomes of tettigidae and acrididae: V-shaped chromosomes and their significance in acrididae, locustidae, and gryllidae: Chromosomes and variation. *Journal of Morphology* **27**: 179–331.

Rouard M, Droc G, Martin G, et al. 2018. Three New Genome Assemblies Support a Rapid Radiation in Musa acuminata (Wild Banana). *Genome Biology and Evolution* **10**: 3129–3140.

Sandler L, Novitski E. **1957**. Meiotic Drive as an Evolutionary Force. *The American Naturalist* **91**: 105–110.

Sano Y. 1990. The genic nature of gamete eliminator in rice. *Genetics* 125: 183–191.

Sardos J, Rouard M, Hueber Y, 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 (KM Olsen, Ed.). *PLOS ONE* **11**: e0154448.

Schröder J, Hsu A, Boyle SE, et al. 2014. Socrates: identification of genomic rearrangements in tumour genomes by re-aligning soft clipped reads. *Bioinformatics* 30: 1064–1072.

Shepherd K. 1999. Cytogenetics of the genus Musa. INIBAP.

Simmonds NW. 1962. The evolution of the bananas. The Evolution of the Bananas.

Simmonds NW, Shepherd K. 1955. The taxonomy and origins of the cultivated bananas. *Botanical Journal of the Linnean Society* **55**: 302–312.

Sweigart AL, Willis JH. **2012**. Molecular evolution and genetics of postzygotic reproductive isolation in plants. *F1000 Biology Reports* **4**.

Tadmor Y, Zamir D, Ladizinsky G. **1987**. Genetic mapping of an ancient translocation in the genus Lens. *Theoretical and applied genetics* **73**: 883–892.

Talbert PB, Henikoff S. 2010. Centromeres Convert but Don't Cross. PLoS Biology 8: e1000326.

Vilarinhos AD. 2004. Cartographie génétique et cytogénétique chez le bananier : caractérisation des translocations.

Vuylsteke DR, Swennen RL, Ortiz R. 1992. Development and performance of balck sigatoka-resistant tetraploid hybrids of plantain (Musa spp., AAB group). *Euphytica* **65**: 33–42.

Wang Z, Miao H, Liu J, et al. 2019. Musa balbisiana genome reveals subgenome evolution and functional divergence. *Nature Plants* **5**: 810–821.

White F. 1962. *Geographic variation and speciation in Africa with particular reference to Diospyros , Ebenaceae*. https://eurekamag.com/research/024/746/024746530.php. 31 Oct. 2019.

Wong C, Kiew R, Argent G, Set O, Lee SK, Gan YY. 2002. Assessment of the Validity of the Sections in Musa (Musaceae) using AFLP. *Annals of Botany* **90**: 231–238.

Yang L, Luquette LJ, Gehlenborg N, *et al.* 2013. Diverse Mechanisms of Somatic Structural Variations in Human Cancer Genomes. *Cell* 153: 919–929.

Zeitouni B, Boeva V, Janoueix-Lerosey I, et al. 2010. SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics* **26**: 1895–1896.

ANNEXES

		Chromosome 2 and 8 structures										C	Chromosome 1 and 9 structure				structur	es						
		M sec	Mate-pair Signature Segment Junctions							M sec	ate-p Jueno	air ing	Si So Ju	gnat egm incti	ure ent ons									
		5/08		8/8T2		Paha	ng HD	I	Bar	ksii	Maia'C	Da		Calc	utta 4	Ļ	ure	60/1		-9/9T1	Ca	lcut	ta 4	ure
	pecies	gous 02	ygous	gous 21	C	h2	C	h8	C	18	Ch8		Ch	2T8	Ch	8T2	d struct	gous 01	ygous	gous 11	Chr1	r9	Chr9T1	d struct
Name	group/s.	Homozy	Heteroz	Homory	a-b	p-c	d-e	e-f	d-e	e-f	d-e	e-t	а-х	x-f	d-y	y-c	Deduced	Homozy	Heteroz	Homory	k1-k3	k3-h	h-k1	Deduced
Banksii (CRB)	M. a. ssp. banksii	1	0	0	115	105			126	149							ST	1	0	0				ST
Hawain 2	M. a. ssp. banksii				148	179			125	190							ST							?
M. a. ssp. banksii ITC0885	M. a. ssp. banksii				128	148	40		146	180							ST							?
M. a. ssp. banksii ITCO897	M. a. ssp. banksii				62	99			68	117							ST							?
Waigu	M. a. ssp. banksii				135	82			130	125							ST							?
Higa (ITC0464)	M. a. ssp. banksii				115	168			145	117							ST							?
Long Tavoy	M. a. ssp. burmannica												53	55	99	23	N1				154			N1
Pisang Prentel	M. a. ssp. burmannica												63	69	78	31	N1				182			N1
Calcutta 4	M. a. ssp. burmannicoïdes	0	0	1									57	61	63	24	N1	0	0	1	97		11	N1
Agutay (ITC1028)	M. a. ssp. errans				155	135			103	157							ST							?
Malaccensis nain	M. a. ssp. malaccensis	1	0	0	103	37					122						ST	1	0	0				ST
Pahang	M. a. ssp. malaccensis	1	0	0	87	106	105	132	34								ST	1	0	0				ST
Pahang (CRB)	M. a. ssp. malaccensis				67	54	28	27	48	59							ST							?
Pisang Karok 391	M. a. ssp. malaccensis				58	129		18	38	63							ST							?
Pisang Kra	M. a. ssp. malaccensis				44	90		19	49	58							ST							?
Pisang serun 400	M. a. ssp. malaccensis				59												?							?
Pisang serun 404	M. a. ssp. malaccensis				76	52		26	54	69	42 3	32					ST							?

Annexe 1 - Supplementary table 1 - Accessions tested for their structure on chromosomes 2, 8, 1 and 9.

Sélangor	M. a. ssp. malaccensis				117	80		28			108	65					ST					?
Bornéo	M. a. ssp. microcarpa	1	0	0	80	70			74	89							ST	1	0	0		ST
EN 13 IDN075	M. a. ssp. microcarpa				89	97	39	12			78	34					ST					?
Microcarpa	M. a. ssp. microcarpa				93	63			48	60	42	51					ST					?
AAs-IDN113	M. a. ssp. microcarpa der.				56	98					67	82					ST					?
Khae Phrae	M. a. ssp. siamea												74	95	92	19	N1				142	N1
Pa Rayong	M. a. ssp. siamea												38	37	108	21	N1				102	N1
Sumatrana	M. a. ssp. sumatrana				122	94		15									ST					?
Truncata	M. a. ssp. truncata				61	88		4									ST					?
Buitenzorg	M. a. ssp. zebrina				117	100		23	48	55	54	35					ST					?
Cici Brésil	M. a. ssp. zebrina				104	83		21			94	96					ST					?
Maia Oa	M. a. ssp. zebrina	1	0	0	120	115					110	132					ST	1	0	0		ST
Monyet	M. a. ssp. zebrina				87	85	56				61	68					ST					?
Pisang Cici Alas	M. a. ssp. zebrina				71	57		30			84	72					ST					?
Pa Songkhla	M. a. ssp. ²				81	82	110	119	22								ST					?
Gulum	AAcv				106	125			110	137							ST					?
Guyod	AAcv				106	109			105	129							ST					?
Kole	AAcv				136	136		36		67	85	69					ST					?
Malaysian Blood	AAcv				122	135	79	37			135	158					ST					?
Pa (Patthalong)	AAcv				59	52											ST					?
Pallen Berry	AAcv				190	97		46	93	68							ST					?
Pisang Pipit	AAcv				73	70	54	30		25	71	32					ST					?
Pisang Sapon	AAcv				74	111	33		34	43	39	38					ST					?
Pisang Tongat	AAcv				84	50			93	96							ST					?
Racadag	AAcv				172	128	81		63	95	80	49					ST					?
Sinwobogi	AAcv				140				140								?					?
Tomolo	AAcv				96	95	38		105	162							ST					?

Veinte Cohol	AAcv				96	107			50	68		38					ST				?
Pisang Berlin	AAcv				95	58		17	121	106							ST				?
Colatina Ouro	AAcv				85	74			54	55	60	41					ST				?
Gwanhour	AAcv				87	90			44	65	36	63					ST				?
Niyarma Yik	AAcv				165	160	69		66	84	79	78					ST				?
Sowmuk	AAcv				119	88	80		81	86	78	81					ST				?
Bebek	AAcv				144	87	38		130	175							ST				?
Beram	AAcv				88	75			49	80							ST				?
Heva	AAcv				125	103	90			52	71	65					ST				?
Pisang Jari Buaya	AAcv				94	84	61	9	56	48							ST				?
IDN 077	AAcv				61	67	65	54	54	48							ST				?
Kirun	AAcv				98	83		18	49	71	58						ST				?
N°110/THA052	AAcv				133	47					66	46					ST				?
Pisang Jaran	AAcv				170	104	104	42			153	172					ST				?
Pisang Mas	AAcv				128	68			47	75	67	39					ST				?
Dibit	AAcv				95	66					23	67					ST				?
Galeo	AAcv	1	0	0	111	91			117	159							ST	1	0	0	ST
SF 265	AAcv				228	182			196	240							ST				?
Katual no.2	AAcv				88	100			96	121							ST				?
Papat	AAcv				148	160			142	146							ST				?
Hom	AAcv				79	29			47	66				50	66	12	ST+N1				?
IDN 110	AAcv	1	0	0	67	50	38	12	35	45	42						ST	1	0	0	ST
Khai Nai on	AAcv				100	53			68	98							ST				?
Manang	AAcv	0	1	0	134	48				103			97	49			ST+N1				?
Pisang Bangkahulu	AAcv				75	33			40	66							ST				?
Pisang Buntal	AAcv				160	106			48	121				2			ST				?
Thong Det	AAcv				83	47	65	64	77	64							ST				?

Tjau Lagada	AAcv					48				79	91		ST				?
Kumburgh	AAcv				102	102			47	61	33	3	ST				?
Pisang Madu	AAcv	1	0	0	118	64		6		37			ST	1	0	0	ST
Akondro mainty	AAcv	1	0	0	92	119			96	128			ST	1	0	0	ST
Aivip	AAcv				76	102			130	121			ST				?
Sena	AAcv				102				100	149			ST				?
Spiral	AAcv				131	137	49		138	141			ST				?
Manameg Red	AAcv				92	100			77	100			ST				?
Vudo Beo	AAcv				149	168			86	101	50	0	ST				?
Sihir	AAcv				83	80			83	95			ST				?
Pisang Lilin	AAcv	1	0	0	129	50		34		39	69		ST	1	0	0	ST
Paka	AAcv	1	0	0	118	79			30	48	75		ST	1	0	0	ST
Pisang Ambon	AAAcv				69	78		24	25	51	44 23	1	ST				?
Hom (Thong Mokho)	AAAcv				116	97			59	136			ST				?
Grande Naine	AAAcv	1	0	0	130	140			90	155			ST	1	0	0	ST
Која	AAAcv				109	133		31	49	41	86 43	3	ST				?
Gros Michel	AAAcv				160	116			135	166			ST				?
Intokatoke	AAAcv				128	154	96	38	61	37	115 13	33	ST				?
Hom (Sakhon Nakhon)	AAAcv				100	94	65	49	90	73			ST				?
Pisang Kayu	AAAcv				81	102			55	95	33	1	ST				?
Figue rose	AAAcv				86	55			31	73	32 37	7	ST				?
Figue rose naine	AAAcv				85	65			29	58	36		ST				?
Pisang Papan	AAAcv				117	71			35	36	57 36	6	ST				?
Merik	AAAcv				118	114	44		69	84	33	3	ST				?
Palang	AAAcv				151	144	35		154	174			ST				?
Wh-o-Gu	AAAcv				88	113			14	33	52 66	6	ST				?
Mnalouki	AABcv				83	61			77	104			ST				?

Luba	AABcv		81	85			31	39	29	ST		?
Laknao	AABcv		69	46			31	68		ST		?
Lady Finger	AABcv		66	44	27	20	64	80		ST		?
Pisang Rajah	AABcv		72	56		6	55	69		ST		?
Pisang Kelat	AABcv		84	59		3	33	21	18 28	ST		?
Mkono Wa Tembo	AABcv		94	75			80	78		ST		?
Prata Ana	AABcv		158	129		18	142	148		ST		?
Maia Maoli	AABcv		100	58			57	73		ST		?
Popoulou CMR	AABcv		76	52			53	76		ST		?
Kingala 1	AABcv		105	82		20	70	77		ST		?
Poteau Géant	ABBcv		79	27			25	39		ST		?
Monthan	ABBcv		87	38	33			33		ST		?
Ice Cream	ABBcv			41			25	39		ST		?
Pelipita	ABBcv		105	25			74	68		ST		?
Namwa Khom	ABBcv		108	26			42			ST		?
Red Yade	ABBcv		218	18			72	85		ST		?
Saba	ABBcv		96	77	17		63	101		ST		?
Kunnan	ABcv		58	40			27	37		ST		?
Safet Velchi	ABcv		39	47	11		45	42		ST		?
Tonton Kepa	AScv		127	187		3	69	69		ST		?
Wompa	AScv		135	136	98				29 64	ST		?
Butuhan	Musa balbisiana		102	12						ST1		ST1
Pisang Klutuk Wulung	Musa balbisiana		71	14						ST1		ST1
Schizocarpa	Musa schizocarpa		131	115						ST1		ST1
Schizocarpa	Musa schizocarpa		140	167	78	4				ST		ST1

Annexe 2 - Supplementary Table 2 - Genomic coordinates of compared sequences.

	'DH-Pahang' reference genome	'Calcutta 4' assembly
Translocation 2/8	chr02:29268836-29274837	utg154:2411918-2417918
	chr08:37722146-37728147	utg170:2941233-2947233
Translocation 1/9	chr01:8219188-8289189	utg94:4743017-4813018
	chr09:11554857-11624857	utg195:5581660-5651463

Annexe 3 - Supplementary Table 3 - Genomic position of the SSJs for the translocation 2/8.

	Chromo	osome 2	Chromosome 8							
SSJ	a-b	b-c	d-e	e-f						
'DH-	chr02:29271222	chr02:29272451	chr08:37723288	chr08:37727005						
Pahang'										
reference										
genome										
	Chromos	ome 2T8	Chromos	ome 8T2						
SSJ	а-х	x-f	d-y	у-с						
'Calcutta 4'	utg154:2414601	utg154:2415237	utg170:2943641	utg170:2944825						
assembly										

Annexe 4 - Supplementary Table 4 - Genomic position of the SSJs for the translocation 1/9.

	Chromos	Chromosome 9T1	
SSJ	k1-k3	k₃-h	h-k1
'Calcutta 4'	utg94:4774156	utg94:4778354	utg195:5616816
assembly			
Annexe 5 - Supplementary figure 2 - Genetic marker statistics along the 11 Musa acuminata chromosomes.



(A) Signature segment junction definition. Considering an accession with a standard chromosome structure (in blue, chr 2 and 3) and an accession with an alternative structure, a reciprocal translocation between chromosomes 2 and 3 result in chr 2T3 and 3T2 (in green). All segments (a,b,c,d) are present in both accessions, but some segment combinations and thus segment junctions are specific to chr 2 (a-b), chr 3(c-d), chr 2T3 (a-d) an chr 3T2 (c-d). They are referred to as specific segment junctions (SSJs).

(B) Accession characterization through paired-end read alignment. Paired-end reads of accessions to be structurally characterized are aligned on the sequence of the accession serving as reference for the standard structure chr 2 and chr 3 and on the sequence of the accession serving as reference for the alternative structures 2T3 and 3T2. When paired reads overlapped a SSJ's position (bold), this indicates that the analysed accession bears the same chromosome structure as the reference accession at this SSJ's position.





(A-B) Representation of marker linkage in 'Calcutta 4' self-progeny along *Musa* acuminata reference chromosomes. Each dot represents linkage between two markers, with the intensity ranging from blue to dark red. (A) Markers are plotted according to their physical distance along the *Musa acuminata* reference genome. Boxes along the axes symbolize scaffolds. A linkage break visible at the beginning of reference chromosome 4 most likely resulted from an absence of segregating markers (see Supplementary Fig. 1C) due to the presence of a homozygous region in 'Calcutta 4'. (B) Markers are plotted consecutively.

(C) Black and white boxes symbolize scaffolds and their orientation along the pseudomolecules. Genetic marker and gene densities are represented in orange and blue, respectively, based on a 100 kb window size. Annexe 7 - Supplementary figure 3 - Paired read mapping of 14 Musa accessions focused on the two reciprocal translocations detected in 'Calcutta 4'.



Circos representation of significant discordant read clusters from 14 accessions compared to the M. acuminata reference sequence assembly with a focus on paired read clusters detected in the targeted regions of chromosomes 2 and 8 (A) and chromosomes 1 and 9 (B). Grey internal lines correspond to paired reads with a correct orientation and insert size, red and blue lines correspond to discordant pairs with a higher insert size and reverse mapping orientation, respective

Annexe 8 - Supplementary figure 4 - Factorial analysis performed on 32 wild *M. acuminata* accessions with projection of 51 cultivated accessions.



The dissimilarity matrix was based on SNPs from WGS data. Pink dots indicate accessions homozygous for chromosomes 2T8 and 8T2 and carrying at least one version of chromosomes 1T9 and 9T1 (see text). Black dots indicate accessions homozygous for chromosomes 2 and 8. Purple dots represent heterozygous accessions for the 2/8 translocation. Grey dots represent accessions with undetermined structure for the 2/8 translocation. Wild accession names are colored according to their subspecies: yellow: *M. a.* ssp. *burmannica, burmannicoides, siamea*; green: *M. a.* ssp. *banksii*; Blue: *M. a.* ssp. *malaccensis;* Red: *M. a.* ssp. *Zebrina*; Purple: *M. a.* ssp. *Microcarpa*; Dark grey: other wild *M. a.* subspecies. **A**: axis 1/3, **B**: axis 2/3. Eigenvalues: axis 1 = 27.27; axis 2 = 21.11; axis 3 = 15.93.

Annexe 9 – Couverture en lecture et ratios alléliques de l'individu BCM 21, porteur d'une aneuploïde segmentale sur le fragment T8 et triploïde pour le chromosome 2 complet.

