



HAL
open science

Analyse des bases génétiques de la tolérance à la pourriture basale du stipe chez le palmier à huile (*Elaeis guineensis* Jacq.)

Aurélie Daval

► To cite this version:

Aurélie Daval. Analyse des bases génétiques de la tolérance à la pourriture basale du stipe chez le palmier à huile (*Elaeis guineensis* Jacq.). Sciences du Vivant [q-bio]. 2018. dumas-01958430

HAL Id: dumas-01958430

<https://dumas.ccsd.cnrs.fr/dumas-01958430>

Submitted on 18 Dec 2018

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

AGROCAMPUS
OUEST

CFR Angers

CFR Rennes



UNIVERSITÉ DE
RENNES 1

 **cirad**
LA RECHERCHE AGRONOMIQUE
POUR LE DÉVELOPPEMENT

Année universitaire : 2017 - 2018

Master Biologie, Agrosiences

Parcours Amélioration, Production,
Valorisation du végétal

Option : Génétique, Génomique et
Amélioration des Plantes

Rapport de stage

d'Ingénieur de l'Institut Supérieur des Sciences agronomiques,
agroalimentaires, horticoles et du paysage

de Master de l'Institut Supérieur des Sciences agronomiques,
agroalimentaires, horticoles et du paysage

d'un autre établissement (étudiant arrivé en M2)

Analyse des bases génétiques de la tolérance à la pourriture basale du stipe chez le palmier à huile (*Elaeis guineensis* Jacq.)

Par : Aurélie Daval

Soutenu à Rennes le 28/06/2018

Devant le jury composé de :

Président : M. Jubault

Maître de stage : S. Tisné

Enseignant référent : M. Jubault

Examineur : M. Manzanares-Dauleux

Rapporteur : M.L. Pilet-Nayel

Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celles d'AGROCAMPUS OUEST et l'université de Rennes 1

Ce document est soumis aux conditions d'utilisation
« Paternité-Pas d'Utilisation Commerciale-Pas de Modification 4.0 France »
disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr>



Fiche de confidentialité et de diffusion du mémoire

Confidentialité

Non Oui si oui : 1 an 5 ans 10 ans

Pendant toute la durée de confidentialité, aucune diffusion du mémoire n'est possible ⁽¹⁾.

Date et signature du maître de stage ⁽²⁾ :

A la fin de la période de confidentialité, sa diffusion est soumise aux règles ci-dessous (droits d'auteur et autorisation de diffusion par l'enseignant à renseigner).

Droits d'auteur

L'auteur⁽³⁾ **Aurélié Daval**

autorise la diffusion de son travail (immédiatement ou à la fin de la période de confidentialité)

Oui Non

Si oui, il autorise

la diffusion papier du mémoire uniquement(4)

la diffusion papier du mémoire et la diffusion électronique du résumé

la diffusion papier et électronique du mémoire (joindre dans ce cas la fiche de conformité du mémoire numérique et le contrat de diffusion)

accepte de placer son mémoire sous licence Creative commons CC-By-Nc-Nd (voir Guide du mémoire Chap 1.4 page 6)

Date et signature de l'auteur :

Autorisation de diffusion par le responsable de spécialisation ou son représentant

L'enseignant juge le mémoire de qualité suffisante pour être diffusé (immédiatement ou à la fin de la période de confidentialité)

Oui Non

Si non, seul le titre du mémoire apparaîtra dans les bases de données.

Si oui, il autorise

la diffusion papier du mémoire uniquement(4)

la diffusion papier du mémoire et la diffusion électronique du résumé

la diffusion papier et électronique du mémoire

Date et signature de l'enseignant :

(1) L'administration, les enseignants et les différents services de documentation d'AGROCAMPUS OUEST s'engagent à respecter cette confidentialité.

(2) Signature et cachet de l'organisme

(3) Auteur = étudiant qui réalise son mémoire de fin d'études

(4) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option)) sera signalée dans les bases de données documentaires sans le résumé

Remerciements

En présentant ce rapport, je souhaite remercier Norbert Billotte, responsable de l'équipe Génome et Sélection du CIRAD, pour son accueil et sa bienveillance.

J'adresse mes sincères remerciements à mon (super) encadrant Monsieur Sébastien Tisné, pour sa disponibilité, son soutien (à toute épreuve), ses encouragements et surtout pour ses enseignements. Je suis particulièrement ravie d'avoir la possibilité de continuer à travailler ensemble et de me former un peu plus à l'univers de la génétique quantitative.

Egalement un grand merci à Sandrine Le Squin, merci de m'avoir formé à ASreml-R, merci pour ses présentations et toutes ses explications.

Merci à PalmElit pour leur accueil, les échanges et leur confiance.

Merci à Virginie Pomiès pour m'avoir appris à dompter les microsatellites.

Je souhaite remercier toutes les personnes présentes lors de ces fameuses pauses-café pour les moments de détente, de décontraction et de rigolade. Merci à Virginie, Marie, Clara, les Benjamin, Najatte, Roselyne (maman de tous les stagiaires), Alexandre, Letizia, Chandro, Juliette, Nolan, Julien et Antoine. Merci de votre accueil.

Liste des tableaux

Tableau I : Estimation des composantes de la variance.

Tableau II : QTL associés à la résistance au *Ganoderma* dans la population étudiée.

Liste des figures

Figure 1 : Surfaces et rendements agricoles dédiés à l'oléiculture des principaux oléagineux (Oil world 2016).

Figure 2 : Principaux pays producteurs d'huile de palme, d'après FAOSTAT 2013 (<http://www.fao.org/faostat/en/#data/QC>)

Figure 3 : Différents types de coques dans les fruits du palmier à huile. La coque est plus ou moins épaisse selon le type de fruit sous contrôle monogénique (Photographie de Singh et al., 2013).

Figure 4 : Echelle symptomatique de la pourriture basale du stipe.

Figure 5 : Visualisation d'une palmeraie atteinte par le *Ganoderma*.

Figure 6 : Mises en évidence des symptômes visibles en pépinière.

Figure 7 : Visualisation du pédigrée via le logiciel Pedimap (Biometris, Wageningen UR).

Figure 8 : Visualisation de l'occurrence d'isolats au sein des séries.

Figure 9 : Occurrence différentielle des géniteurs des groupes hétérotiques.

Figure 10 : Représentation graphique de la modélisation de l'aire sous la courbe de progression de la maladie (AUDPC).

Figure 11 : Méioses informatives moyennes par marqueurs.

Figure 12 : Carte génétique des pedigrees La Mé, Yangambi et intégrée.

Figure 13 : Distribution de la proportion d'individus atteints par le *Ganoderma*.

Figure 14 : Cinétique et AUDPC.

Figure 15 : Distribution des BLUP par origine de croisement.

Figure 16 : Profils des modèles QTL.

Figure 17 : Génotypes estimés des fondateurs aux QTL.

Figure 18 : Boxplot de l'effet des génotypes aux QTL.

Figure 19 : Haplotypage et effet des haplotypes aux marqueurs avoisinant le QTL du groupe de liaison 13 chez les autofécondations de L1.

Liste des abréviations utilisées

AGC	Aptitude générale à la combinaison
ASC	Aptitude spécifique à la combinaison
AUDPC	Area under the disease progress curve
BF	Bayes factors
BLUP	Best Linear Unbiased Precrion
CIRAD	Centre de coopération Internationale en Recherche Agronomique pour le Développement
EST	expressed sequence tag
FELDA	Federal Land Development Authority
GBS	Genotyping by Sequencing
GLMM	Generalized linear mixed model
IBD	Identical by descends
IBS	Identical by states
IPM	Integrated Pest Managment
IRD	institut de Recherche pour le Développement
IRHO	Institut de Recherche sur les Huiles et les Oleagineux
PPS	Porportion totale de jeune Palmier Symptomatique
QTL	Quantitative Trait Loci
SNP	Single nucleotid polymorphism
SRR	Sélection réccurrente réciproque
SSR	Simple sequence repeat
USDA	United States Department of Agriculture

Table des matières

1.	Introduction.....	1
1.1.	Le palmier à huile une plante agronomique d'intérêt	1
1.2.	Différentes populations de palmier à huile.....	1
1.3.	L'amélioration du palmier à huile par CIRAD-PalmElit.....	2
1.4.	Présentation du pathogène.....	3
1.5.	Présentation des tests précoces en pépinière	3
1.6.	Essais génétiques et modélisations.....	4
1.7.	La détection de QTL dans une population en pedigree	4
1.8.	Problématique et objectifs du stage	5
2.	Matériel et méthode	6
2.1.	Matériel génétique.....	6
2.2.	Données phénotypiques.....	6
2.2.1.	Dispositif des tests précoces	6
2.2.2.	Acquisition des données.....	6
2.2.3.	Modélisation du phénotype	7
2.2.4.	Mod _A , Mod _B et Mod _C effet de l'apparentement sur la prédiction des BLUP	7
2.2.5.	Mod _D : prise en compte de la cinétique de la maladie par l'AUDPC.....	8
2.2.6.	Estimation des composantes de la variance	9
2.3.	Données moléculaires	10
2.4.	Construction de la carte génétique	10
2.5.	Détection de QTL.....	11
3.	Résultats	12
3.1.	Polymorphisme des marqueurs et cartographie génétique	12
3.2.	Analyse des données brutes de phénotypage	12
3.3.	Modélisation différentielle du phénotype	13
3.4.	Modèles QTL, génotypes et effets associés aux QTL.....	13
4.	Discussion	14
4.1.	Cartographie génétique d'une population en pedigree.....	14
4.2.	Utilisation des BLUP pour la cartographie de QTL, approche bayésienne	15
4.3.	Comparaison des résultats avec la littérature.....	17
4.4.	La sélection du palmier à huile, ce qu'apporte cette étude.....	18
5.	Conclusion	19
6.	Bibliographie.....	20

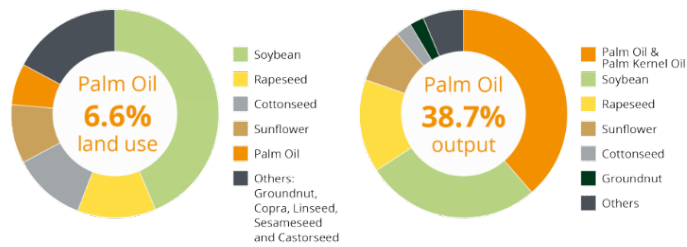


Figure 1 : Surfaces et rendements agricoles dédiés à l'oléiculture des principaux oléagineux (Oil world 2016).

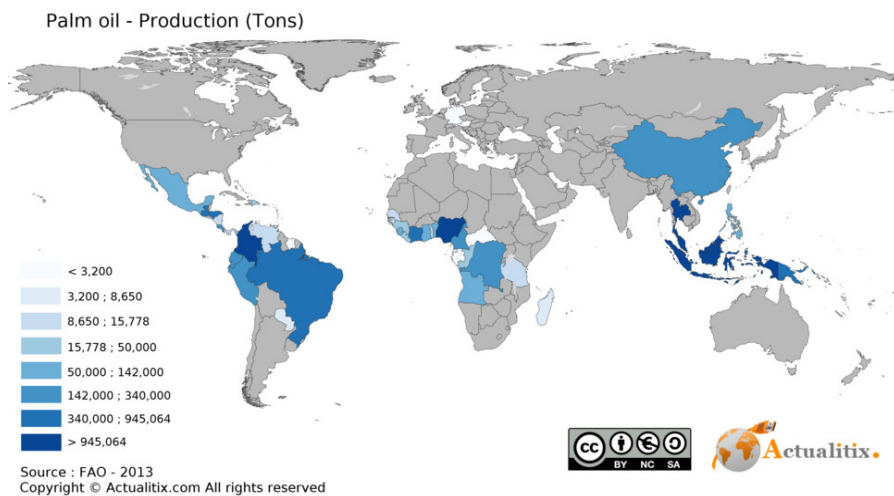


Figure 2 : Principaux pays producteurs d'huile de palme, d'après FAOSTAT 2013 (<http://www.fao.org/faostat/en/#data/QC>)

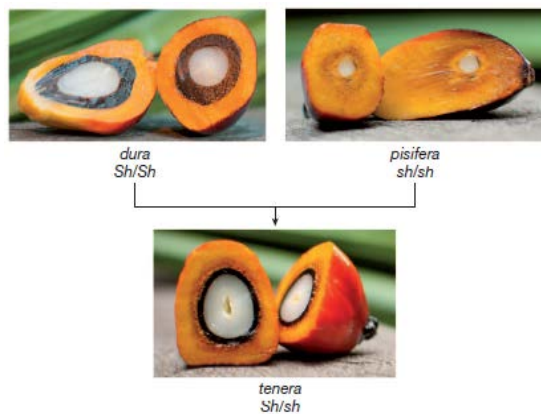


Figure 3 : Différents types de coques dans les fruits du palmier à huile. La coque est plus ou moins épaisse selon le type de fruit sous contrôle monogénique (Photographie de Singh et al., 2013).

1. Introduction

1.1. Le palmier à huile une plante agronomique d'intérêt

Les palmiers apparaissent comme les plantes les plus emblématiques des régions tropicales par leur morphologie caractéristique, leur importance écologique, ethnologique et économique (Annexe I). Le palmier dattier, le cocotier et le palmier à huile sont trois représentants bien connus de cette famille. Bien qu'aujourd'hui l'image de ce dernier ait ternie celle de ses cousins, sa production n'en demeure pas moins importante à la fois pour les populations locales, mais également en réponse à la demande mondiale croissante en oléagineux avec une augmentation nécessaire de plus de 30 % (United States Department of Agriculture, USDA 2015, www.fas.usda.gov). Cette demande croissante en huile végétale est liée à la démographie croissante de la population et la diversification des usages de l'huile de palme qui touche deux types de secteurs, le secteur agroalimentaire et cosmétique et le secteur des bioénergies (Oil World 2016, www.oilworld.biz).

La production d'huile de palme est la plus élevée de toutes les huiles végétales avec 3.8 tonnes par hectare en 2016 contre 0.8 tonnes par hectare d'huile de soja, seconde production mondiale d'huile végétale (Oil World 2016). Le département des États-Unis de l'agriculture (USDA) estime la production globale d'huile de palme en 2017 à 62.88 million de tonnes avec la plus petite surface agricole dédiée à la production d'huile végétale (Figure 1). La production d'huile de palme couvre 39% des productions d'huiles végétales (FAOSTAT 2013). Les cinq principaux pays producteurs en 2013-2014 sont l'Indonésie, la Malaisie, la Thaïlande, le Nigeria et la Colombie fournissant 85 % de la production d'huile de palme (Jacquemard 2012; Zulkifli *et al.* 2017) (Figure 2).

Le palmier à huile, *Elaeis guineensis* Jacq. (2n = 32) est une monocotylédone pérenne, allogame, monoïque (dichogame), appartenant à la famille des Arecacées (Jacquemard 2012). Les premières récoltes du palmier à huile se font entre 4 et 6 ans après germination et son exploitation peut durer entre 25 et 35 ans (Prevot 1962).

Le genre *Elaeis* ne concerne que deux espèces tropicales *Elaeis guineensis* Jacq. et *Elaeis oleifera* (Kunth) Cortés., ayant pour centre d'origine deux régions différentes, respectivement l'Afrique (Guinée) et l'Amérique latine (Amazonie)(Cochard *et al.* 2009). *E. guineensis* présente un intérêt agro-alimentaire et économique par ses fruits chargés en huiles de palme et de palmiste (Figure 3, Annexe II) largement utilisées dans le monde. Le fruit du palmier à huile est une drupe composée d'un péricarpe, d'un mésocarpe (ou pulpe), d'un endocarpe (ou coque) et d'une amande. La pulpe est chargée d'huile de palme et l'amande d'huile de palmiste. Il existe 3 types de fruits chez le palmier à huile (Figure 3) :

- Type *dura* : fruit contenant une coque épaisse (jusqu' à 7 mm).
- Type *pisifera* : fruit ne contenant quasiment pas de coque et un maximum de pulpe (femelle stérile).
- Type *tenera* : hybride de *dura* et *pisifera* avec une épaisseur de coque inférieure à 2 mm (utilisé pour l'exploitation agricole).

L'épaisseur de coque est sous contrôle monogénique par le gène SHELL, le génotype Sh/Sh est retrouvé chez les *dura*, Sh/sh chez les *tenera* et sh/sh chez les *pisifera* (Beirnaert 1941; Singh *et al.* 2013). D'un point de vue de l'optimisation des rendements, le maintien des palmiers *dura* ou *pisifera* (utilisé comme géniteurs mâles) permet l'obtention de 100 % de *tenera* en production de semences pour l'exploitation agricole.

1.2. Différentes populations de palmier à huile

Chez le palmier à huile, il a été constaté des différences importantes entre les matériels de diverses origines génétiques à la fois pour les caractéristiques végétatives mais aussi pour la production et la qualité des régimes. En revanche d'un point de vue intra origine la variabilité est relativement faible (Meunier and Gascon 1972). L'importance des croisements entre palmiers d'origines différentes n'a été évaluée qu'à la suite d'une initiative nommée « Expérience Internationale » en 1949 par l'Institut

de Recherche sur les Huile et Oléagineux (IRHO) et consistant en une comparaison des matériels de différentes provenances géographiques (dites origines) et de leurs hybridations (Cao 1995). C'est en ayant ces informations sur la structuration de la diversité génétique qu'un véritable gain de productivité a été observé (Cochard *et al.* 2009).

À la suite de ces observations, les origines parentales ont été rassemblées (notamment par le Centre de coopération Internationale en Recherche Agronomique pour le Développement, CIRAD) en deux groupes de caractéristiques de production complémentaires nommés groupes hétérotiques A et B. Le groupe A rassemble les origines produisant un petit nombre de gros régimes. Ce groupe est composé en majorité de palmiers (originaire principalement de Déli en Indonésie) dont l'origine remonte à l'importation de quatre plantules *dura* dans le jardin botanique de Bogor (Java, Indonésie) en 1848 (Corley and Tinker 2003). Par la suite ce groupe d'origine génétique restreinte a été enrichi par des introductions de palmiers ayant des caractéristiques de production semblables, notamment provenant d'Angola en Afrique. Le groupe B rassemble des origines génétiques produisant un grand nombre de petits régimes et pour lesquelles l'allèle *Sh* ségrége, ce qui permet d'obtenir des géniteurs *pisifera*. Les origines utilisées dans les programmes d'amélioration du CIRAD et de sa filiale PalmElit sont majoritairement les origines La Mé (Côte d'Ivoire) et Yangambi (Congo) (Meunier and Gascon 1972; Jacquemard 2012).

L'étude de la diversité et de la structuration génétique d'une population palmier à huile de Cochard *et al.* (2009) confirme l'existence de deux groupes génétiques Africains, le premier regroupant les populations d'Afrique de l'Ouest dont celle de Côte d'Ivoire (Groupe I) et le second les populations du Bénin, du Nigéria et de l'Afrique centrale et introduites au Brésil, (Groupe II). La population Déli formerait un troisième groupe plus proche du groupe II que du groupe I. Cette étude souligne le gain de productivité des hybrides issues de croisements entre deux origines génétiquement éloignées, comme entre les La Mé (Côte d'Ivoire) et les Déli (Indonésie) qui sont les croisements à fort rendement en huile de palme.

1.3. L'amélioration du palmier à huile par CIRAD-PalmElit

En 1957, le CIRAD et ses partenaires mettent au point un schéma de sélection récurrente réciproque (cité SRR dans la suite du rapport) se basant sur l'exploitation de l'hétérosis existant entre les deux groupes hétérotiques (Cao 1995). Un cycle de SRR s'organise en 3 étapes majeures, partant des populations initiales correspondant aux groupes hétérotiques A et B. La première étape permet de sélectionner des individus au sein des populations (sélection massale). La seconde étape mène à évaluer chaque individu en croisement avec des individus de l'autre groupe (Meunier and Gascon 1972). En effet, la SRR utilise chaque groupe de population comme testeur l'un pour l'autre (Annexe III). Les individus ou parents de chaque groupe seront alors sélectionnés pour leur bonne aptitude générale à la combinaison (AGC) ou leur bonne aptitude spécifique à la combinaison (ASC). Et enfin une étape de recombinaison des parents de chaque groupe sélectionné est effectuée (Durand-Gasselin *et al.* 1999; Jacquemard 2012).

L'amélioration variétale du palmier à huile doit tenir compte de plusieurs critères pour optimiser la durabilité de la production d'huile de palme. La sélection suit alors trois axes de développement majeurs : la sélection pour la durée de production, pour le rendement et pour la résistance aux maladies (Durand-Gasselin *et al.* 2010).

L'amélioration génétique joue un grand rôle dans le gain de productivité et la recherche de résistance aux maladies. Le gain de productivité imputable à la sélection génétique a été égal à 1% par an pendant 50 ans, les enjeux actuels se sont déplacés sur la résistance aux maladies qui permettrait d'atteindre un rendement optimal (Jacquemard 2012). Sur ce dernier point, il est essentiel pour le palmier à huile de déterminer des sources génétiques de résistances aux maladies, ceci s'intégrant dans une volonté générale de produire de l'huile de palme durable (« *The Global initiative for Sustainable Palm Oil Production* », www.rspo.org), en utilisant des méthodes alternatives pour lutter contre les pathogènes et les ravageurs (Integrated Pest Management, IPM).



Figure 4 : Echelle symptomatique de la pourriture basale du stipe. Différents symptômes du *Ganoderma* visibles sur les palmiers à huile au champ. Ces symptômes forment une échelle de notation de la maladie, non obligatoire et non ordonnés dans le temps. Crédit photo Cirad.



Figure 5 : Visualisation d'une palmeraie atteinte par le *Ganoderma*. Capture d'écran d'un zone d'exploitation d'huile de palme à Tanah Gambus en Indonésie. A gauche de la photo une densité moyenne en palmier à huile, à droite de la photo un manque important de palmiers à huile du entre autre au *Ganoderma* (Google maps 2018).

Les programmes d'amélioration du palmier à huile mis en place par le Cirad et PalmElit se renforcent dans la sélection de géotypes tolérants aux maladies (fusariose, *Ganoderma* et pourriture du cœur) qui deviennent une menace majeure dans la totalité des aires de production. La fusariose est principalement retrouvée dans les palmeraies Africaines tandis que le *Ganoderma* est largement répandu en Asie du Sud-Est (Franqueville *et al.* 1990) et la pourriture du cœur en Amérique latine. L'obstacle majeur à la durabilité de la production d'huile de palme en Asie du Sud-Est est la pourriture basale du stipe maladie causé par le champignon pathogène *Ganoderma boninense*.

1.4. Présentation du pathogène

Ganoderma boninense (nommé *Ganoderma* dans la suite du rapport) est un basidiomycète pathogène héli-biotrophe responsable de la pourriture basale du stipe chez le palmier à huile. Le *Ganoderma* est un champignon tellurique lignivore qui peut engendrer une combinaison de symptômes, mais aucun échelle symptomatique n'a encore été définie (Chong *et al.* 2017). Au champ, les symptômes peuvent prendre 5 formes (Figure 4) : la présence de pourriture due à la dégradation des composés de la paroi cellulaire du stipe du palmier à huile, la présence de sporophores (structure de dissémination des spores du pathogène) le long du stipe, le dessèchement des palmes, le non-développement des jeunes palmes et la mort du palmier (Cooper *et al.* 2011).

Il y a trois voies possibles d'inoculation de la maladie à un palmier : contact entre un débris compact du sol infecté par le *Ganoderma* et les racines du palmier à huile (considéré comme inoculum primaire lors de la première infection parcellaire), contact racinaire entre deux palmiers et dissémination par germinations des spores sur débris végétaux (Pilotti *et al.* 2003).

Le cycle du *Ganoderma*, bien que demeurant encore incertain, est établi comme suit : la germination d'une spore s'effectue dans le sol, le mycélium est considéré comme inoffensif et est monocaryotique. La rencontre entre deux mycéliums monocaryotiques engendre une étape de plasmogamie à l'origine du mycélium secondaire qui sera hétérocaryotique et infectieux. Les études n'ont pas encore départagé le moment où le pathogène colonise un nouveau substrat ligneux (Chong *et al.* 2017).

Cette maladie est à l'origine de nombreuses pertes de rendement et de palmiers au champ (Lim *et al.* 1992) (Figure 5). Le pathogène n'est pas spécifique du palmier à huile mais peut entraîner des dommages considérables. Jusqu'à 50 % de pertes sont observées au champ dans certains fonds génétiques (Durand-Gasselin *et al.* 2005), malgré les efforts importants pour mettre au point des techniques de lutte contre le pathogène, notamment en Indonésie (Corley and Tinker 2003). C'est pourquoi l'étude du déterminisme génétique de la résistance au *Ganoderma* est si intéressante pour les centres de recherche tels que le CIRAD qui s'investissent dans ces programmes d'amélioration.

Certaines origines génétiques ont pu être évaluées comme étant plus tolérantes que d'autres au champ. Afin de déterminer précisément la sensibilité des géniteurs utilisés en sélection et d'accélérer l'obtention de variétés résistantes, un dispositif standardisé basé sur l'inoculation précoce de plantules de palmiers en pépinière a été développé (Breton *et al.* 2006).

1.5. Présentation des tests précoces en pépinière

Le développement des tests précoces en pépinière (Breton *et al.* 2006) a d'abord permis de déterminer les paramètres des essais afin de rendre l'infection artificielle quantifiable et reproductible. Lors des tests précoces (Breton *et al.* 2009), la sensibilité des croisements, et non des individus, à la maladie est mesurée. Chaque croisement est représenté par un nombre fixé de descendants et c'est le suivi du nombre de descendants atteints par la maladie qui est réalisé tout au long de l'essai (figure 6).

Les symptômes notés pendant l'essai sont également standardisés. Deux types de symptômes sont relevés, les symptômes externes (figure 6A) relevés tous les mois pendant l'essai et les symptômes internes (figure 6B) à la fin de l'essai. Les symptômes externes peuvent prendre plusieurs formes, soit un dessèchement des plantules, soit l'apparition de sporophores (figure 6A). Finalement, les

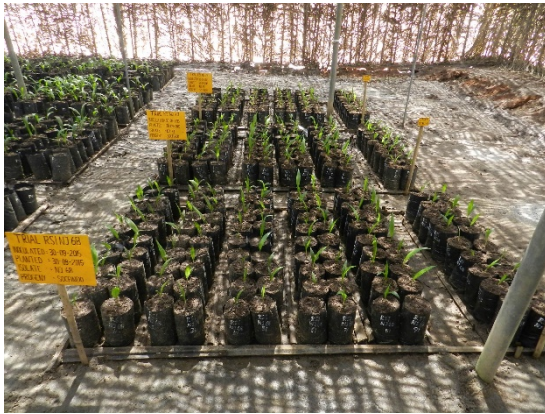


Figure 6 : Mises en évidence des symptômes visibles en pépinière. Photographies de tests précoces en pépinière, des symptômes externes (A) visibles en fin de tests (sporophores et dessèchement des petites palmes) et des symptômes internes (B). Crédit photo Cirad et PalmElit.

croisements étudiés peuvent être discriminés sur leur proportion de descendants atteints par le *Ganoderma* ou sur leur cinétique de progression de la maladie.

1.6. Essais génétiques et modélisations

Les essais génétiques en pépinière ont pour but de sélectionner les géniteurs les plus résistants au *Ganoderma* au sein des familles utilisées en amélioration. Le traitement des données obtenues en pépinière a pour objectif d'évaluer l'effet génétique de chaque géniteur en croisement. Il a été vu précédemment que le palmier à huile est une plante pérenne pouvant être exploitée pendant une trentaine d'années, par conséquent, la généalogie des palmiers utilisés en amélioration est connue (Durand-Gasselin *et al.* 2010).

Les essais génétiques évaluent les géniteurs engagés dans plusieurs croisements. Plus un géniteur est engagé dans de nombreux croisements, mieux son effet sera prédit. Un des changements majeurs en génétique quantitative des plantes cultivées a été l'utilisation de modèles mixtes, notamment pour les populations en pedigree, permettant de prendre en compte la non indépendance des individus étudiés (Lynch and Walsh 1998). En effet, le palmier à huile ne peut pas subir un trop grand nombre d'autofécondations (une ou deux supportées), puisqu'il est rapidement soumis aux effets de dépression de consanguinité (Durand-Gasselin *et al.* 2005) ce qui ne permet l'obtention de lignées pures. Le modèle animal permet lors de l'étude de population en pedigree (présence de familles multigénérationnelles interconnectées) d'estimer les composantes de la variance et de prédire les valeurs génétiques associées à chaque géniteur tout en tenant compte de l'apparentement (Kruuk 2004). Ce modèle a servi de point de départ à la présente étude.

Pour l'analyse des données acquises dans le cadre de programme d'amélioration, le CIRAD et PalmElit se sont inspiré d'un modèle dit modèle animal, appliqué à l'évaluation génétique des bovins laitiers (Ducrocq 1990) pour l'appliquer au palmier. La modélisation des données de plusieurs tests précoces par un modèle animal permet l'obtention d'un BLUP (Best Linear Unbiased Prediction) pour chaque géniteur utilisé en croisement (communication personnelle, PalmElit). L'utilisation d'un tel modèle permet de classer les géniteurs les plus résistants au *Ganoderma*.

1.7. La détection de QTL dans une population en pedigree

La recherche concernant le déterminisme génétique de la résistance des plantes aux pathogènes est une branche active et importante des instituts de recherche ayant pour but de réduire à la fois l'utilisation des produits phytosanitaires et les pertes agricoles imputables aux maladies. On peut citer par exemple le cas de la recherche pour la lutte contre *Plasmodiophora brassicae* chez le colza (Some *et al.* 1996) ou encore contre *Ralstonia solanacearum* chez la tomate (Carmeille *et al.* 2006). Deux types de résistances ont été historiquement distinguées (van der Plank 1968). Une résistance dite qualitative, sous contrôle mono- ou oligogénique (gène-R), est souvent associée à des réactions d'hypersensibilités (HR) et déclenchée par des facteurs d'avirulence. Une résistance dite quantitative sous contrôle polygénique corrobore à un effet génétique global. Le grand nombre d'étude sur le déterminisme génétique des maladies des plantes cultivées rapporte l'importance des QTL pour la durabilité des productions et des résistances à la maladie (Poland *et al.* 2009).

Cartographier les locus de caractères quantitatifs (*Quantitative Trait Loci*, QTL) le long du génome en association avec les marqueurs utilisés, permet aux chercheurs et aux sélectionneurs d'utiliser la sélection assistée par marqueurs pour optimiser l'amélioration variétale (Dekkers and Hospital 2002). Les généticiens et sélectionneurs doivent développer ou utiliser des méthodes statistiques pour identifier des QTL, en utilisant des marqueurs moléculaires. Le design des populations en pedigree ne permet pas d'utiliser les méthodes de détection de QTL dédiées aux populations biparentales (Würschum 2012), en considérant leur faible effectif par famille. Deux types d'approches sont possibles pour l'analyse QTL en pedigree, une approche fréquentiste nommée « *two-step variance component* » développé par (George *et al.* 2000) et une approche Bayésienne toutes deux initialement développées sur des populations en pedigree animales ou humaines.

(Bink *et al.* (2002) ont développé une méthode bayésienne pour la cartographie de QTL dans des populations en pedigree associant robustesse et flexibilité quant aux structures d'apparentement et au nombre de représentant par famille. Cette méthode a été utilisée entre autres, sur des populations de pommier en pedigree (Bink *et al.* 2008) mais aussi dans une population multiparentales de fraise (Mangandi *et al.* 2017). La méthode bayésienne prédit les informations d'identité par descendance (IBD identity-by-descend) ce qui permet de visualiser l'origine des allèles dans la descendance et de savoir de quel fond génétique provient un QTL putatif. Cette méthode représente un avantage pour la sélection, la ségrégation des allèles d'intérêts étant ainsi observée.

1.8. Problématique et objectifs du stage

Dans le but de maximiser et d'accélérer l'amélioration du palmier à huile par la sélection assistée par marqueur, il est nécessaire de connaître les fonds génétiques résistants et les régions associées à la tolérance à la maladie. L'objectif de ce stage a été de réaliser une analyse des bases génétiques de la pourriture basale du stipe au sein de représentant du groupe hétérotique B du palmier à huile.

À la vue du pedigree de la population d'étude, les questions posées au début de ce travail étaient de savoir par quels outils et quelles modélisations est-il possible de traiter des données multigénérationnelles ? Quels sont les régions QTL associées à la tolérance au *Ganoderma* ? Ont-ils déjà été mis en évidence dans des populations à design différent ou dans des conditions expérimentales différentes ?

Pour mener à bien cette étude, plusieurs étapes ont été réalisées. Tout d'abord l'étude bibliographique a permis de prendre en considération l'approche des auteurs dans l'étude de population en pedigree puis de rapprocher les résultats obtenus avec ceux de la présente étude. Ensuite, il a été nécessaire d'étudier le génotypage des marqueurs microsatellites polymorphes pour chacun des individus de la population. Cette étape a permis la construction d'une carte génétique propre à la population de l'étude. Les données de phénotypage ont été analysées et modélisées afin de prédire les valeurs génétiques des individus de l'étude. Ces étapes ont permis d'accéder à l'identification des modèles QTL par l'utilisation d'un logiciel Bayésien utilisé pour l'analyse de données multigénérationnelles interconnectées. Une discussion est proposée abordant les différents choix réalisés dans cette étude, leurs conséquences et les perspectives d'étude d'un tel dispositif.

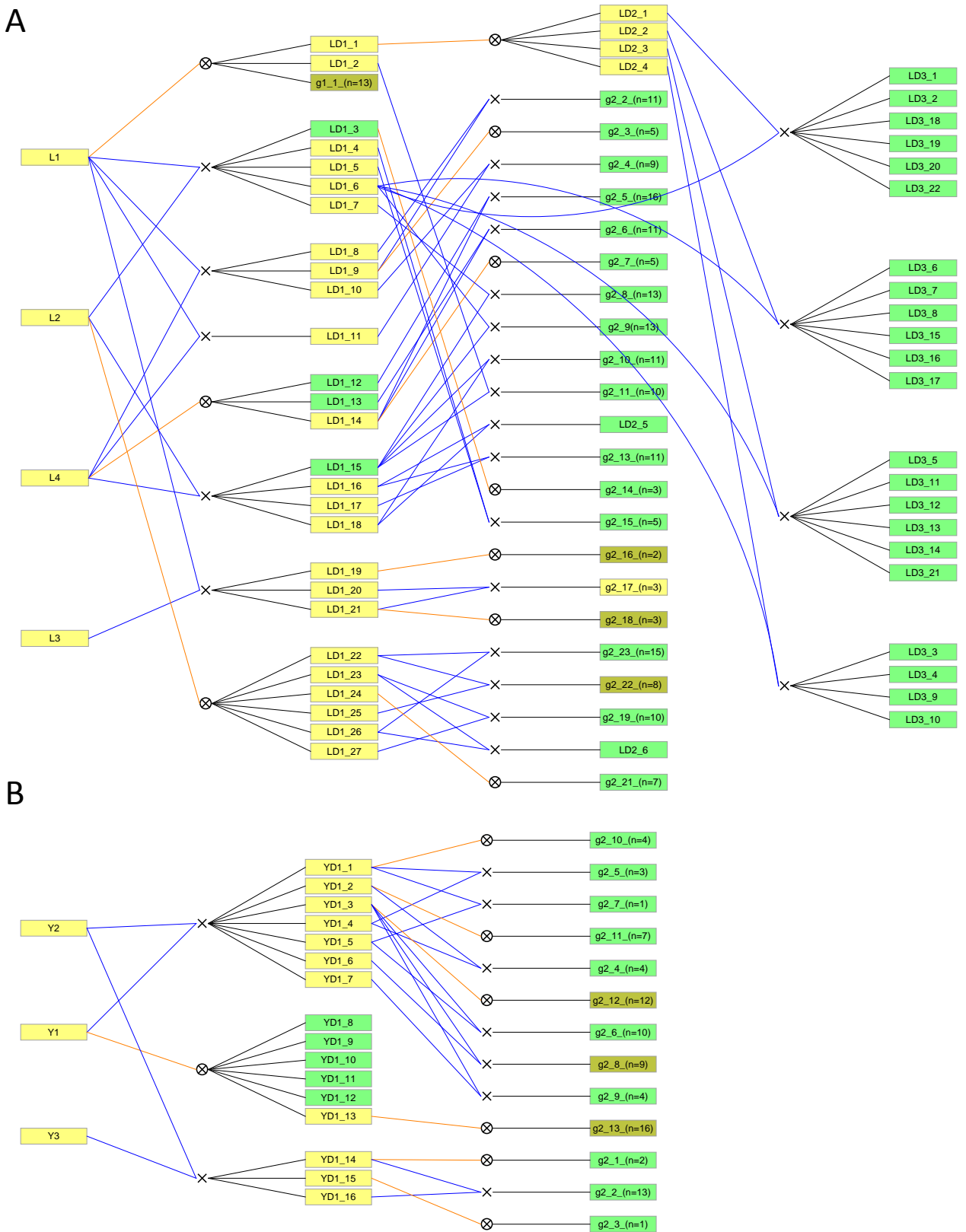


Figure 7 : Visualisation du pédigrée via le logiciel Pedimap (Biometris, Wageningen UR). Les individus les plus à gauche représentent les fondateurs utilisés dans les programmes d'amélioration du palmier à huile au CIRAD. Les traits bleus représentent les croisements biparentaux tandis que les traits oranges représentent les autofécondations. Pour plus de lisibilité l'ensemble des individus du pédigrée n'a pas été représenté ici, les individus plein frères étant regroupés sous les termes g1, g2 désignant respectivement génération 1, génération 2. La figure 1A représente les palmiers (n = 239) ayant pour origine LaMé (Côte d'Ivoire) et la figure 1B représente les palmiers (n = 95) ayant pour origine Yangambi (RDC). Un code couleur a été établi permettant de repérer les palmiers phénotypés: la couleur verte représente un palmier ou une famille de palmiers phénotypée, la couleur kaki représente une famille de palmiers partiellement phénotypée, la couleur jaune représente les palmiers non phénotypés.

2. Matériel et méthode

2.1. Matériel génétique

Les individus choisis pour cette étude font partie du programme d'amélioration du palmier à huile, réalisé par le CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement) et sa filiale PalmElit. Ce programme suit un schéma de sélection récurrente réciproque (SRR) (Meunier and Gascon 1972) (Annexe III).

Pour rappel, le groupe hétérotique B rassemble les populations africaines, et plus particulièrement pour notre étude, les populations ayant pour origine génétique La Mé (Côte d'Ivoire) ou Yangambi (République Démocratique du Congo) avec pour caractéristiques de produire beaucoup de petits régimes et des fruits de tous types (*dura*, *tenera* et *pisifera*) (Gascon and De Berchoux 1964).

Le pedigree de la population de palmier à huile (*Elaeis guineensis* Jacq.) étudié ici comprend 48 familles de demi-frères et de plein-frères interconnectées faisant partie du groupe B. Ces familles sont issues de différents types de croisements entre apparentés provenant de 7 génotypes fondateurs. Au total 328 génotypes ont été étudiés en plus de celui de chacun des 7 fondateurs (Figure 7). Il a volontairement été choisi de représenter les deux origines génétiques de la population sur deux pedigrees séparés car il n'y a pas de croisement inter-origine. Le logiciel Pedimap 1.2 (Biometris, UR Wageningen) a permis de construire une représentation du pedigree (Voorrips *et al.* 2012).

2.2. Données phénotypiques

2.2.1. Dispositif des tests précoces

Afin d'optimiser les cycles de sélection du palmier à huile, des tests précoces standardisés de résistance au *Ganoderma* ont été mis en place au CIRAD (Breton *et al.*, 2006). Ils sont réalisés à Tanah Gambus en Indonésie et permettent de classer les génotypes selon leur degré de résistance au pathogène. Chaque test (nommé série) s'organise en trois parties.

La partie 1 (3 mois) concerne la préparation du matériel de l'essai. Cette partie a pour objectif l'obtention de blocs d'hévéa stériles inoculés avec une souche de *Ganoderma*. La partie 2 (1 mois) concerne l'inoculation des graines de palmier à huile par le *Ganoderma* par la mise en place des blocs d'hévéa inoculés au fond des polybags, l'ajout de la terre, puis le repiquage des graines. L'ensemble des graines du dispositif est repiqué le même jour. Dans une série, 100 croisements différents sont testés. Ils sont représentés chacun par 100 individus (5 répétitions de 20 plants). La partie 3 (7 à 8 mois) concerne le suivi des symptômes. Dans un premier temps la bonne germination des graines est observée. L'apparition de nouveaux palmiers symptomatiques est notée tous les mois à toutes les deux semaines selon l'avancée dans l'essai. Dans cette étude, c'est la proportion de palmiers symptomatiques pour chaque croisement testé qui est pris en compte (Annexe IV).

Pour garantir la répétabilité des tests ainsi développés, des génotypes dits « standards » ont été étudiés avec un grand nombre de souches pathogènes et ceux-ci sont employés dans chaque test pour calibrer l'essai. Sur les 100 croisements étudiés par séries, 20% sont des croisements standards et servent d'étalonnage puisqu'ils sont constitués de croisements de sensibilité connue (sensibles, intermédiaires et résistants) à la maladie. Afin d'observer un maximum de diversité de réponse aux pathogènes, les tests sont déclarés terminés lorsque 30 % des génotypes standards sont atteints par le *Ganoderma* (Breton *et al.* 2009).

2.2.2. Acquisition des données

L'acquisition du phénotype a été réalisée à la suite de 102 séries standardisées sur 4017 croisements différents (AxA, AxB, BxB). Ces essais ont été réalisés entre 2007 et 2017. Dans une série, un seul isolat pathogène est utilisé pour l'ensemble des croisements testés (Figure 8). Sur les 4017 croisements différents, il y a 3910 croisements A x B. Seul ce type de croisement a été considéré pour la prédiction des valeurs génétiques. Les 3910 croisements A x B sont issues du croisement de 361 parents du groupe B avec 2069 parents du groupe A, générant des familles de pleins frères ou de

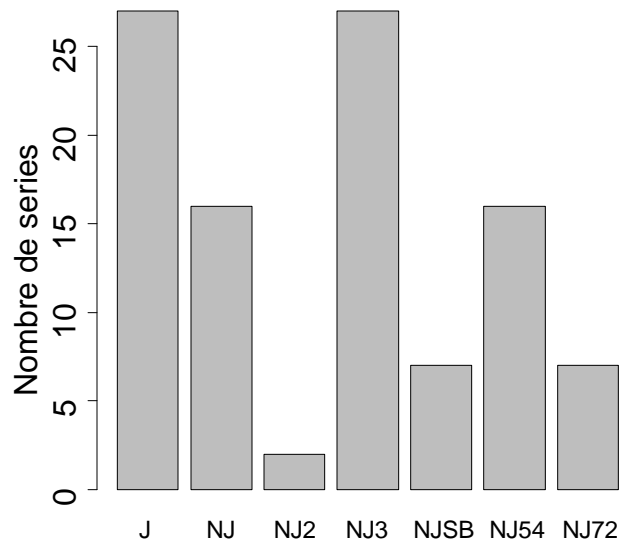


Figure 8 : Visualisation de l'occurrence d'isolats au sein des séries. Sept isolats du *Ganoderma* d'agressivité jugé similaire ont été utilisés au sein des 102 séries de tests précoces.

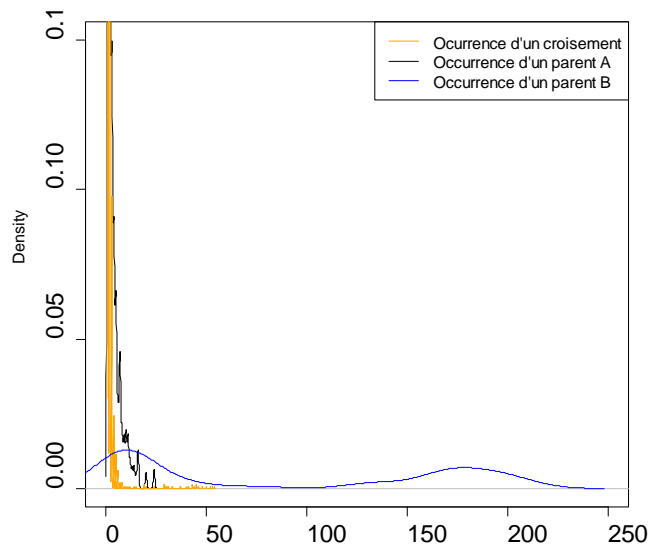


Figure 9 : Occurrence différentielle des géniteurs des groupes hétérotiques. Visualisation de l'occurrence hétérogène d'un parent en croisement (courbe bleue et noire), visualisation du nombre de fois qu'un même croisement intervient dans une série différente (courbe orange).

deux frères. La fréquence d'utilisation d'un parent au sein d'un croisement est très hétérogène (Figure 9). Les 102 séries ont permis de déterminer dans un premier temps la proportion d'individus d'un même croisement infectés puis de prédire la valeur génétique des 335 individus.

2.2.3. Modélisation du phénotype

Dans cette étude, il est question d'identifier et de prendre en compte les facteurs (série, isolat, croisement, parent A, parent B) influençant la probabilité qu'un palmier d'être infecté par le *Ganoderma*. Les modèles développés mod_A , mod_B et mod_C ont pris en compte la proportion totale de jeune palmier symptomatique (PPS) se basant sur les proportions totales de fin d'essai. Le mod_D a considéré la cinétique de la maladie.

Les trois premiers modèles testent par un modèle animal ou un modèle dit *Sir and Dam* une prise en compte différentielle de l'apparentement entre individus. Le dernier modèle teste l'effet de la prise en compte de la cinétique de la maladie par un modèle *Sir and Dam* sans apparentement.

Une matrice d'apparentement a été construite par la commande « asreml.Ainverse » de dimension $335 * 335$. La diagonale correspond aux coefficients d'apparentement entre individu (Coefficient de consanguinité). Une matrice d'identité est également calculée par ASREML et entre dans la modélisation des valeurs génétiques.

Pour prédire les valeurs génétiques des individus étudiés associées aux effets aléatoires du modèle, l'utilisation des modèles linéaires mixtes généralisés permet la prédiction d'analogues des valeurs génétiques nommés BLUP (Best Linear Unbiased Predictor) (Isik *et al.* 2017). La valeur génétique ainsi prédite représente le phénotype « résistance au *Ganoderma* » des génotypes étudiés qui est utilisé pour la détection de QTL. Dans l'ensemble des modèles les facteurs série (S_k), isolat (I_s), croisement (C_j) interviendront dans la modélisation du phénotype. La modélisation du phénotype a été réalisée sur R.studio via le package ASREML-R et est décrite dans les paragraphes suivants (Butler *et al.* 2007).

2.2.4. Mod_A , Mod_B et Mod_C effet de l'apparentement sur la prédiction des BLUP

Dans le modèle Mod_A y_{jks} correspondant à la proportion de plants infectés par le *Ganoderma* est une variable suivant une loi binomiale (infecté ou non). Un modèle linéaire mixte généralisé est alors utilisé pour prédire la valeur génétique des parents, prenant en compte la nature non normale des données avec pour fonction de lien, la fonction logit. Le modèle animal est utilisé ici.

Soit y_{jks} la proportion de plants infectés par le *Ganoderma*, pour le croisement j , de la série k , inoculé avec l'isolat s .

$$y_{jks} \sim \beta(N_j, \pi)$$

N_j représente le nombre de plants évalué pour le croisement j . En théorie N_j égale 100 et π la probabilité d'être infecté.

$$\eta_{jks} = \log\left(\frac{\pi_{jks}}{1-\pi_{jks}}\right) \text{ (Eq.1)}$$

$$\eta_{jks} = \mu + I_s + (S_k) + (C_j) + \varepsilon_{ijkl} \text{ (Eq.2)}$$

Où :

- I_s désigne l'effet fixe associé à l'isolat s
- S_k désigne l'effet aléatoire associé à la série k , $S_k \sim N(0, \sigma_s^2)$
- C_j désigne l'effet aléatoire associé au croisement j , $C_j \sim N(0, A\sigma_c^2)$

A désigne la matrice d'apparentement généalogique reliant les plants de palmiers observés et leurs ascendants présents dans le pedigree. Les éléments de la matrice A sont les coefficients de parentés

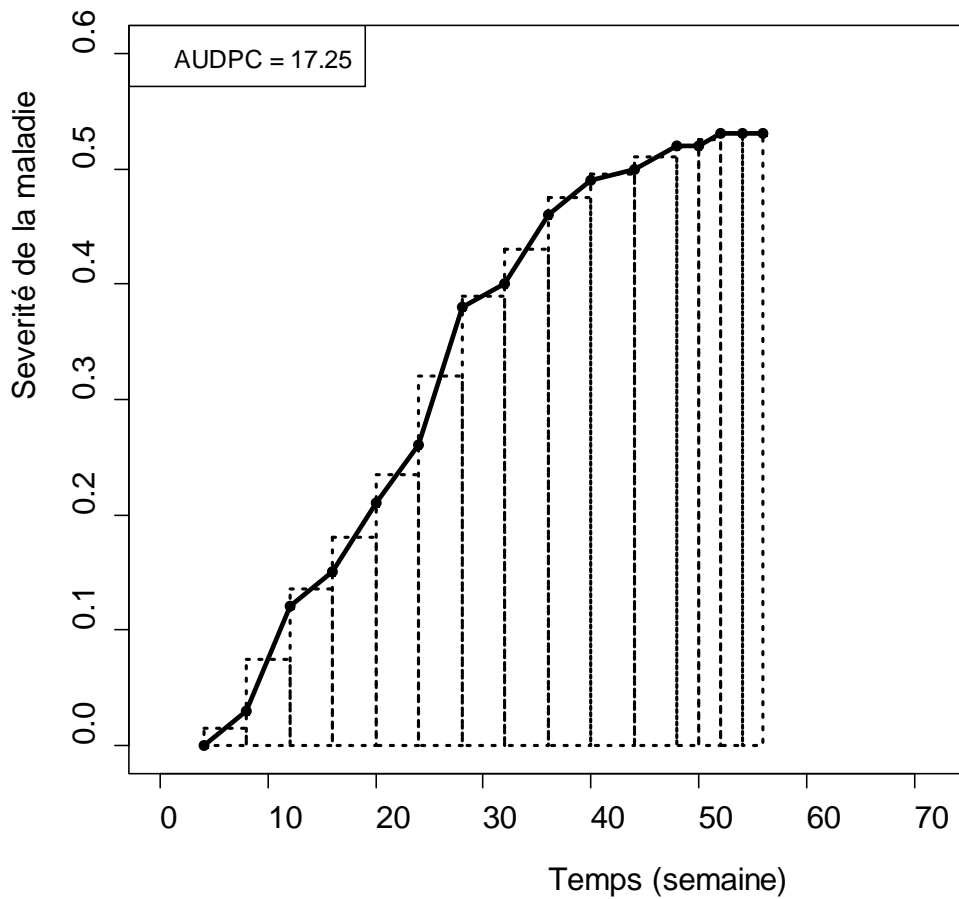


Figure 10 : Représentation graphique de la modélisation de l'aire sous la courbe de progression de la maladie (AUDPC). L'aire de chaque rectangle est équivalente à l'aire du trapèze (en dessous de la courbe) à un même pas de temps puisque l'aire du triangle au-dessus de la courbe est égale à l'aire du triangle en dessous de la courbe. Figure réalisée selon Simko, Piepho, (2012) avec les données de la présente étude.

entre individus. Ainsi la non-indépendance des individus ayant des ancêtres communs est prise en compte dans ce modèle.

De la même manière que précédemment, pour les modèles Mod_B et Mod_C, la modélisation du phénotype passe par l'utilisation d'un modèle linéaire mixte généralisé pour une variable suivant une loi binomiale de paramètres N_j et π. La même fonction de lien est utilisée pour relier les données phénotypiques aux probabilités d'être infecté. Le modèle *Sire and Dam* est ici pris en compte.

Soit y_{abjks} la proportion de plants infectés par le *Ganoderma*, pour le croisement j, entre le parent a et le parent b, de la série k, inoculé avec l'isolat s.

$$\eta_{abjks} = \mu + I_s + (S_k) + (C_j) + (Pa) + (Pb) + \varepsilon_{ijklts} \text{ (Eq.3)}$$

Les facteurs I_s et S_k sont les mêmes que précédemment.

Pour le modèle mod_B :

- C_j désigne l'effet aléatoire associé au croisement j, C_j ~ N(0, Iσ_c²)
- P_a désigne l'effet aléatoire associé au parent a, P_a ~ N(0, Aσ_{Pa}²)
- P_b désigne l'effet aléatoire associé au parent b, P_b ~ N(0, Iσ_{Pb}²)

A désigne la matrice d'apparentement généalogique reliant les plants de palmiers observés du groupe A et leurs ascendants présents dans le pedigree (du groupe A uniquement). Le but de ce modèle est d'expliquer au plus juste la variabilité induite par l'effet du parent A, et de considérer l'ensemble des parents B comme indépendant.

Pour le modèle mod_C :

- C_j désigne l'effet aléatoire associé au croisement j, C_j ~ N(0, Iσ_c²)
- P_a désigne l'effet aléatoire associé au parent a, P_a ~ N(0, Iσ_{Pa}²)
- P_b désigne l'effet aléatoire associé au parent b, P_b ~ N(0, Iσ_{Pb}²)

L'apparentement entre individu n'est pas considéré, les parents sont jugés comme indépendants. C'est le modèle le plus simple de l'étude.

2.2.5. Mod_D : prise en compte de la cinétique de la maladie par l'AUDPC

L'AUDPC ou l'aire sous la courbe de progression de la maladie (*Area Under the Disease Progress Curve*) est une méthode qui permet de prendre en compte l'aspect cinétique de la progression de la maladie selon une méthode dite « méthode trapézoïdale ». Cette méthode consiste à approximer l'aire sous la courbe à chaque pas de temps de la cinétique par l'aire d'un rectangle. L'aire de ce rectangle est représentée par la durée du pas de temps (en semaine) et par la sévérité moyenne de la maladie du génotype considéré au pas de temps donné (Madden *et al.* 2017). La valeur d'AUDPC totale représente la somme de l'aire de chaque rectangle à chaque pas de temps (Figure 10).

Sachant que l'apparition de jeunes palmiers symptomatiques a été suivie dans le temps pour chaque croisement, il est possible de réaliser la cinétique de la maladie. L'AUDPC illustre la cinétique de progression de la maladie (en proportion de jeune palmier symptomatique PPS) entre le moment de l'inoculation de la graine de palmier à huile et la dernière mesure.

Soit A(t_k), l'AUDPC au temps t = t_k, représentant la PPS accumulé jusqu'à t = t_k :

$$A_k = \sum_{i=1}^k \frac{(y_{i-1} + y_i)}{2} (t_i - t_{i-1}) \text{ (Eq. 4)}$$

Où y_i représente la PPS accumulé à la i^{ème} observation. Dans la mesure où les notations n'ont pas été effectués aux mêmes temps, ni pendant la même durée pour chaque série, l'AUDPC A_k a été normalisée et corrigée.

$$AUDPC_{cc} = \frac{A_k}{c*n} \text{ (Eq.5)}$$

Où c représente la durée de l'essai le plus long et n la durée de l'essai du croisement considéré.

L'AUDPC a été utilisé comme donnée de phénotypage et puisque l'individu statistique est le croisement, il a fallu modéliser les données pour obtenir la valeur génétique associé aux parents B. Un modèle linéaire mixte généralisé a été utilisé pour cette modélisation.

Soit y_{abjks} l'AUDPCcc, du croisement j, entre le parent a et le parent b, de la série k, inoculé avec l'isolat s.

$$y_{abjks} = \mu + I_s + (S_k) + (C_j) + (Pa) + (Pb) + \varepsilon_{abjks} \text{ (Eq.5)}$$

$$y_{abjks} \sim N(0, \sigma^2)$$

L'apparentement entre individus n'est pas considéré ici, les parents sont considérés comme indépendant.

2.2.6. Estimation des composantes de la variance

A l'échelle d'une population, il est intéressant d'estimer tous les paramètres qui caractérisent la transmissibilité d'un caractère à la descendance, soit les différentes composantes de la variance génétique, ainsi que les héritabilités.

Les modèles utilisés font intervenir les effet aléatoires associés à chacun des deux parents et à leur interaction. Ces effets sont respectivement des estimateurs des AGC (aptitude générale à la combinaison) et des ASC (aptitude spécifique à la combinaison), sachant que chaque géniteur est évalué par la sensibilité de ses descendants au *Ganoderma*. L'AGC ou l'effet gamétique moyen transmis par le géniteur à la descendance est égale à la moitié des effets additifs des locus plus le quart des interactions simples additive x additive entre les locus (Isik *et al.* 2017).

Soit un géniteur si on considère n locus bi alléliques:

$$AGC = \frac{1}{2}A + \left(\frac{1}{2}\right)^2 A \times A + \dots + \left(\frac{1}{2}\right)^n A \times A \times A \text{ (n fois)} \text{ (Eq.6)}$$

Comme les croisements sont réalisés entre les groupes A et B, comprenant chacun des individus fortement apparentés, il peut alors être écrit que la variance génétique totale est égale à la variance des AGC plus la variance de l'ASC.

Ainsi seule l'héritabilité au sens large pourra être calculée pour les modèles faisant intervenir l'effet des parents (modèle sire and dam).

$$h_{sl}^2 = \frac{V_g}{V_p} = \frac{\sigma_{AGC_{pA}}^2 + \sigma_{AGC_{pB}}^2 + \sigma_{ASC}^2}{\sigma_k^2 + \sigma_{AGC_{pA}}^2 + \sigma_{AGC_{pB}}^2 + \sigma_{ASC}^2 + \sigma_e^2} = \frac{\sigma_A^2 + \sigma_B^2 + \sigma_j^2}{\sigma_A^2 + \sigma_B^2 + \sigma_j^2 + \sigma_e^2} \text{ (Eq.7)}$$

L'héritabilité au sens strict a été calculée pour le modèle animal dont la variance de l'effet aléatoire associé au pedigree correspond à la variance additive du modèle (Kruuk 2004).

$$h_{ss}^2 = \frac{\sigma_j^2}{\sigma_j^2 + \sigma_e^2} = \frac{V_a}{V_p} \text{ (Eq. 8)}$$

2.3. Données moléculaires

Le CIRAD possède des chambres froides de stockage des ADN de plantes d'intérêts récoltés lors de missions antérieures. Les ADN peuvent être conservés à plus ou moins long terme selon qu'ils soient sous forme de suspension, feuille sèche ou broyat. L'ADN est extrait à partir de broyat de feuille lyophilisée de palmier à huile en suivant le protocole d'extraction MATAB du CIRAD (Annexe V) qui permet d'obtenir une suspension d'ADN mère à 50ng/µl.

La population génotypée se trouve être formée de 335 géniteurs d'intérêt du groupe B, parents de croisement évalués dans les essais en pépinière. Les individus sélectionnés ont été génotypés avec 199 marqueurs microsatellites (SSR, *simple sequence repeat*): 177 marqueurs SSR ont été développés par le CIRAD (Billotte *et al.* 2005), 4 marqueurs SSR ont été développés par *the Malaysian Palm Oil Board* (MPOB) (Zaki *et al.* 2012) et 18 marqueurs microsatellites ETSs *expressed sequence tags* ont été développés par l'Institut de Recherche pour le Développement (IRD) et le CIRAD (Tranbarger *et al.* 2012). Les marqueurs ont été choisis à la suite de l'étude de (Cochard *et al.* 2015) pour couvrir l'ensemble du génome avec une densité homogène.

L'amplification des fragments microsatellites a été réalisée sur plaque 384 en multiplex avec 5 ng d'ADN dans un volume final de 10 µl contenant 0.5 µl de Taq polymérase maison (2U), 1µl de Tampon Tris (10X, pH = 8.3), 1µl de dNTP (2 mM), 0.12 µl de MgCl₂ (50 mM), 0.08 µl d'amorce sens possédant une queue complémentaire à M13 (10µM), 0.1 µl d'amorce antisens (10 µM), 0.1 µl d'amorce M13 à fluorescence marqué (dyes : FAM, VIC, PET, NED) et 2.1µl d'eau. Les conditions de la PCR suivies sont celles mentionnés par Tisné *et al.* 2015 (Applied Biosystem, USA). Les produits PCR ont été rassemblés dans un même puits pour un même ADN, 2µl du produit a été mélangé à 10 µl avec un marqueur de taille standard (GeneScan™ 600 LIZ ® à 0.0012 %) dans formamide Hi-Di™ pour migration dans un 3500xL Genetic Analyser (Applied Biosystems, USA). Le protocole d'extraction, d'amplification et de révélation de l'ADN correspond à celui utilisé par (Cochard *et al.* 2015; Tisné *et al.* 2015).

De la même manière que dans les articles cités précédemment, le logiciel GeneMapper© V4.1 (Applied Biosystems, USA) a été utilisé pour déterminer la taille des allèles à chaque marqueur. Via l'interface de ce logiciel les allèles de la population ont été déterminés pour chaque marqueur. A l'issue de cette étape, les données ont été soumises à une Macro « maison » réalisée via Excel, qui permet d'identifier les allèles non-légitimes de la descendance en prenant en compte les allèles des parents à chaque marqueur. Ainsi une descendance non légitime a été écartée (10 individus retirés) et les erreurs de génotypage et/ou de lecture ont été corrigées.

2.4. Construction de la carte génétique

Une carte génétique intégrée basée sur le pedigree des deux groupes hétérotiques A et B du palmier à huile a été réalisée précédemment au CIRAD (Cochard *et al.* 2015). La même méthode a été employée ici pour la réalisation de la carte génétique du groupe B et plus spécifiquement des sous-populations La Mé et Yangambi étudiées ici.

Au total trois cartes génétiques ont été réalisées, une carte intégrée de la population d'étude contenant les sous-populations La Mé et Yangambi, et une carte pour chaque sous population.

La carte génétique a été calculée par le logiciel CRI-MAP version 2.4 (Green *et al.* 1990). CRI-MAP permet la construction d'une carte génétique multilocus en prenant en compte les relations d'apparentement entre individus.

Dans ce cas le pedigree La Mé compte 239 individus apparentés et le pedigree Yangambi 95 individus également apparentés. Les marqueurs dont l'emplacement était déjà connu ont été rassemblés sur leur groupe de liaison correspondant (Billotte *et al.* 2010; Cochard *et al.* 2015). Les marqueurs pour lesquels il n'y avait pas d'information ont été assignés à un groupe de liaison par l'option TWOPOINT de CRI-MAP. Cette option calcule les taux de recombinaison et les LOD score entre toutes les paires de marqueurs. Ainsi pour un LOD score supérieur à 3, c'est-à-dire pour une probabilité 1000 fois supérieur d'avoir deux marqueurs liés que deux marqueurs non liés, les

marqueurs encore non étudiés ont pu être ajoutés à un groupe de liaison. Les marqueurs ont été ordonnés dans les groupes de liaison en utilisant la commande BUILD qui ordonne les marqueurs les plus informatifs. Les marqueurs restants ont été placés avec la commande ALL en testant pour chacun toutes les positions possibles. Des permutations entre 2 ou plusieurs marqueurs ont été réalisées avec la commande FLIPS afin de mesurer la vraisemblance d'ordres alternatifs. Les distances génétiques de Kosambi en centimorgans entre deux marqueurs ont été calculées par la commande FIXED pour obtenir la carte finale. La carte par groupe de liaison a été représentée par le logiciel MapChart 2.0 (Voorrips 2002).

2.5. Détection de QTL

L'intérêt d'utiliser la méthode Bayésienne implémentée dans le logiciel FlexQTLTM pour la détection de QTL réside dans le fait que c'est une méthode qui prend en compte l'apparement et qui propose les modèles multi QTL les plus probables à la suite d'un grand nombre de test de modèle QTL différents.

La cartographie de QTL a été réalisée selon une méthode bayésienne utilisant l'algorithme de Monte Carlo par Chaîne de Markov (MCMC) qui permet d'augmenter le pouvoir d'estimation de l'emplacement du QTL sur le chromosome et la contribution des gènes affectant les traits mesurés. Ces méthodologies d'abord développées pour l'étude de caractères quantitatifs chez l'humain s'appliquent bien à l'étude de caractères agronomique chez les plantes pérennes. Cette analyse a été réalisée par l'interface de FlexQTLTM qui est un logiciel sous ligne de commande linux développé par (Bink *et al.* 2002) (Biometris, Wageningen UR).

Les valeurs des BLUP prédits précédemment sont utilisées comme phénotypes dans cette partie pour la modélisation des QTL. Cette approche Bayésienne permet l'exploration de différents modèles QTL expliquant la variabilité phénotypique. Le nombre de QTL est traité en effet aléatoire et leur distribution est estimée (Bink *et al.* 2014). L'analyse débute avec un jeu de paramètres prédéfini en amont (Annexe VI) et les données de génotypage, le phénotype associé et la carte génétique des marqueurs. Chaque analyse comprend un grand nombre d'itérations testant un maximum de combinaisons de régions QTL et leur probabilité associée jusqu'à convergence vers les régions QTL les plus probables.

Deux analyses distinctes sont réalisées pour un même set de données (deux pour chacun des modèles linéaires mixtes généralisés utilisés) en utilisant une graine de départ différente. Aussi, les sous-groupes de l'étude ont été considérés ensemble et indépendamment dans l'analyse QTL en séparant les données La Mé des Yangambi. Une région QTL a été identifiée lorsque dans deux analyses une même région a obtenu un facteur de Bayes supérieur à 2 ($2 \cdot \ln BF > 2$), similaire à un dépassement du seuil de vraisemblance (Bink *et al.* 2002, 2008, 2014). La position la plus probable a alors été déterminée en faisant la médiane de chaque analyse significative au locus considéré (Howard *et al.* 2018).

Le modèle bayésien considère chaque QTL comme étant biallélique, ainsi 3 génotypes se distinguent, QQ, Qq et qq ayant pour valeurs génotypiques respectives +a, d et -a. Le phasage des marqueurs est réalisé par FlexQTLTM à l'aide du pedigree et permet d'obtenir l'identité des allèles aux marqueurs à savoir l'identité par descendance (IBD : Identity by descent). Cette étape de phasage permet de prédire les allèles aux QTL en fonction du fond génétique considéré (Howard *et al.* 2017, 2018).

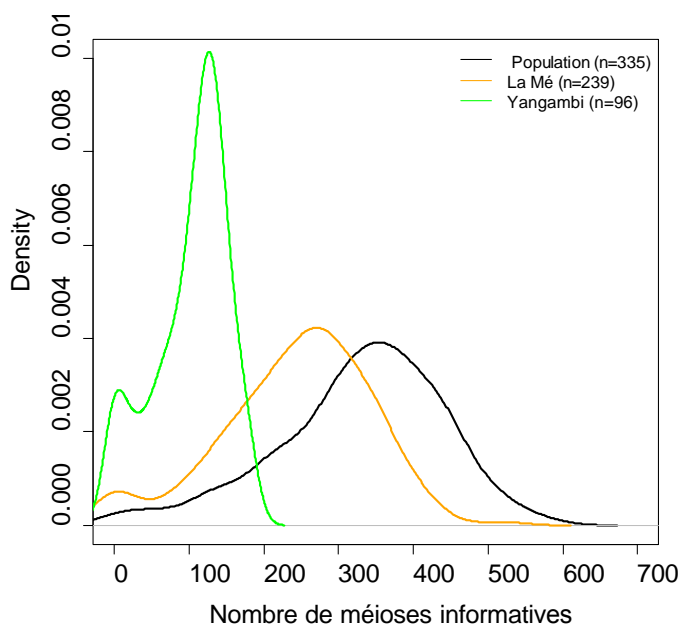


Figure 11 : Méioses informatives moyennes par marqueurs. Distribution des méioses informatives moyennes par marqueurs dans la population totale en noire, la sous population La Mé en orange et la sous population Yangambi en vert. L'effectif des populations est mentionné dans l'encadré en légende.

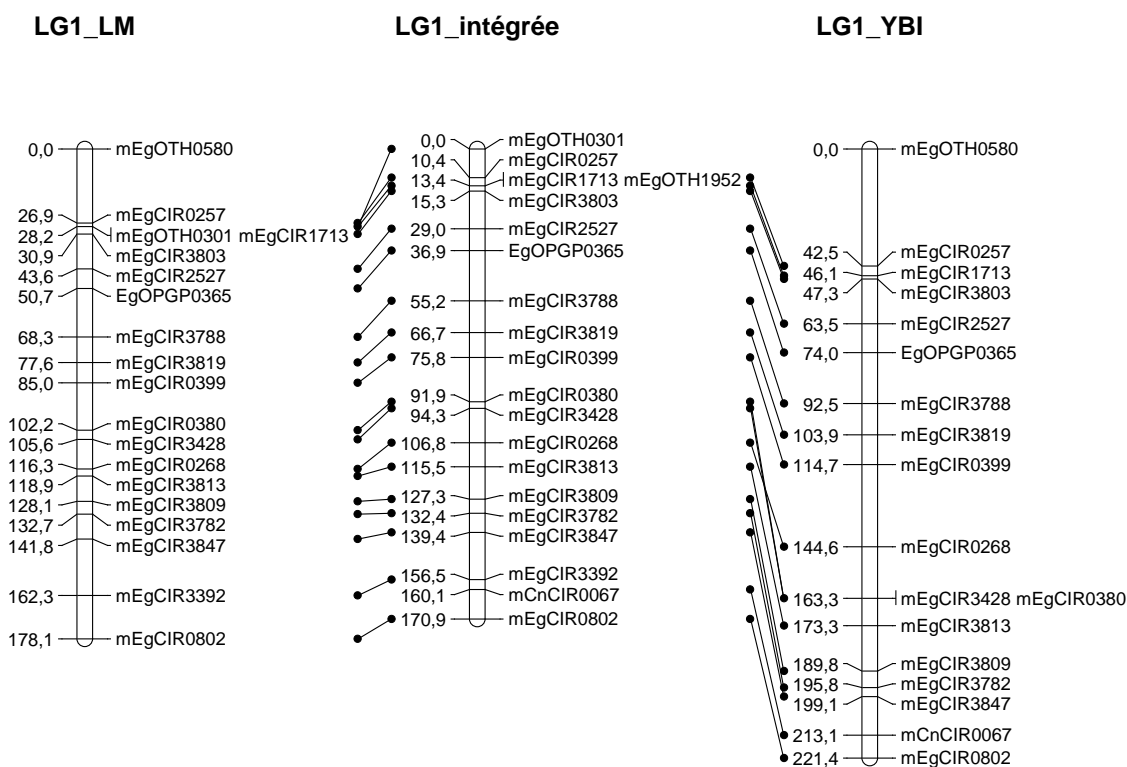


Figure 12 : Carte génétique des pedigrees La Mé, Yangambi et intégrée. Carte génétique du groupe de liaison 1 de la population de palmier à huile étudiée. Trois cartographies sont visibles, à gauche la carte propre aux individus du pedigree La Mé, à droite la cartographie propre aux individus du groupe Yangambi et au centre la cartographie intégrée des deux pedigrees.

3. Résultats

3.1. Polymorphisme des marqueurs et cartographie génétique

Les deux pedigrees d'origine génétique différentes constituant la population d'étude ont été analysés. La richesse allélique et le nombre de méioses informatives ont été calculés dans la population totale et dans chaque sous-groupe. Un jeu de 199 marqueurs microsatellites a été utilisé pour le génotypage des 335 palmiers. En moyenne 4,62 allèles ont été observés pour ces 199 marqueurs dans la population étudiée. Au total, 2 à 8 allèles sont observés par marqueurs. Au sein des pedigrees La Mé et Yangambi de l'étude sont observés respectivement 2.36 et 2.81 allèles en moyenne par marqueur. Le nombre moyen de méioses informatives en considérant les deux sous-populations étudiées est de 328 et est plus élevé dans le pedigree La Mé que dans le pedigree Yangambi en accord avec le fait qu'il y ait moins d'individu chez les Yangambi. Néanmoins, le nombre de méioses informatives est très variable d'un marqueur à l'autre Figure 11.

A partir des 199 marqueurs disponibles pour réaliser la cartographie, 190 ont pu être utilisés pour la cartographie de la population complète, 181 pour la cartographie du pedigree La Mé et 171 pour la cartographie du pedigree Yangambi. Parmi les marqueurs non utilisés, 2 marqueurs présentaient un LODscore toujours inférieur à 3 lors des calculs par la fonction TWOPOINT et le reste des marqueurs était monomorphe dans la population d'étude.

La carte génétique montre 16 groupes de liaisons pour les deux sous-populations, obtenues en utilisant les informations de recombinaison sur tout le pedigree (Figure 12, Annexe VII). Les trois cartes génétiques obtenues ont des tailles comparables, 1931 cM pour la population totale, 1802 cM pour le pedigree La Mé et 1908 cM pour le pedigree Yangambi. La densité moyenne des marqueurs est d'un marqueur tous les 10,2 cM, avec une densité supérieure dans le pedigree La Mé que Yangambi (respectivement tous les 10 cM et tous les 11,22 cM).

Les cartes obtenues pour les deux pedigrees de l'étude ont été comparées à la carte génétique du groupe B de (Cochard *et al.* 2015). Dans la présente étude, la carte couvre 1931 cM contre 1845 pour la carte comparée. 64 % des marqueurs de la carte comparée ont été utilisés ici pour la cartographie des sous-groupe La Mé et Yangambi.

L'ordre des marqueurs sur les 16 groupes de liaison est globalement conservé sauf dans le cas des groupes de liaisons 3, 6, 8, 9 et 13 où il est noté soit une inversion entre plusieurs marqueurs, soit un intervalle plus grand entre deux marqueurs.

3.2. Analyse des données brutes de phénotypage

La résistance à la pourriture basale du stipe a été mesurée en moyenne pendant 34,2 semaines sur 3910 croisements A x B. 361 parents du groupes B dont 258 génotypés ont été suivis en croisement avec 2069 parents du groupe A. En moyenne les parents du groupe B sont engagés dans 85 croisements différents (au minimum dans un croisement, au maximum dans 206 croisements) (Figure 13).

La distribution de la proportion de jeunes palmiers atteints par le *Ganoderma* montre qu'à la fin des essais en moyenne 30 % des individus sont symptomatiques avec un minimum 0% et un maximum 92,5% par croisement. (Figure 14). Le calcul de l'AUDPC est utilisé comme données de phénotypage en plus de la proportion de jeunes palmiers atteint par le *Ganoderma*.

La représentation graphique de la cinétique des quatre valeurs d'AUDPC les plus faibles, les intermédiaires et les plus fortes permet de visualiser différents profils cinétiques (Figure 14). Les croisements les plus résistants auront un profil cinétique semblable à ceux représentés par la Figure 14A, avec l'apparition des premiers symptômes tardivement (à plus de 15 semaines), et une faible occurrence de la maladie tout au long de la durée de l'essai. Les croisements moins tolérants à la maladie auront un profil cinétique semblable à ceux représentés par la Figure 14B, avec l'apparition des premiers symptômes autour de 10 semaines et une progression du nombre de plants atteints par

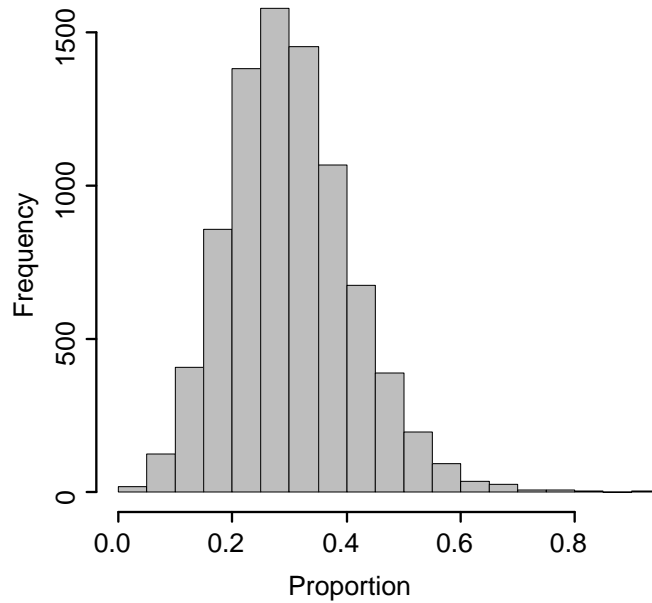


Figure 13 : Distribution de la proportion d'individus atteints par le *Ganoderma*. Exploration de 10 ans de données de phénotypage issues de tests précoces en pépinière, 3910 croisements A x B ont été considérés.

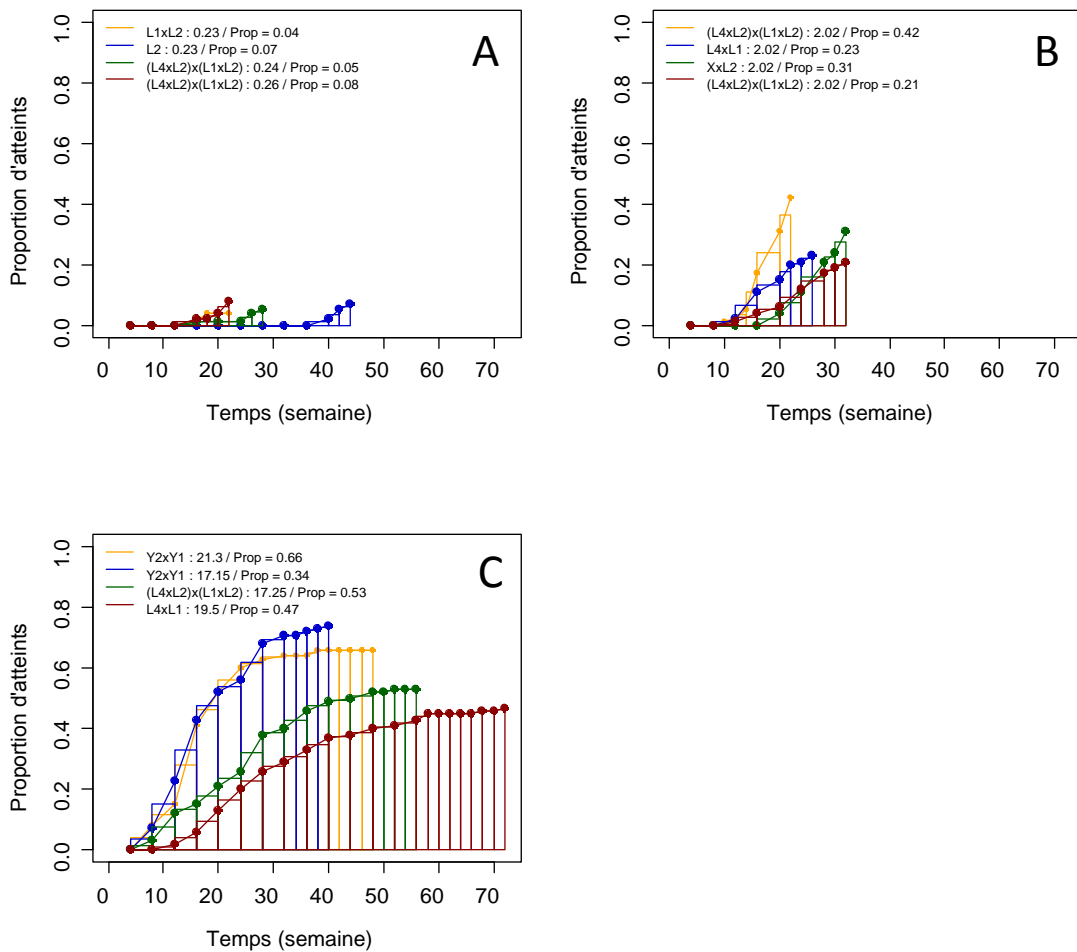


Figure 14 : Cinétique et AUDPC. Cinétique des quatre croisements ayant les valeurs les plus faibles d'AUDPC (A), les valeurs médianes d'AUDPC (B) et les valeurs les plus forte d'AUDPC (C).

le *Ganoderma* tout au long de l'essai. Les croisements les plus sensibles (figure 14C) développent des symptômes précocement (5 semaines) et comptent le plus de palmiers atteints par la maladie. Lorsque les valeurs d'AUDPC sont observées pour chaque type de cinétique, les valeurs les plus faibles d'AUDPC sont associées à une faible proportion de malades en fin de tests et inversement. Cependant, une même valeur d'AUDPC peut correspondre à plusieurs valeurs de proportion de malades en fin de test. Par exemple pour le croisement ayant pour origine de croisement du parent B (L4xL2) x (L1xL2) (figure 14B) sa valeur d'AUDPC est de 2.02 et sa PPS en fin de test est de 0.21, pour la même valeur d'AUDPC, le croisement dont le parent B a la même origine a une PPS en fin de test de 0.42.

3.3. Modélisation différentielle du phénotype

Pour visualiser l'effet des génotypes parentaux, la distribution des valeurs génétiques par famille et par modèle a été réalisée (Figure 15). Indépendamment du modèle utilisé, les croisements ayant pour origine le fondateur Y1 du pedigree Yangambi ont les valeurs génétiques les plus élevées, et sont donc plus les plus sensibles. De la même façon, les croisements ayant pour origine le fondateur L1 du pedigree La Mé ont les valeurs génétiques les plus faibles et sont donc les plus résistants.

L'estimation des composantes de la variance a été réalisée pour chaque modèle utilisé. L'héritabilité au sens strict du modèle mod_A est 0.25. L'héritabilité au sens large estimée par les modèles mod_B , mod_C et mod_D est égale à 0.09, 0.11 et 0.22 respectivement (Tableau I). Le modèle animal incluant l'apparentement entre individus estime une variance pour la composante du croisement plus forte qu'avec les autres modèles. Les modèles « *Sire and Dam* » mod_B et mod_C , sans structure d'apparentement sur la composante du groupe B, estiment une variance plus faible pour l'effet de ce groupe génétique.

3.4. Modèles QTL, génotypes et effets associés aux QTL

Pour l'inférence du nombre de QTL, les QTL dépassant fortement le seuil de présence ($2\ln BF > 5$) et les QTL dépassant simplement le seuil de présence ($2\ln BF > 2$) n'ont pas été distingués. Tous les QTL mis en évidence dans deux analyses sur le phénotype issu du même modèle et dépassant le seuil de présence ($2\ln BF > 2$) ont été considérés (Tableau II). Au total, 10 régions QTL ont été identifiées toutes analyses confondues (tableau II), les QTL ayant un intervalle de confiance qui se chevauche étant considérés identiques. Six régions QTL ont été exclusivement trouvées chez les LM, 2 QTL chez les YBI et 2 QTL lors de l'analyse conjointe des deux jeux de données. Ces régions QTL sont distribuées sur 6 groupes de liaison avec les groupes de liaison 5, 10 et 13 qui comptent deux régions QTL. Seuls les régions QTL mises en évidence sur le groupe de liaison 13 ont un $2\ln BF > 5$, dépassent fortement le seuil de présence. Le modèle mod_A permet de détecter 8 QTL sur les 10, soit les QTL 5@17.94, 8@164.39, 9@95.31, 10@16.93, 10@138.96, 13@88.45, 13@27.92 et 16@10.49. Le modèle mod_B détecte une seule région QTL soit 10@94.66 qui n'a pas été mise en évidence par mod_A . Les modèles mod_C et mod_D mettent chacun en évidence deux QTL dont un en commun avec le mod_A , soit 9@88.55 et 13@90.85 pour mod_C et 5@9.46 et 13@87.87 pour mod_D . Le profil du modèle mod_B montre la faiblesse de sa capacité d'estimation d'un modèle QTL (Figure 16 montrant une seule analyse par modèle, annexe VIII). La région QTL mise en évidence par ce modèle n'est mise en évidence dans aucun autre. C'est le modèle pour lequel la valeur de l'héritabilité est la plus faible. Les profils des modèles mod_C et mod_D sont assez similaires, mais la probabilité de présence des QTL du mod_D est deux fois supérieure à la probabilité de présence du mod_C (Figure 16). La même relation est observée pour leur héritabilité respective.

Les données comprennent 48 familles interconnectées de plein-frères et de demi-frères qui dérivent de croisement entre 7 fondateurs. La présence d'un QTL est obtenue par la ségrégation des allèles d'intérêts dans une ou plusieurs familles. En conséquence, le génotype inféré au QTL au sein des individus du pedigree a permis de prédire le génotype au QTL des fondateurs en utilisant les informations d'IBD prédites par le modèle (Figure 17).

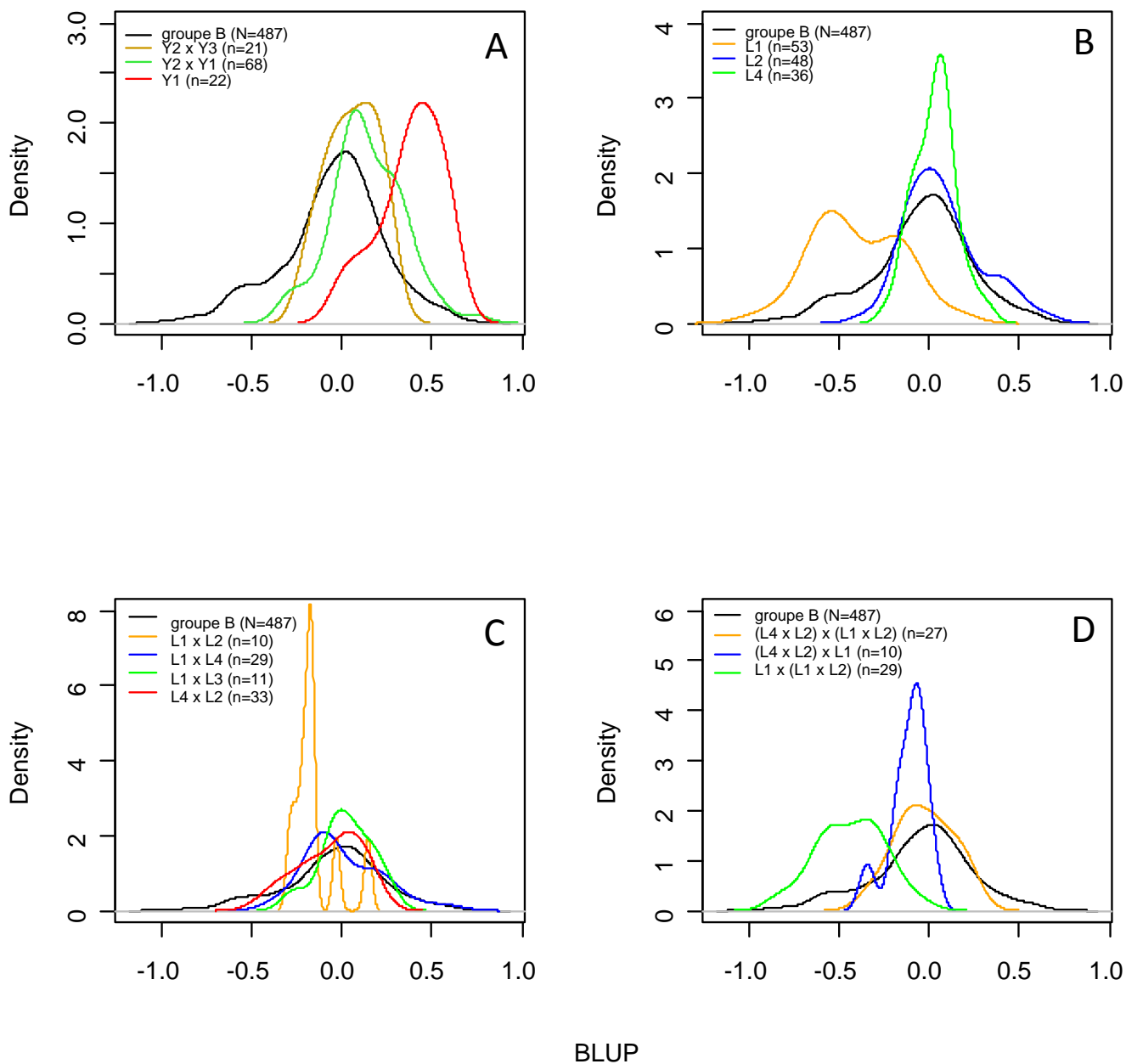


Figure 15 : Distribution des BLUP par origine de croisement. Chaque croisement a été considéré selon son origine génétique ce qui permet de visualiser directement l'impacte des fondateurs des pedigree. (A) recombinaisons entre fondateurs du pedigree Yangambi, (B) autofécondations des fondateurs La Mé, (C) et (D) recombinaisons entre fondateurs La Mé.

Le génotype le plus fréquent au QTL est « QQ » chez les fondateurs, soit associés au phénotype sensible. La forme homozygote « qq » est la moins fréquente, seuls trois fondateurs possèdent au moins une fois ce génotype. La forme hétérozygote « Qq » est bien présente pour les QTL 5@17.94, 8@164.39, 9@95.31, 10@16.93, complètement absente pour le QTL 10@138.96 et complètement absente chez les fondateurs Yangambi (Y1, Y2 et Y3) qui ont l'allèle « Q » fixé pour les QTL 13@27.92, 13@88.45, 16@10.49.

Pour chaque individu de la population étudiée, en associant son génotype au QTL avec sa valeur génétique, il peut être étudié l'effet moyen de chaque génotype au QTL (Figure 18, Annexe IX). Le génotype « qq » est associé aux valeurs génétiques les plus basse et inversement pour le génotype « QQ ».

L'allèle d'intérêt pour le QTL 13@88.45, le QTL majeur de l'étude, semble être associé à une diminution nette de la valeur génétique du parent B et provient du parent L1 du sous-groupe La Mé.

Lors de l'étude des haplotypes des marqueurs dans la région QTL du groupe de liaison 13 du parent L1, deux haplotypes sont observés H_0 et H_1 . Les descendants de L1 par autofécondation ont alors obtenu soit deux copies de H_0 (H_0/H_0), soit deux copies de H_1 (H_1/H_1), soit une copie des deux (H_0/H_1), soit constitués de gamètes recombinés (R) (Figure 19A et B).

L'association de la valeur génétique des descendants avec leur haplotype respectif permet de visualiser l'effet des haplotypes parentaux, des hétérozygotes et des recombinants (Figure 19 C).

L'haplotype H_1 est associé à une diminution de la valeur génétique de l'individu porteur, en considérant une fois de plus les valeurs génétiques du modèle mod_C , soit du cas le plus simple.

Cet haplotype sera, entre autres, lié à la diminution de la sensibilité des descendants ayant obtenus une copie de cet haplotype de L1.

4. Discussion

4.1. Cartographie génétique d'une population en pedigree

Les logiciels de cartographies génétiques sont généralement destinés aux populations biparentales en ségrégation telles que F2, backcross ou encore RIL (*recombinant imbreed line*). Le logiciel JoinMap (Ooijen and Voorrips 2002) par exemple, permet la réalisation de cartes génétiques dans de telles populations, ou de cartes génétiques consensus pour des populations multiparentales constituées de grandes familles. Dans le cadre du palmier à huile, les contraintes biologiques rendent impossible l'obtention de telles populations. Ainsi l'étude des populations en pedigree à faible effectif par famille doit passer par l'étude de la ségrégation des allèles au sein du pedigree. La réalisation d'une carte génétique par origine chez le palmier à huile est permise par l'utilisation d'un logiciel spécifique qui intègre les informations d'apparentements pour étudier la ségrégation des allèles. Le logiciel CRI-MAP a été développé par Green et al., (1990) pour la création de carte génétique dans des populations humaines et animales constituées de plusieurs familles plus ou moins interconnectées et ayant chacune un effectif limité.

Dans cette étude, une carte génétique a été construite pour chaque groupe de palmiers étudiés, soit des palmiers ayant pour origine La Mé, soit Yangambi. Ceux-ci sont organisés en familles de demi-frères et plein-frères dans un pedigree complexe qui est utilisé en améliorations variétales. Moins de marqueurs se sont avérés informatifs chez les Yangambi mais la richesse allélique moyenne des deux sous-groupes est sensiblement la même. Les géniteurs du groupe Yangambi sont fortement apparentés : le pedigree démarre par quatre fondateurs qui ont été croisés les uns les autres pendant plusieurs générations pour donner les trois géniteurs (Y1, Y2 et Y3) à l'origine du sous-groupe Yangambi de l'étude. Contrairement au pedigree Yangambi, le pedigree La Mé part de deux fondateurs qui ont été croisés une fois pour donner les quatre géniteurs à l'origine du sous-groupe La Mé de l'étude (L1, L2, L3 et L4). Ainsi, les fondateurs du pedigree Yangambi de l'étude sont plus consanguins ($F=0.55$) mais possèdent une base génétique plus grande que ceux de La Mé ($F=0.5$)

Tableau I : Estimation des composantes de la variance. Les variances génétiques dont la variance génétiques additive, la variance des AGC, de l'ASC et la variance phénotypique ont permis d'estimer les héritabilités au sens large et au sens strict.

Modèle	Vg	Va	V _{AGCpB}	V _{AGCpA}	V _{ASC}	Vp	h ² _{sl}	h ² _{ss}
Mod _A	-	0.3445	-	-	-	1.3648	-	0.2524
Mod _B	0.1057	-	0.0228	0.0414	0.0932	1.1816	0.0894	-
Mod _C	0.1382	-	0.0214	0.0271	0.0896	1.1739	0.1177	-
Mod _D	0.0005	-	0.00015	0.00025	0.0001	0.0023	0.2227	-

Tableau II : QTL associés à la résistance au *Ganoderma* dans la population étudiée. Indépendamment du modèle utilisé, les QTL qui ont été considérés avaient un facteur de Bayes supérieur à 2 ($2\ln\text{BF} > 2$).

Groupe de liaison	Population	Modèle	Nom du QTL	Région QTL (cM)	Position (cM)	$2\ln\text{BF} > 2$
5	GB	Mod _A	5@17.94	0.54 – 35.34	17,94	4.1
	YBI	Mod _D	5@9.46	0.52 – 18.40	9,46	3.2
8	LM	Mod _A	8@164.39	151.70 – 177.08	164,39	2.1
9	YBI	Mod _A	9@95.31	86.59 – 104.04	95,31	3.0
	YBI	Mod _C	9@88.5	72.55 – 104.45	88,5	2.6
10	LM	Mod _A	10@16.93	0.52 – 33.34	16,93	4.2
	LM	Mod _A	10@138.96	119.71 – 158.21	138,96	2.2
	LM	Mod _B	10@94.66	67.20 – 122.13	94,66	2.0
13	LM	Mod _A	13@27.92	17.58 – 38.27	27,92	7.55
	LM	Mod _A	13@88.45	60.58 - 118.7	89,64	26.3
	LM	Mod _C	13@88.45	67.66 – 114.05	90,85	6.2
	LM	Mod _D	13@88.45	65.60 – 110.14	87,87	4.0
16	GB	Mod _A	16@10.49	0.62 – 20.37	10,49	3.0

issus de seulement deux fondateurs, ce qui peut expliquer leur plus grande richesse allélique mais leur plus faible nombre de marqueurs polymorphes.

Billotte et al., (2010) et Cochard et al., (2015) ont travaillé sur la carte génétique du palmier à huile soit en créant une carte consensus sur JoinMap à partir de populations à design spécifique (Milan et al 2010) soit en utilisant CRI-MAP respectivement. Une bonne cohérence est observée quant à l'ordre des marqueurs, le nombre de groupes de liaison et la couverture du génome.

Les marqueurs microsatellites sont utilisés pour leur polymorphisme attendu important, ce qui est particulièrement intéressant lors d'un travail sur un pedigree complexe avec beaucoup d'allèles en ségrégation au travers des générations (Cochard *et al.* 2015). L'utilisation de 190 marqueurs informatifs a permis d'étudier la ségrégation des allèles au sein de la descendance, mais leur densité moyenne (un marqueur tous les 10,2 cM) peut s'avérer insuffisante pour capter un maximum de recombinaisons. Ainsi, pour la population d'étude de 335 individus, un marqueur tous les 10 cM représente environ 34 recombinaisons entre deux marqueurs. L'utilisation de données de génotypage haut débit (GBS, SNP array) pourrait être plus informative quant à l'étude de la ségrégation des allèles au sein du pedigree par la densité supérieure des marqueurs obtenus. Cependant la richesse allélique des SSR (*simple sequence repeat*) à un locus donné est plus importante que celle des SNP (*single nucleotide polymorphism*) qui compte au maximum quatre allèles différents. Dans le cas de cette étude, la meilleure alternative serait donc de combiner les informations de SSR et de SNP dans une même carte génétique. Par la suite, les régions QTL pourraient être densifiées en marqueur pour réaliser une cartographie fine des régions d'intérêts et permettre une meilleure précision afin d'identifier les gènes sous-jacents (Tisné *et al.* 2017).

4.2. Utilisation des BLUP pour la cartographie de QTL, approche bayésienne

La génétique quantitative est une discipline qui s'intéresse au déterminisme génétique de traits phénotypiques complexes. L'hypothèse de base est de dire que les traits quantitatifs Gaussiens (normalement distribués) sont sous contrôle polygénique, mais de nombreux traits à distribution discrète peuvent l'être également (de Villemereuil 2018). Pour normaliser les traits non Gaussiens, il existe des fonctions de lien, comme la fonction « logit » qui a été utilisée dans cette étude.

La modélisation des valeurs phénotypiques via un modèle linéaire mixte généralisé (GLMM) permet d'estimer l'effet de chaque variable explicative sur la variabilité du phénotype observée. Une telle analyse est utilisée pour estimer les composantes de la variance et ainsi quantifier la portion héritable du caractère mesuré (Kruuk 2004). L'utilisation d'un modèle animal pour quantifier la valeur génétique individuelle dans le but de cartographier des QTL représente également un point de départ significatif pour de nombreuses études (Marchal et al., 2016 ; Bink et al., 2014). Dans cette analyse les valeurs génétiques, prédites par la méthode des BLUP (Isik *et al.* 2017), ont été utilisées comme phénotype pour la cartographie de QTL car les individus génotypés n'ont pas été directement phénotypés.

Afin de tester l'effet de la prise en compte de l'apparentement dans l'estimation des composantes de la variance et dans la prédiction des modèles QTL, trois modèles intégrant différemment l'apparentement ont été construits. Une plus faible part de variance génétique a été captée dans les deux derniers modèles, qui prennent moins en compte l'apparentement, le mod_A apparaissant le plus précis pour la prédiction des valeurs génétiques. Afin de tester l'effet de la cinétique *via* l'utilisation de l'AUDPC comme valeur phénotypique, le modèle mod_D a été construit avec une prise en compte de l'apparentement similaire au mod_C. Une plus grande part de la variance génétique est prise en compte dans le modèle AUDPC que dans le modèle mod_C.

Ces méthodes contraignent fortement la prédiction des valeurs génétiques pour les individus apparentés (inclusion de matrice de variance-covariance) ce qui est problématique pour l'analyse QTL. Il existe des moyens de dé-régresser les valeurs génétiques obtenues afin d'exploiter toute la

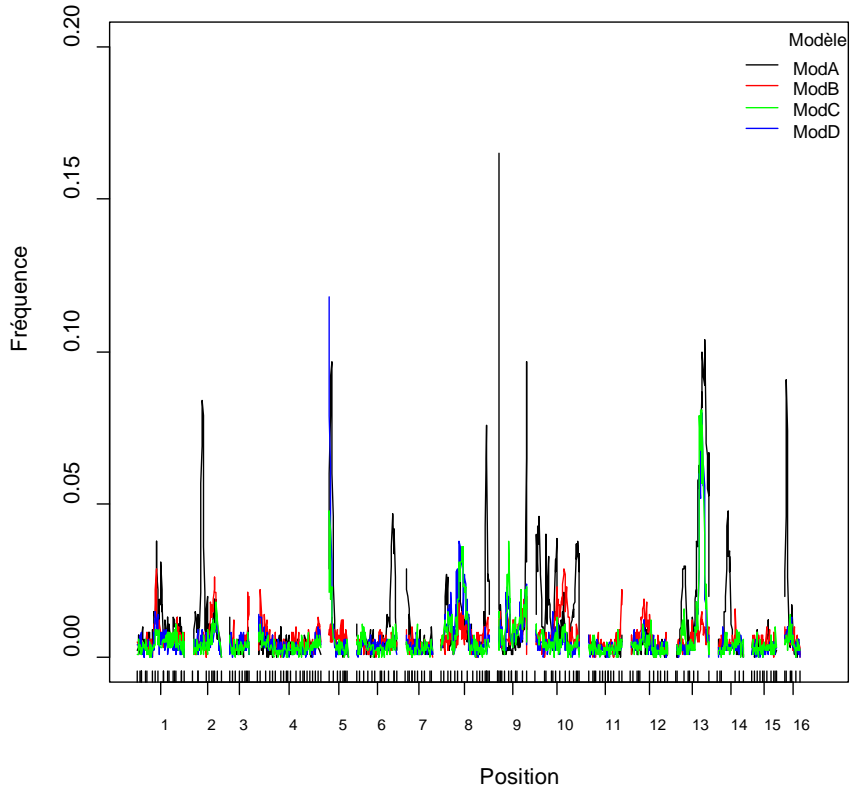


Figure 16 : Profils des modèles QTL. Probabilité à *posteriori* de la position des QTL issues des sorties de FlexQTL™ pour la résistance au *Ganoderma*. Le nom de chromosome est référencé au-dessus des graphiques. La position des marqueurs est annotée par les traits au dessus de l'axe des abscisses. Le numéro associé représente la position de chaque fin de chromosome (en cM).

variabilité associée à l'effet génétique mesuré (Garrick *et al.* 2009). De plus c'est à l'échelle « logit » que les composantes de la variance sont estimées. Récemment, (de Villemereuil 2018) ont mis en évidence un moyen de corriger les valeurs génétiques des modèles linéarisés et contraints dans le cas de données non linéaires. Cette étude propose une méthode pour prédire les valeurs génétiques sans biaiser les prédictions et permet de passer des modèles linéaires mixtes aux données réelles. Ces approches devront être testés dans la continuité de cette étude pour comparer les méthodes d'estimations des composantes de la variance.

Identifier des QTL en utilisant l'appui de l'information moléculaire permet, aux généticiens et aux sélectionneurs, de connaître les régions du génome d'étude impliqués dans la variabilité du trait d'intérêt. Historiquement, la détection de QTL a été initiée par l'analyse de liaison sur des populations biparentales (Würschum 2012). Ces populations étant souvent éloignés d'une population de sélection, d'autres méthodes d'analyses ont été développées. La cartographie génétique de QTL dans une population de sélection offre la possibilité d'accéder à l'architecture génétique du trait d'intérêt dans la population et d'identifier la localisation des allèles d'intérêts (Würschum 2012).

Le logiciel FlexQTL™ développé par (Bink *et al.* 2002) offre la possibilité et la flexibilité d'étudier de nombreuses familles interconnectées en intégrant les effets de pedigree. Ce logiciel utilise une approche bayésienne multi-QTL basée sur le pedigree qui fonctionne pour l'étude de traits agronomiques chez les espèces cultivées (Bink *et al.* 2014; Howard *et al.* 2017; Mangandi *et al.* 2017).

Dans la présente étude qui modélise le phénotype avant l'analyse QTL, l'exploration des modèles QTL sur des données déjà fortement contraintes (mod_A) a été difficile. En effet lors des analyses avec des valeurs initiales différentes, les modèles QTL les plus probables étaient différents, avec peu de régions communes. En revanche, les modèles mod_B , mod_C et mod_D moins précis dans l'estimation de l'effet génétique, ont laissé la possibilité au modèle Bayésien d'explorer un plus grand nombre de modèle QTL, et ont donné des résultats plus robustes. Ainsi, si l'objectif est de prédire au mieux l'effet des géniteurs utilisés en sélection, un modèle intégrant la totalité des effets d'apparentement devrait être réalisé, alors que si l'objectif est de réaliser une cartographie de QTL, une prédiction de l'effet des géniteurs au plus proche de la variation du phénotype devrait être réalisée.

Dans le cas de cette étude où une étape supplémentaire est nécessaire pour obtenir la valeur génétique des individus étudiés, l'utilisation de l'information IBD à chaque position du génome et modélisée en effet aléatoire pourrait être envisagée pour la détection de QTL. Cette approche dite « *two-step variance component* » a été appliquée au palmier à huile par (Tisné *et al.* 2017). Elle permet en une seule étape de relier les données phénotypiques à des effets QTL aléatoires en modélisant les matrices de variance-covariance par les probabilités IBD à chaque locus (George *et al.* 2000).

ModA	Génotype au QTL															
Fondateurs	5@17.94		8@164.39		9@95.31		10@16.93		10@138.96		13@27.92		13@89.64		16@10.49	
L1	Blue	Blue	Grey	Grey	Green	Green	Green	Green	Blue	Blue	Blue	Blue	Green	Green	Green	Green
L2	Orange	Orange	Orange	Orange	Orange	Orange	Green	Green	Blue	Blue	Green	Green	Orange	Orange	Orange	Orange
L3	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Green	Green	Orange	Orange	Grey	Grey
L4	Grey	Grey	Grey	Grey	Orange	Orange	Green	Green	Orange	Orange	Green	Green	Green	Green	Orange	Orange
Y1	Orange	Orange	Orange	Orange	Orange	Orange	Green	Green	Blue	Blue	Orange	Orange	Orange	Orange	Orange	Orange
Y2	Green	Green	Green	Green	Green	Green	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange
Y3	Green	Green	Orange	Orange	Orange	Orange	Orange	Orange	Blue	Blue	Orange	Orange	Orange	Orange	Orange	Orange

ModB	Génotype au QTL	
Fondateurs	10@94.66	
L1	Blue	Blue
L2	Orange	Orange
L3	Grey	Grey
L4	Grey	Grey
Y1	Green	Green
Y2	Blue	Blue
Y3	Blue	Blue

ModC	Génotype au QTL			
Fondateurs	9@88.5		13@90.85	
L1	Orange	Orange	Green	Green
L2	Orange	Orange	Orange	Orange
L3	Orange	Orange	Orange	Orange
L4	Orange	Orange	Orange	Orange
Y1	Orange	Orange	Orange	Orange
Y2	Orange	Orange	Orange	Orange
Y3	Green	Green	Orange	Orange

ModD	Génotype au QTL			
Fondateurs	5@9.46		13@87.87	
L1	Orange	Orange	Green	Green
L2	Orange	Orange	Orange	Orange
L3	Orange	Orange	Orange	Orange
L4	Orange	Orange	Orange	Orange
Y1	Orange	Orange	Orange	Orange
Y2	Green	Green	Orange	Orange
Y3	Blue	Blue	Orange	Orange

Figure 17 : Génotypes estimés des fondateurs aux QTL. Estimation *a posteriori* des génotypes aux QTL pour les fondateurs des pedigree La Mé (L1, L2, L3 et L4) et les fondateurs du pedigree Yangambi (Y1, Y2 et Y3) selon le modèle de prédiction des valeurs génétiques utilisés.

Les individus bleus sont homozygotes favorables, les individus vert sont hétérozygotes et les individus rouges sont homozygotes défavorables.

4.3. Comparaison des résultats avec la littérature

Dans de nombreuses analyses de la tolérance à la maladie pour des populations de plantes cultivées, des gènes de résistance majeurs sont généralement mis en évidence (Ayliffe *et al.* 2008; Laperche *et al.* 2017). Il y a deux types de résistance aux maladies : les résistances complètes conditionnées par un seul gène (résistance qualitative : gène R entraînant une réponse d'hypersensibilité de l'hôte) et les résistances incomplètes conditionnées par plusieurs gènes à effet partiels dites résistance quantitative (Poland *et al.* 2009). L'utilisation des résistances qualitatives sous contrôle des gènes R présente un réel intérêt en amélioration variétale puisqu'il est plus aisé de les localiser sur le génome et de les introduire dans un fond génétique d'intérêt. Cependant, leur efficacité varie selon le pathosystème, les motifs de reconnaissance hôte-pathogène sont extrêmement spécifiques, l'hôte sera résistant à un type de pathogène donné et donc facilement contournable, notamment dans le cas d'un pathogène à fort potentiel évolutif (McDonald and Linde 2002). De plus, il s'avère que la résistance qualitative est plus efficace vis-à-vis des pathogènes biotrophes où la réaction hypersensible va bloquer sa progression, alors que dans le cas de pathogènes nécrotrophes, la réaction d'hypersensibilité sera favorable à sa progression (Glazebrook 2005).

Dans la présente étude, le pathogène est hémi-biotrophe, une résistance sous contrôle polygénique était attendue, d'autant qu'aucune résistance majeure n'a été identifiée au niveau phénotypique (Pilotti *et al.* 2003). Trois régions QTL consensus ont été mises en évidence par les modèles QTL, un QTL chez les La Mé (sur le groupe de liaison 13) et deux QTL chez les Yangambi (sur les groupes de liaisons 5 et 9) bien qu'ils soient considérés comme plus sensibles à la vue de la distribution de leurs valeurs génétiques par rapport à celle du groupe B. C'est une première approche pour l'étude du déterminisme génétique de la résistance au *Ganoderma* réalisée au sein des géniteurs du groupe hétérotique B. Les régions QTL mises en évidence devront être validées par d'autres analyses et d'autres expérimentations.

Une étude menée sur trois familles d'hybrides AxB (Déli x Yangambi) constituée au maximum de 34 individus a également été réalisée sur des données de pépinières par l'institut FELDA (*Federal Land Development Authority*) en Malaisie (Hama-Ali *et al.* 2015). Parmi les 58 marqueurs SSR développés par le CIRAD qui ont été utilisés, deux ont été identifiés comme étant associés au niveau de résistance à la maladie. Ces résultats n'ont pas pu être retrouvés dans la présente étude mais les effectifs et le design de la population, le nombre de marqueurs et les analyses statistiques effectuées ne permettent pas de conclusion robuste.

Une autre étude menée cette fois sur des données plein champ a mis en évidence des QTL de résistance au *Ganoderma* (Tisné *et al.* 2017). La population d'étude était une population multiparentale de 1200 hybrides AxB suivie pendant 25 ans dont le fond génétique des géniteurs du groupe était commun à la présente étude. Cette étude a montré quatre QTL de résistance, deux QTL sur le groupe de liaison 1, un sur le groupe de liaison 3 et un dernier sur le groupe de liaison 14. Cette étude bien que robuste et reproductible d'un point de vue modélisation statistique ne montre pas les mêmes QTL de résistance que la présente analyse. Cependant, ni le stade végétatif des palmiers, ni le trait phénotypique mesuré, ni l'environnement de développement des palmiers n'ont été les mêmes (en plein champ, une échelle symptomatique est utilisée pour noter le stade d'infection du palmier) et peuvent expliquer les résultats différents. Il sera nécessaire de développer une étude qui vise à comparer les résultats obtenus en pépinière avec ceux obtenus plein-champ. Afin de mesurer la pertinence des essais en pépinière, (Breton *et al.* 2006, 2009) ont déjà montré une bonne corrélation entre l'effet des fondateurs des groupes hétérotiques au champ et en pépinière.

Ces trois études ont montré une diversité de bases génétiques à faible effet visibles dans les mêmes fonds génétiques, sans aucun QTL majeur ressorti des analyses, ce qui renforce l'aspect polygéniques de la résistance au *Ganoderma*. Les résistances quantitatives sont considérées comme étant plus durables ou moins facilement contournable en raison de leur caractère polygénique

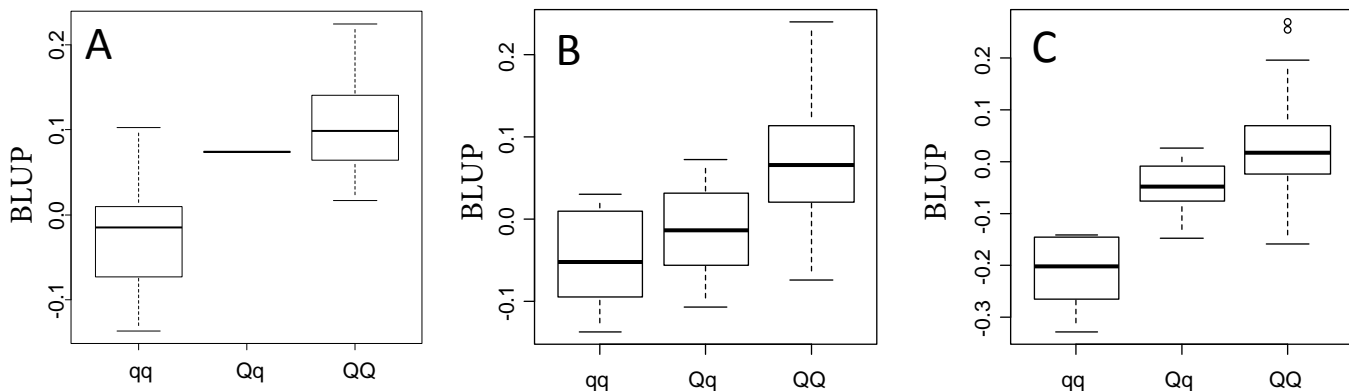


Figure 18 : Boxplot de l'effet des génotypes aux QTLs. Effet des QTL identifiés à l'issue des analyses, les QTL 5@9.46 (A) 9@88.5 (B) et 13@88.45 (C) sont représentés ici. L'allèle « q » étant l'allèle associé à la résistance à la maladie. Une diminution des valeurs génétiques représente un gain de résistance.

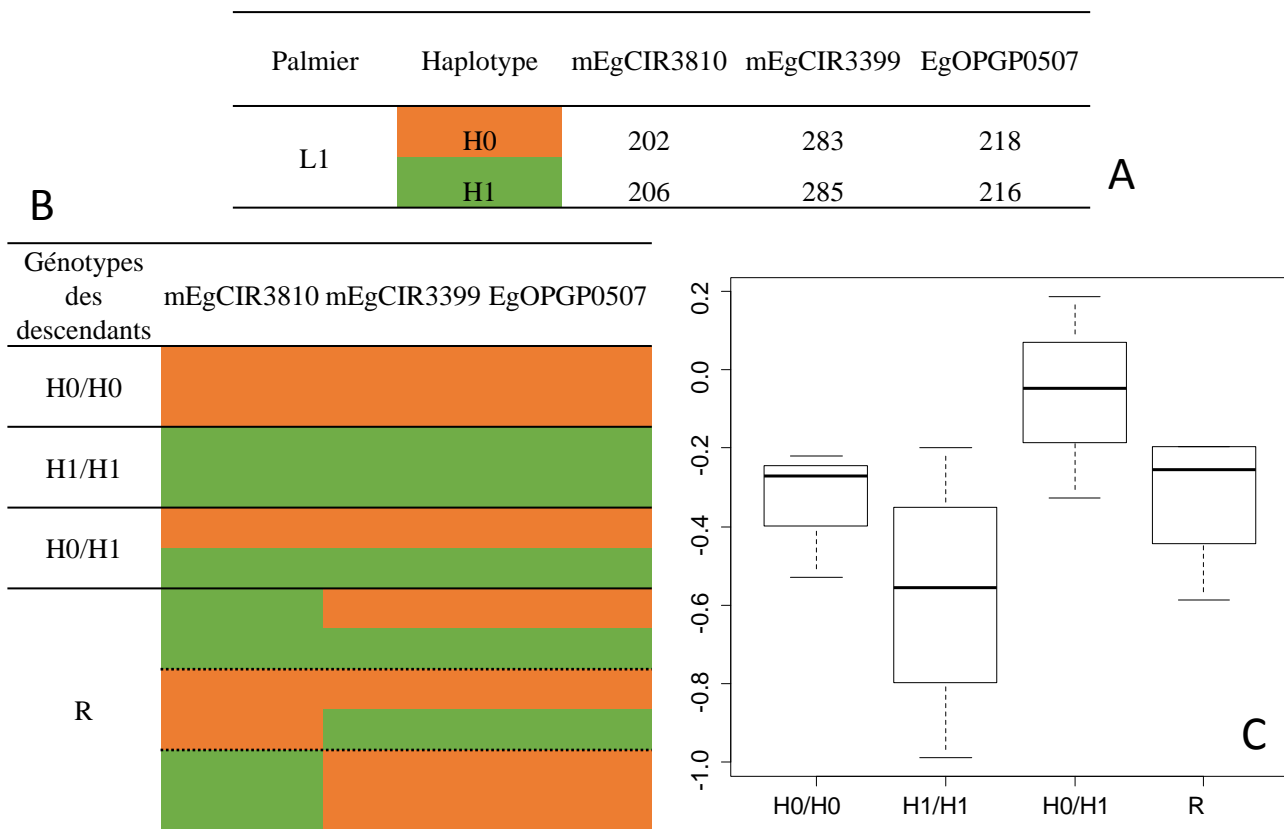


Figure 19 : Haplotypage et effet des haplotypes aux marqueurs avoisinant le QTL du groupe de liaison 13 chez les autofécondations de L1. Les haplotypes du fondateur L1 pour les marqueurs mEgCIR3810, mEgCIR3399 et EgOPGP0507 (A). Observation de la transmission des haplotypes au sein des autofécondations de L1, les haplotypes parentaux (homozygotes et hétérozygotes) et les haplotypes recombinants ont été visibles (B). Effet des haplotypes au sein des autofécondations de L1 (C).

(Palloix *et al.* 2009). Poland *et al.*, (2009) annoncent que la force de la résistance quantitative demeure dans le fait de l'implication d'une large diversité de mécanismes de résistance, mais ce qui la rend plus difficile à sélectionner.

4.4. La sélection du palmier à huile, ce qu'apporte cette étude

L'étude de populations de palmiers à huile en pedigree a permis l'identification d'allèles de résistance et de sensibilité au *Ganoderma* dans deux origines génétiques distinctes faisant parti du groupe hétérotique B. L'amélioration génétique du palmier à huile sur le critère de la tolérance à la maladie passe aujourd'hui par l'utilisation des géniteurs les moins sensibles au *Ganoderma* à l'issu des tests en pépinière et au champ. Le schéma de sélection utilisé est un schéma de sélection récurrente réciproque qui exploite l'effet d'hétérosis entre les deux groupes hétérotiques ; afin de maximiser les cycles de sélection, le but serait de cumuler au sein des géniteurs utilisés dans les deux groupes des allèles de résistance dont leurs effets seraient additifs dans la descendance. Dans cette étude, seul l'effet des individus du groupe hétérotique B a été considéré, mais indépendamment du modèle utilisé, la variance des ASC est plus grande que celles des AGC des parents A et des parents B. Il y aurait sans doute des combinaisons AxB plus favorables à la résistance au *Ganoderma*.

L'avantage de l'utilisation de FlexQTLTM est de pouvoir visualiser la ségrégation des allèles favorables au sein des fonds génétiques. Les individus porteurs des allèles favorables pourraient être recombinaisonnés entre eux pour cumuler les QTL de résistance. L'identification d'allèles favorables à la tolérance à la maladie pourrait mener au développement de méthodes de sélection assistée par marqueur pour accélérer et optimiser les programmes d'amélioration. Il reste néanmoins une étape de validation des résultats issus de pépinières, la validité des déterminismes génétiques des maladies en pépinières et au champ n'ayant pas encore été démontrée (Breton *et al.* 2006).

Le CIRAD et PalmElit s'orientent de plus en plus sur la sélection multicritère du palmier à huile, en effet la production d'huile de palme ne subit pas uniquement la pression du *Ganoderma* mais également facteurs biotiques (pathogènes, ravageurs) ou abiotiques (sécheresse, températures). Par exemple, en Afrique, la Fusariose, le *Ganoderma* et la sécheresse sont des facteurs limitant fortement les rendements. Ce qui est intéressant avec le programme d'amélioration du palmier à huile au CIRAD, c'est que les mêmes fonds génétiques sont disponibles en Afrique et en Asie du Sud, les tests d'Asie du Sud Est étant plutôt dédiés à la résistance au *Ganoderma* tandis que ceux d'Afrique à la résistance au stress hydriques et à la Fusariose. L'utilisation d'un même pedigree dans plusieurs sites peut permettre de rapprocher les résultats obtenus et d'identifier les QTL pour différents caractères dans les mêmes fonds génétiques. De plus, la sélection de QTL de résistance aux maladies a déjà montré (Poland *et al.* 2009) que cela pouvait contre-sélectionner des traits en faveur du rendement. Il serait intéressant de répondre à ces problématiques en identifiant des caractères de tolérance aux maladies et leur ségrégation au sein des populations utilisées en sélection mais aussi d'évaluer les compromis phénotypiques qu'engendrerait la sélection des caractères de tolérance.

5. Conclusion

L'analyse du déterminisme génétique de la résistance au *Ganoderma* dans une population de palmier à huile en pedigree n'a pas de méthode d'étude prédéfinie. Les logiciels utilisés que ce soit pour la visualisation du pedigree, la cartographie génétique, la prédiction des valeurs génétiques, mais aussi pour la détection de modèle QTL sont initialement dédiés à l'étude des populations animales ou humaines mais finalement s'appliquent bien dans le cadre de ce design de population.

Un compromis doit être fait entre la prise en compte des structures d'apparement dans un modèle linéaire mixte généralisé et l'exploration bayésienne de modèle QTL en utilisant les valeurs génétiques prédites comme phénotypes.

L'étude d'une population de sélection a directement permis d'identifier les QTL et leurs allèles de résistance au *Ganoderma* au sein des géniteurs du groupe hétérotique B étudiés ainsi que leur ségrégation dans le pedigree. Trois QTL consensus ont été trouvés parmi les deux groupes hétérotiques, dont deux au sein du pedigree Yangambi qui en moyenne était plus sensible que le pedigree La Mé, qui a mis en évidence un seul QTL. Ainsi trois fondateurs des pedigree La Mé et Yangambi utilisés dans les programmes d'amélioration du palmier à huile ont été identifiés comme porteur d'allèles favorables.

Après une étape de validation de l'utilisation de tels modèles statistiques, une étape de validation de la cohérence entre les résultats de pépinière et ceux au champ devra être réalisée. Cette étape sera essentielle et permettra véritablement d'optimiser l'amélioration du palmier à huile sur la résistance au *Ganoderma*.

6. Bibliographie

- Ayliffe M., R. Singh, and E. Lagudah, 2008 Durable resistance to wheat stem rust needed. *Curr. Opin. Plant Biol.* 11: 187–192. <https://doi.org/10.1016/j.pbi.2008.02.001>
- Beirnaert A. D. F., 1941 *Contribution à l'étude génétique et biométrique des variétés d'Elaeis Guineensis Jacquin*. East African Standard.
- Billotte N., N. Marseillac, A.-M. Risterucci, B. Adon, P. Brottier, *et al.*, 2005 Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 110: 754–765. <https://doi.org/10.1007/s00122-004-1901-8>
- Billotte N., M. F. Jourjon, N. Marseillac, A. Berger, A. Flori, *et al.*, 2010 QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis* Jacq.). *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 120: 1673–1687. <https://doi.org/10.1007/s00122-010-1284-y>
- Bink M., P. Uimari, J. Sillanpää, G. Janss, and C. Jansen, 2002 Multiple QTL mapping in related plant populations via a pedigree-analysis approach. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 104: 751–762. <https://doi.org/10.1007/s00122-001-0796-x>
- Bink M. C. A. M., M. P. Boer, C. J. F. ter Braak, J. Jansen, R. E. Voorrips, *et al.*, 2008 Bayesian analysis of complex traits in pedigreed plant populations. *Euphytica* 161: 85–96. <https://doi.org/10.1007/s10681-007-9516-1>
- Bink M. C. A. M., J. Jansen, M. Madduri, R. E. Voorrips, C.-E. Durel, *et al.*, 2014 Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theor. Appl. Genet.* <https://doi.org/10.1007/s00122-014-2281-3>
- Breton F., Y. Hasan, Hariadi, Z. Lubis, and H. De Franqueville, 2006 Characterization of parameters for the development of an early screening test for basal stem rot tolerance in oil palm progenies. *Technol. Breakthr. Commer. Way Forw. Proc. PIPOC 2005 Int. Palm Oil Congr. Agric. Biotechnol. Sustain.* 25-29 Sept. 2005 Petaling Jaya Malays.
- Breton F., R. Miranti, Z. Lucis, Z. Hayun, S. UMI, *et al.*, 2009 Ganoderma disease of the oil palm : hypothesis on natural infection and implementation of an early artificial inoculation test to screen oil palm progenies for their level of resistance. 30.
- Butler D., B. R. Cullis, A. R. Gilmour, and B. J. Gogel, 2007 ASReml-R reference manual. 145.
- Cao T. V., 1995 Organisation de la variabilité génétique chez le palmier à huile (*Elaeis guineensis* Jacq.) : conséquences pour l'amélioration des populations et la création variétale
- Carmeille A., C. Caranta, J. Dintinger, P. Prior, J. Luisetti, *et al.*, 2006 Identification of QTLs for *Ralstonia solanacearum* race 3-phylo-type II resistance in tomato. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 113: 110–121. <https://doi.org/10.1007/s00122-006-0277-3>
- Chong K. P., J. Dayou, and A. Alexander, 2017 Pathogenic Nature of *Ganoderma boninense* and Basal Stem Rot Disease, pp. 5–12 in *Detection and Control of Ganoderma boninense in Oil Palm Crop*, Springer International Publishing, Cham.
- Cochard B., B. Adon, S. Rekima, N. Billotte, R. D. de Chenon, *et al.*, 2009 Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding. *Tree Genet. Genomes* 5: 493–504. <https://doi.org/10.1007/s11295-009-0203-3>

- Cochard B., C. Carrasco-Lacombe, V. Pomiès, J.-F. Dufayard, E. Suryana, *et al.*, 2015 Pedigree-based linkage map in two genetic groups of oil palm. *Tree Genet. Genomes* 11. <https://doi.org/10.1007/s11295-015-0893-7>
- Cooper R. M., J. Flood, R. W. Rees, and web-support@bath.ac.uk, 2011 *Ganoderma boninense* in oil palm plantations: current thinking on epidemiology, resistance and pathology. *The Planter* 87: 515–526.
- Corley R. H. V., and P. B. Tinker, 2003 *The oil palm*. John Wiley & Sons, Hoboken, NJ.
- Cros D., 2014 Etude des facteurs contrôlant l'efficacité de la sélection génomique chez le palmier à huile (*Elaeis guineensis* Jacq)
- Dekkers J. C. M., and F. Hospital, 2002 The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* 3: 22–32. <https://doi.org/10.1038/nrg701>
- Ducrocq V., 1990 *Inra Productions Animales - Les techniques d'évaluation génétique des bovins laitiers*
- Durand-Gasselín T., L. Baudouin, B. Cochard, B. Adon, and T. V. Cao, 1999 Stratégies d'amélioration génétique du palmier à huile. *Plant. Rech. Dév.* 6: 344–358.
- Durand-Gasselín T., H. Asmady, A. Flori, J.-C. Jacquemard, Z. Hayun, *et al.*, 2005 Possible sources of genetic resistance in oil palm (*Elaeis guineensis* Jacq.) to basal stem rot caused by *Ganoderma boninense*—prospects for future breeding. *Mycopathologia* 159: 93–100.
- Durand-Gasselín T., L. Blangy, C. Picasso, H. de Franqueville, F. Breton, *et al.*, 2010 Sélection du palmier à huile pour une huile de palme durable et responsabilité sociale. *Ol. Corps Gras Lipides* 17: 385–392. <https://doi.org/10.1051/ocl.2010.0343>
- Franqueville D., H. (Centre de C. I. en R. A. pour le Développement, D. (Cote d'Ivoire) I. I. de R. pour les H. et Oleagineux, P. E. R. Michaux), and J. L. Renard, 1990 Improvement of oil palm vascular wilt tolerance - results and development of the disease at the R. Michaux plantation. *Ol. Fr.*
- Garrick D. J., J. F. Taylor, and R. L. Fernando, 2009 Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol. GSE* 41: 55. <https://doi.org/10.1186/1297-9686-41-55>
- Gascon J. P., and C. De Berchoux, 1964 Caractéristiques de la production de quelques origines d'*Elaeis guineensis* (Jacq.) et de leurs croisements : application à la sélection du palmier à huile. *Oléagineux* 19: 75–84.
- George A. W., P. M. Visscher, and C. S. Haley, 2000 Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics* 156: 2081–2092.
- Glazebrook J., 2005 Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. *Annu. Rev. Phytopathol.* 43: 205–227. <https://doi.org/10.1146/annurev.phyto.43.040204.135923>
- Green P., K. Falls, and S. Crooks, 1990 CRIMAP Documentation
- Hama-Ali E. O., J. M. Panandam, S. G. Tan, S. S. R. S. Alwee, T. J. Sheong, *et al.*, 2015 Association between basal stem rot disease and simple sequence repeat markers in oil palm, *Elaeis guineensis* Jacq. *Euphytica* 202: 199–206. <https://doi.org/10.1007/s10681-014-1270-6>

- Howard N. P., E. van de Weg, D. S. Bedford, C. P. Peace, S. Vanderzande, *et al.*, 2017 Elucidation of the “Honeycrisp” pedigree through haplotype analysis with a multi-family integrated SNP linkage map and a large apple (*Malus domestica*) pedigree-connected SNP data set. *Hortic. Res.* 4: 17003. <https://doi.org/10.1038/hortres.2017.3>
- Howard N. P., E. van de Weg, J. Tillman, C. B. S. Tong, K. A. T. Silverstein, *et al.*, 2018 Two QTL characterized for soft scald and soggy breakdown in apple (*Malus domestica*) through pedigree-based analysis of a large population of interconnected families. *Tree Genet. Genomes* 14: 2. <https://doi.org/10.1007/s11295-017-1216-y>
- Isik F., J. Holland, and C. Maltecca, 2017 *Genetic Data Analysis for Plant and Animal Breeding*. Springer International Publishing.
- Jacquemard J.-C., 2012 *Le Palmier A Huile*. Quae CTA, Versailles; Wageningen; Gembloux; Presses agronomiques de Gembloux.
- Kruuk L. E. B., 2004 Estimating genetic parameters in natural populations using the “animal model”. *Philos. Trans. R. Soc. B Biol. Sci.* 359: 873–890. <https://doi.org/10.1098/rstb.2003.1437>
- Laperche A., Y. Aigu, M. Jubault, M. Ollier, S. GUICHARD, *et al.*, 2017 Clubroot resistance QTL are modulated by nitrogen input in *Brassica napus*. *Theor. Appl. Genet.* 130: 669–684. <https://doi.org/10.1007/s00122-016-2842-8>
- Lim T. K., G. F. Chung, and W. H. Ko, 1992 Basal stem rot of oil palm caused by *Ganoderma boninense*. *Plant Pathol. Bull.* 1: 147–152.
- Lynch M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer.
- Madden L. V., G. Hughes, and F. van den Bosch, 2017 *The Study of Plant Disease Epidemics*. The American Phytopathological Society.
- Mangandi J., S. Verma, L. Osorio, N. A. Peres, E. van de Weg, *et al.*, 2017 Pedigree-based analysis in a multiparental population of octoploid strawberry reveals QTL alleles conferring resistance to *Phytophthora cactorum*. *G3 Genes Genomes Genet.* 7: 1707–1719.
- Marchal A., A. Legarra, S. Tisné, C. Carasco-Lacombe, A. Manez, *et al.*, 2016 Multivariate genomic model improves analysis of oil palm (*Elaeis guineensis* Jacq.) progeny tests. *Mol. Breed.* 36. <https://doi.org/10.1007/s11032-015-0423-1>
- McDonald B. A., and C. Linde, 2002 Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* 40: 349–379. <https://doi.org/10.1146/annurev.phyto.40.120501.101443>
- Meunier J., and J. P. Gascon, 1972 Le schéma général d’amélioration du palmier à huile à l’IRHO. *Oléagineux* 27: 1–12.
- Ooijen J. W., and R. E. Voorrips, 2002 *JoinMap version 3.0: software for the calculation of genetic linkage maps*. University and Research Center, Wageningen.
- Palloix A., V. Ayme, and B. Moury, 2009 Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. *New Phytol.* 183: 190–199. <https://doi.org/10.1111/j.1469-8137.2009.02827.x>

- Pilotti C. A., F. R. Sanderson, and E. A. B. Aitken, 2003 Genetic structure of a population of *Ganoderma boninense* on oil palm. *Plant Pathol.* 52: 455–463.
- Van der Plank J. E., 1968 *Disease resistance in plants*. Academic Press.
- Poland J. A., P. J. Balint-Kurti, R. J. Wisser, R. C. Pratt, and R. J. Nelson, 2009 Shades of gray: the world of quantitative disease resistance. *Trends Plant Sci.* 14: 21–29. <https://doi.org/10.1016/j.tplants.2008.10.006>
- Prevot P., 1962 *Physiologie des plantes tropicales cultivées*
- Singh R., E.-T. L. Low, L. C.-L. Ooi, M. Ong-Abdullah, N.-C. Ting, *et al.*, 2013 The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* 500: 340–344. <https://doi.org/10.1038/nature12356>
- Some A., M. J. Manzanares, F. Laurens, F. Baron, G. Thomas, *et al.*, 1996 Variation for virulence on *Brassica napus* L. amongst *Plasmodiophora brassicae* collections from France and derived single-spore isolates. *Plant Pathol.* 45: 432–439. <https://doi.org/10.1046/j.1365-3059.1996.d01-155.x>
- Tisné S., M. Denis, D. Cros, V. Pomiès, V. Riou, *et al.*, 2015 Mixed model approach for IBD-based QTL mapping in a complex oil palm pedigree. *BMC Genomics* 16. <https://doi.org/10.1186/s12864-015-1985-3>
- Tisné S., V. Pomiès, V. Riou, I. Syahputra, B. Cochard, *et al.*, 2017 Identification of *Ganoderma* Disease Resistance Loci Using Natural Field Infection of an Oil Palm Multiparental Population. *G3 Genes Genomes Genet.* 7: 1683–1692. <https://doi.org/10.1534/g3.117.041764>
- Tranbarger T. J., W. Kluabmongkol, D. Sangsrakru, F. Morcillo, W. James Tregear, *et al.*, 2012 SSR markers in transcripts of genes linked to post-transcriptional and transcriptional regulatory functions during vegetative and reproductive development of *Elaeis guineensis*. *BMC Plant Biol.* 12: 1. <https://doi.org/10.1186/1471-2229-12-1>
- Villemereuil P. de, 2018 Quantitative genetic methods depending on the nature of the phenotypic trait: Quantitative genetics for non-classical traits. *Ann. N. Y. Acad. Sci.* 1422: 29–47. <https://doi.org/10.1111/nyas.13571>
- Voorrips R. E., 2002 MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *J. Hered.* 93: 77–78. <https://doi.org/10.1093/jhered/93.1.77>
- Voorrips R. E., M. C. A. M. Bink, and W. E. van de Weg, 2012 Pedimap: software for the visualization of genetic and phenotypic data in pedigrees. *J. Hered.* 103: 903–907. <https://doi.org/10.1093/jhered/ess060>
- Würschum T., 2012 Mapping QTL for agronomic traits in breeding populations. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 125: 201–210. <https://doi.org/10.1007/s00122-012-1887-6>
- Zaki N. M., R. Singh, R. Rosli, and I. Ismail, 2012 *Elaeis oleifera* Genomic-SSR Markers: Exploitation in Oil Palm Germplasm Diversity and Cross-Amplification in Arecaceae. *Int. J. Mol. Sci.* 13: 4069–4088. <https://doi.org/10.3390/ijms13044069>
- Zulkifli Y., A. Norziha, M. H. Naquiuddin, A. M. Fadila, N. Azwani, *et al.*, 2017 Designing the oil palm of the future. *J. Oil Palm Res.* 29: 440–455.

Sitographie

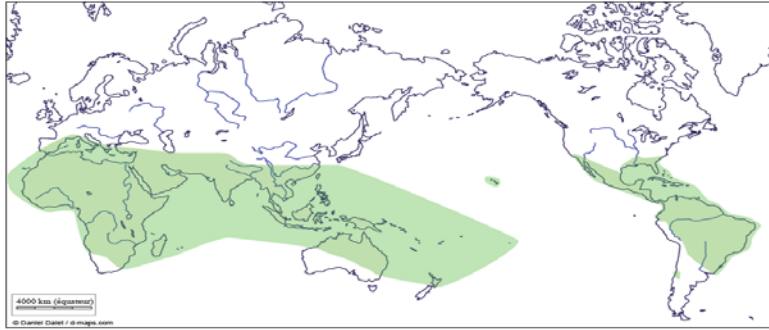
United States Department of Agriculture, USDA 2015, www.fas.usda.gov

Oil World 2016, www.oilworld.biz

FAOSTAT 2013, www.fao.org/faostat/en/#home

RSPO, Roundtable on sustainable Palm Oil « *The Global initiative for Sustainable Palm Oil Production* », www.rspo.org

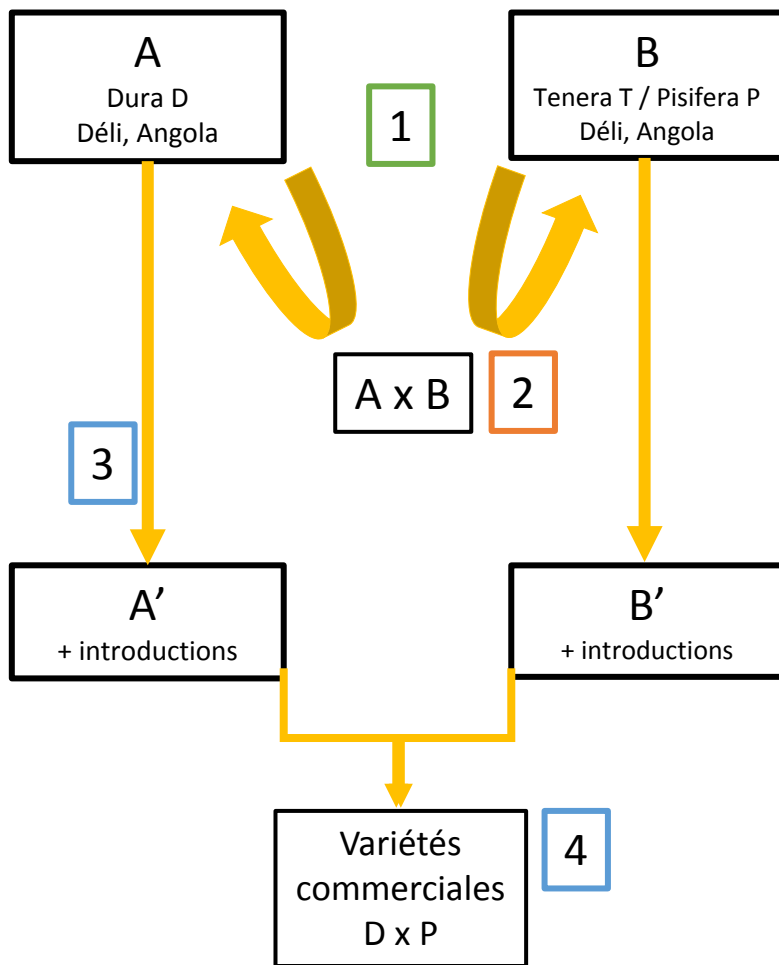
Annexe I : Distribution actuelle des palmiers, d'après (Dransfield et al., 2008)



Annexe II : Régime d'un palmier à huile chargé en fruits. La partie orange du fruit contient l'huile de palme, la partie blanche à l'intérieur de la noix contient l'huile de palmiste. Photographie du Cirad.



Annexe III : Schéma de sélection récurrente réciproque appliquée au palmier à huile depuis 1957.
D'après Cros,,(2015)



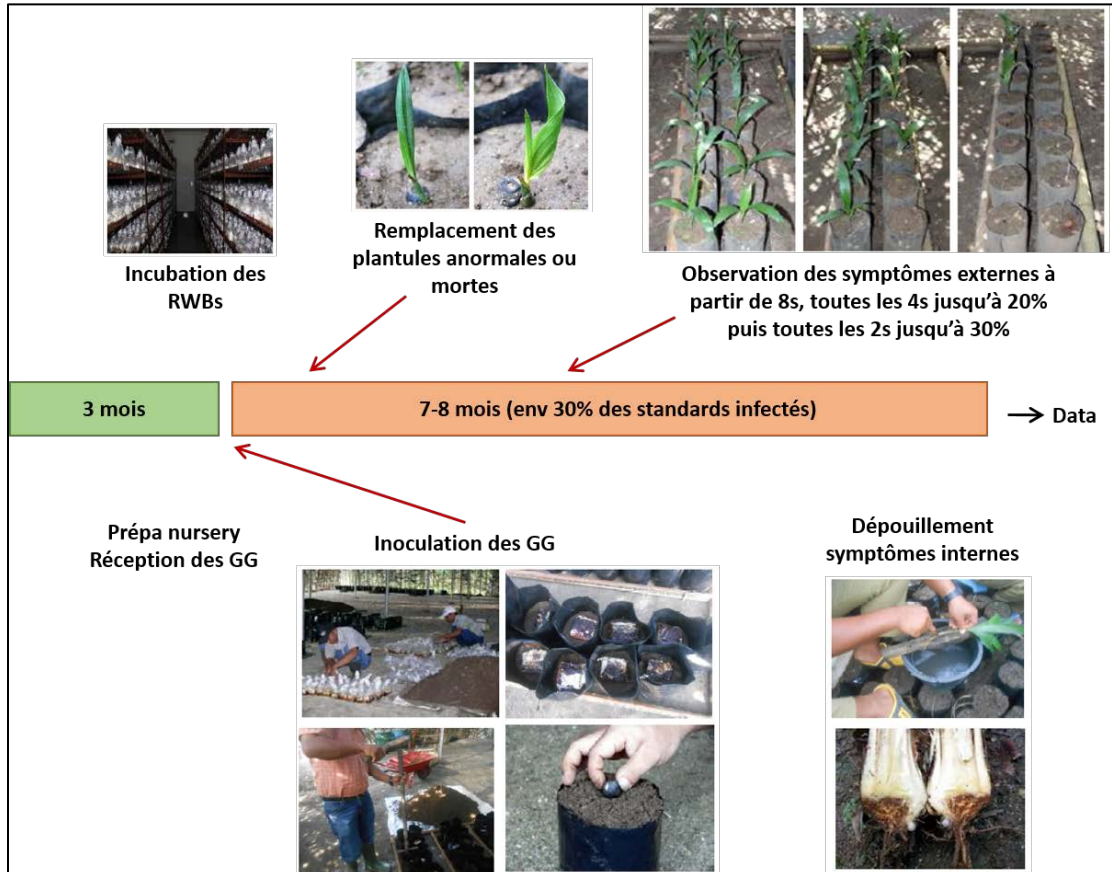
Un cycle de sélection dure au moins 12 ans et comporte 3 étapes :

1 : Les populations A et B de départ sont soumises à une sélection massale (sur caractères les moins influencés par l'environnement)

2 : Les individus A et B sélectionnés sont croisés entre eux et évalués dans des essais génétiques. La sélection finale est faite sur les aptitudes à la combinaison.

3 : Les individus sélectionnés sont autofécondés et recombinaés au sein de chaque groupe pour former 2 populations A et B améliorées qui serviront de point de départ au cycle suivant et à produire des semences commerciales (**4**)

Annexe IV : Principales étapes de mises en place des tests précoces en pépinières et suivi de l'apparition des symptômes sur les jeunes palmiers. Figure réalisée d'après PalmElit 2018. La première étape consiste à inoculer et incuber des blocs d'hévéas (RWB, *rubber wood block*) par des souches de *Ganoderma*. La seconde étape est marquée par l'inoculation des graines germées (GG) de palmiers à huile avec les RWB dans des polybags. La dernière étape consiste au suivi des symptômes externes et au dépouillement des symptômes internes en fin de test.



Extraction MATAB d'ADN de Palmier à huile

Broyer 500 mg au mortier dans N2 liquide

Peser entre 150mg et 250mg max de broyat de feuilles lyophilisées

Transférer dans un tube de 15 ml

1)* ajouter 5 ml de tampon d'extraction, préchauffé à 74 °c

Ajouter 15 microlitres de RNase T1: 25mg/ml à 20u/mg (facultatif)

2)* Vortexer 10 sec

3)* Incuber 20 mn à 74°C, agiter légèrement de temps en temps (mi-temps)

4)* Refroidir à température ambiante

ajouter 5 ml de CIAA (Chloro IsoAmylAlcohol) (20ml alcool isoamylique + 480ml chloroforme)

agiter par retournement (~100 fois)

5)* Centrifuger 15 mn à 9000 g ou 30 min à 4500g

6)* Transférer la phase aqueuse dans nouveaux tubes de 15 ml

7)* Ajouter 5 ml d'isopropanol

8)* Agiter légèrement jusqu'à formation de la pelote d'ADN, ou centrifuger si nécessaire.

30 min à 4000g – vider le tube ajouter 1ml d'éthanol 70%

20 min à 4500g – vider et faire sécher le culot

(Note : on peu conserver les tubes à -20°C si pas le temps de poursuivre)

9)* Reprendre la pelote dans 400 microlitres de TE pour microsattellites

Agitateur O/N

Doser plus tard, 1 mois dans le frigo et dilution 200X pour usat

Tampon d'extraction MATAB

Pour 500 ml:

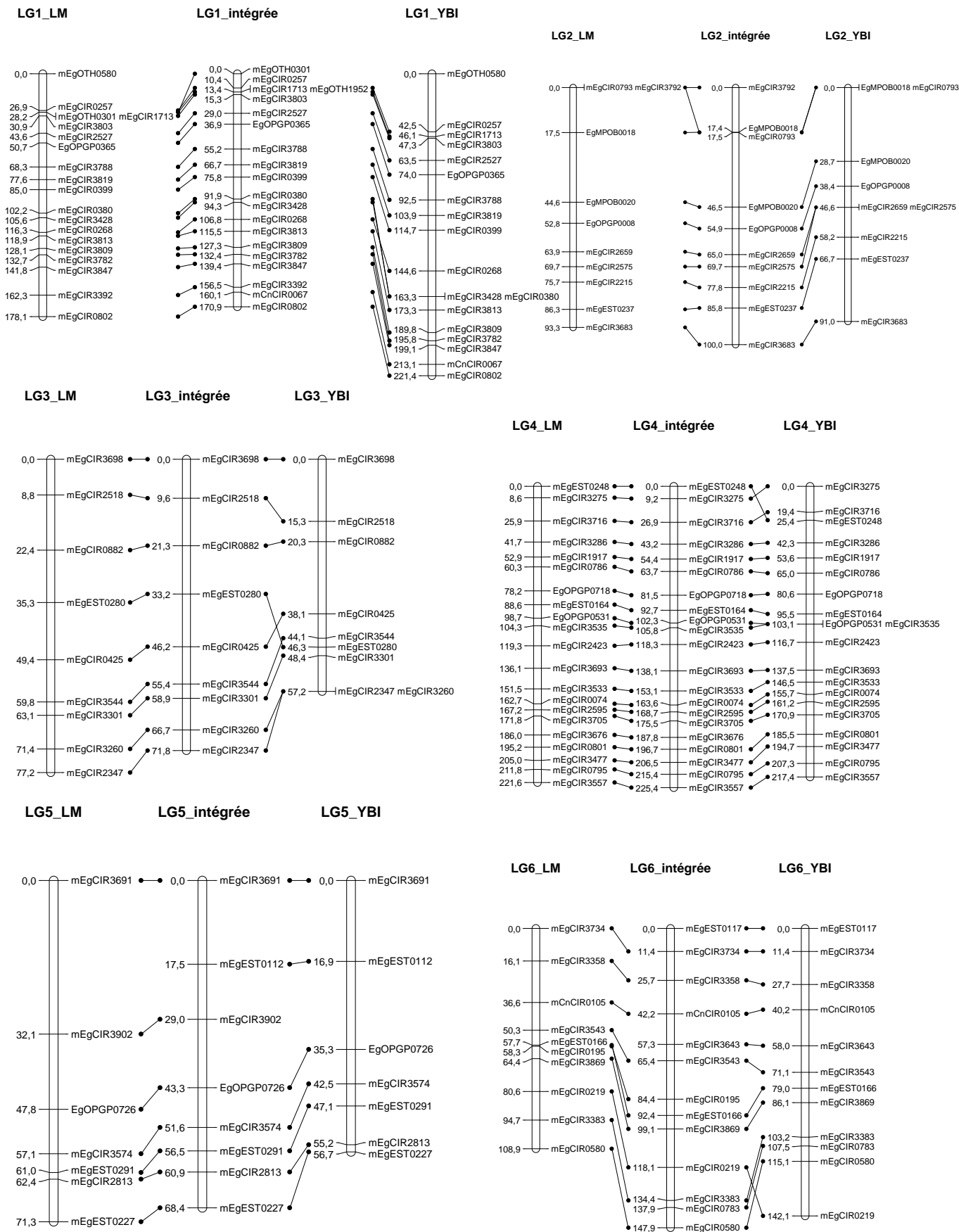
Concentration finale

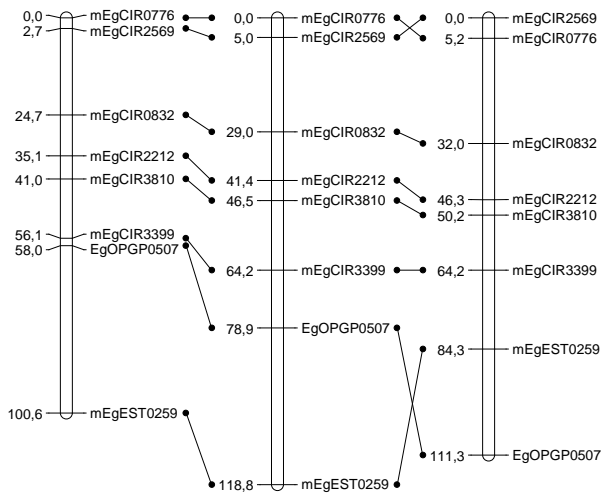
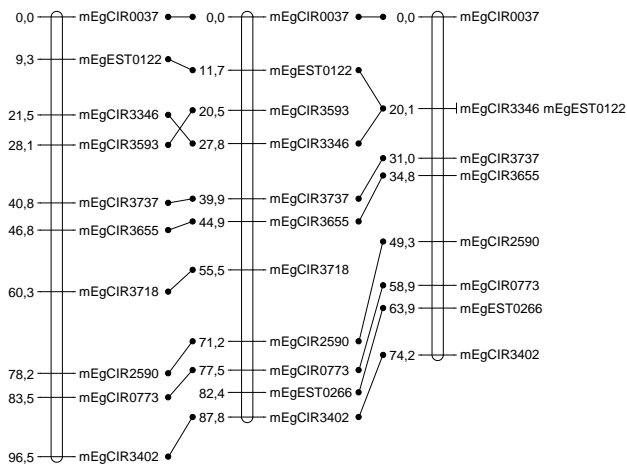
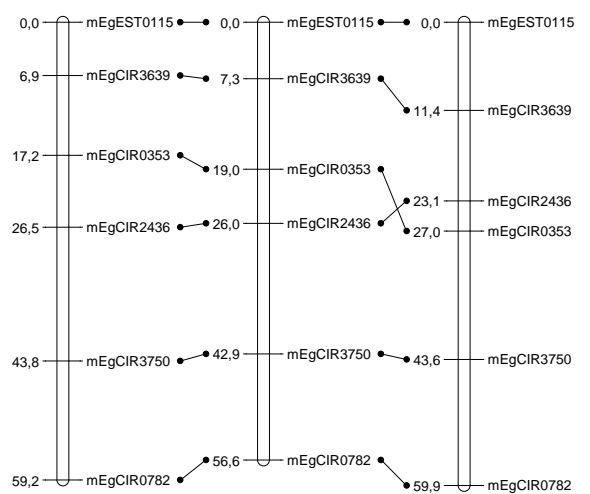
-100 mM Tris-cl, ph 8.0	50 ml, 1M
-1,4 M NaCl	140 ml, 5M
-20 mM EDTA	20 ml, 500mM
-2% MATAB	10 gr
-1% PEG 6000	5 gr
-0,5% sulfite de sodium	2,5 gr

Annexe VI : Tableau présentant les paramètres pour la cartographie de QTL avec le logiciel FlexQTL™ (inspiré de **Mangandi et al., (2017)**). Liste non exhaustive, seuls des paramètres principaux sont renseignés ici. D'autres paramètres en lien avec les analyses statistiques bayésiennes ont été utilisés mais ne sont pas nécessaire à la compréhension générale de l'étude.

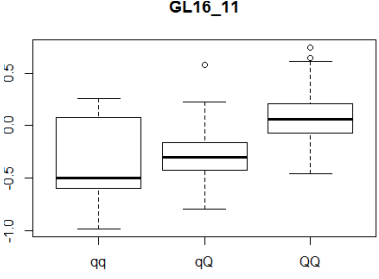
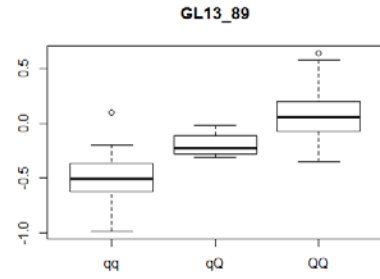
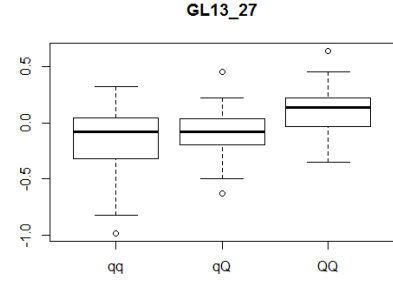
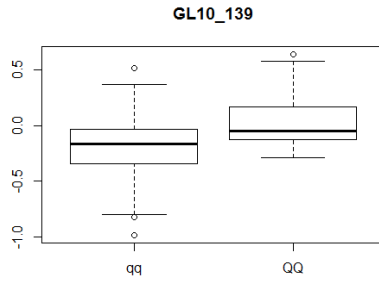
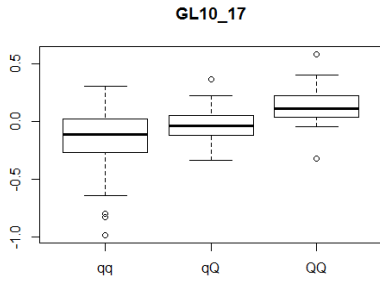
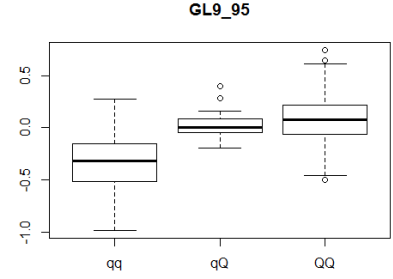
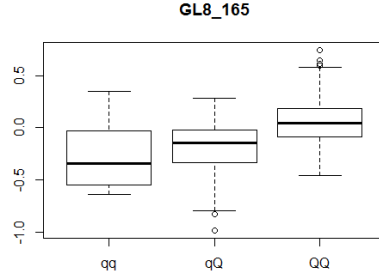
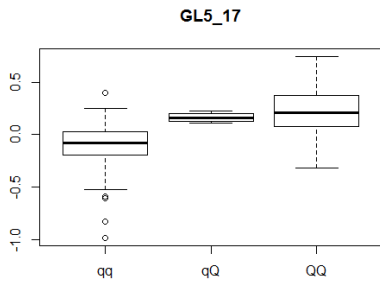
Paramètres FlexQTL™	Descriptions	Valeurs	Commentaires
indiN	Indicateur de variable de nuisance. Permet d'ajouter un facteur à l'analyse. Ex : appartenance à une sous population etc.	0	La valeur phénotypique a dû être modélisée en amont. Seule l'effet du génotype est analysé ici.
nchrom	Nombre de chromosomes à analyser dans le jeu de donnée	16	16 groupes de liaisons ont été identifiés suite à la réalisation de la carte génétique.
indiC	Indication sur les chromosomes à inclure dans l'analyse	1	0 = exclu 1 = inclu
mmrkrC	Nombre de marqueurs par groupe de liaison présents dans le jeu de donnée et sur la carte génétique	n (max = 21, min = 6)	Varie selon les groupes de liaisons
indicQTL	Indication pour une analyse sur les QTL ou sur les IBD	1	1 = analyse QTL 2 = analyse IBD
maximQTL	Nombre maximal de QTL qui doit être inclus dans le modèle	20	Jusqu'à 20 QTL sont permis à chaque analyse
priorQTL	Distribution <i>a priori</i> du nombre de QTL, la variable est d'une loi de Poisson.	5	Jusqu'à 5 QTL retrouvés entre deux analyses.
addGENE	Indication pour les effets additifs de chaque QTL	5	Six options sont disponibles sur FlexQTL™ allant de 0 à 5 et simulant les effets d'additivité ou de dominance des QTL. L'option 5 correspond à la présente étude : Effet additif des QTL avec une distribution <i>a priori</i> normale. (Bink et al., 2014).
domGENE	Indication pour les effets de dominance de chaque QTL	0	Dominance exclue du modèle (idem commentaire addGENE)
Length	Longueur de la chaîne de Markov	1 000 000	1 000 000 simulations de modèles QTL
Thinning	Raccourcissement de la chaîne de Markov, écriture dans les sorties FlexQTL™	1000	Longueur de chaîne de 1 000 000, écriture dans les fichiers de 1 000 000 / 1 000 = 1 000. 1 000 échantillons dans les fichiers de sorties.

Annexe VII : Carte génétique des 16 groupes de liaison du palmier à huile réalisée par le logiciel CRI-MAP et visualisée par MapChart. 3 cartes génétiques par groupe de liaison sont présentés ici, celles concernant le sous groupe La Mé, celles concernant le sous-groupe Yangambi et la carte intégrée des deux sous-populations.

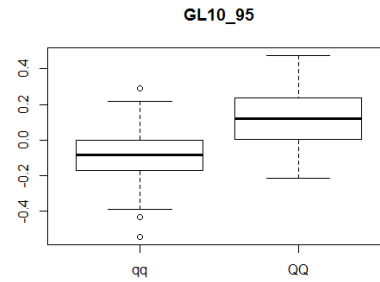


LG13_LM**LG13_intégrée****LG13_YBI****LG14_LM****LG14_intégrée****LG14_YBI****LG15_LM****LG15_intégrée****LG15_YBI****LG16_LM****LG16_intégrée****LG16_YBI**

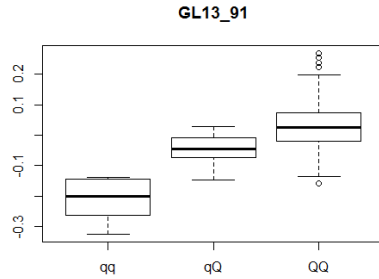
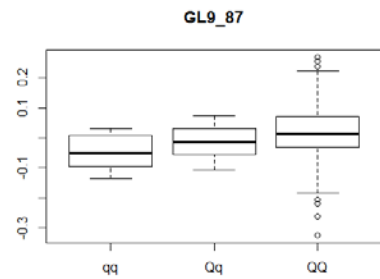
Annexe VIII : Effets des génotypes aux QTL selon leur modélisation. L'axe des ordonnées représente la valeur génétique prédite des palmiers étudiés selon le modèle utilisé.



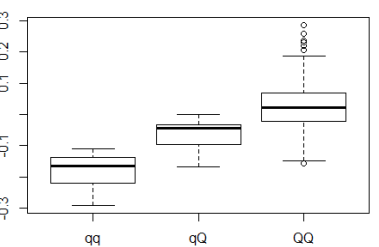
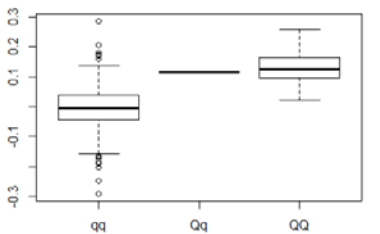
mod_A



mod_B

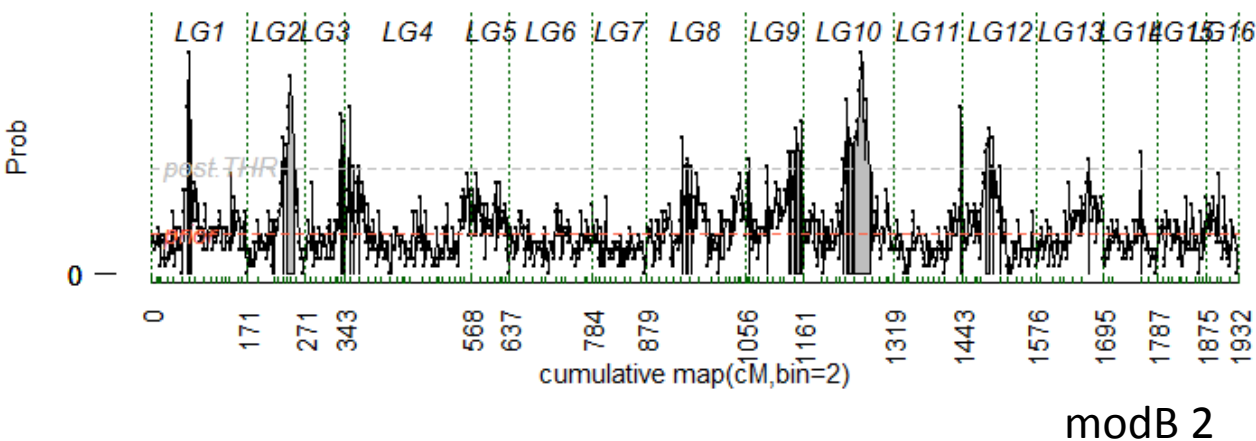
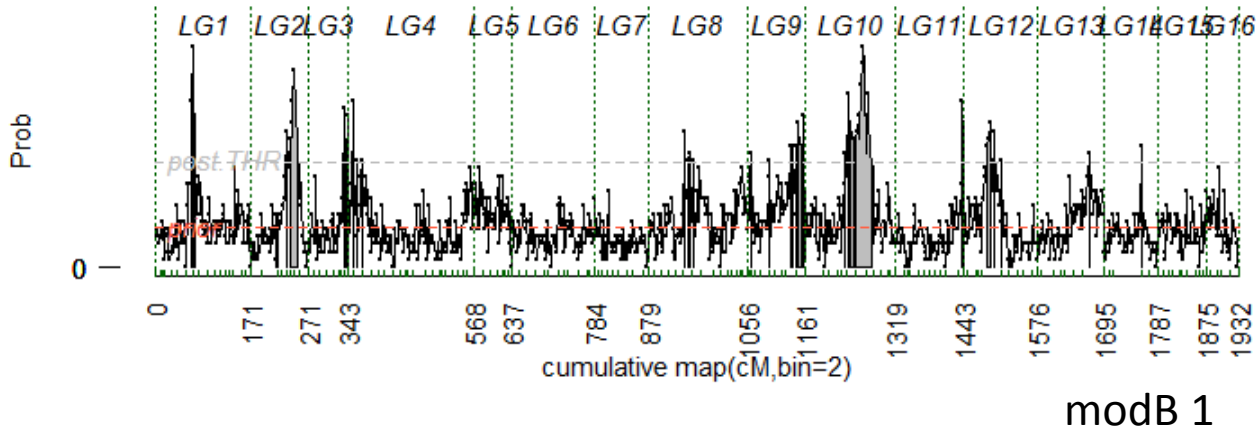
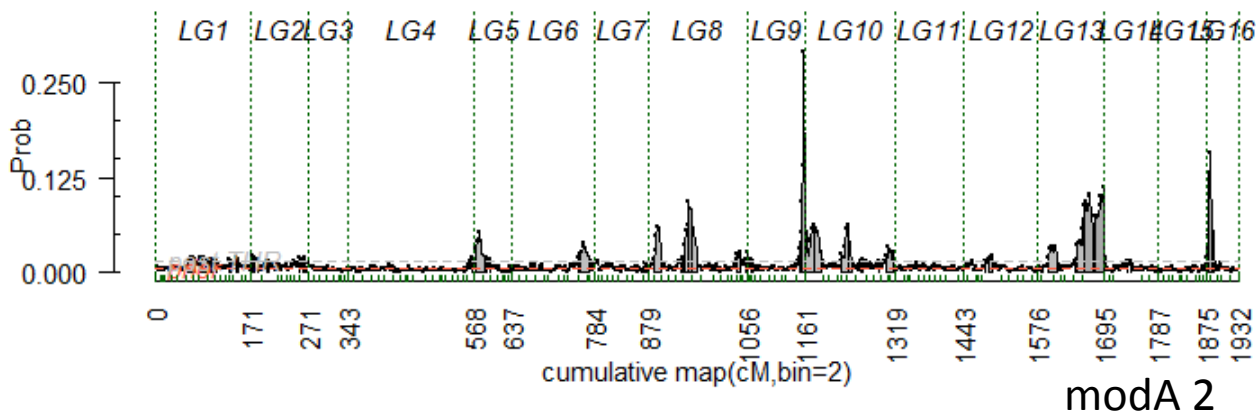
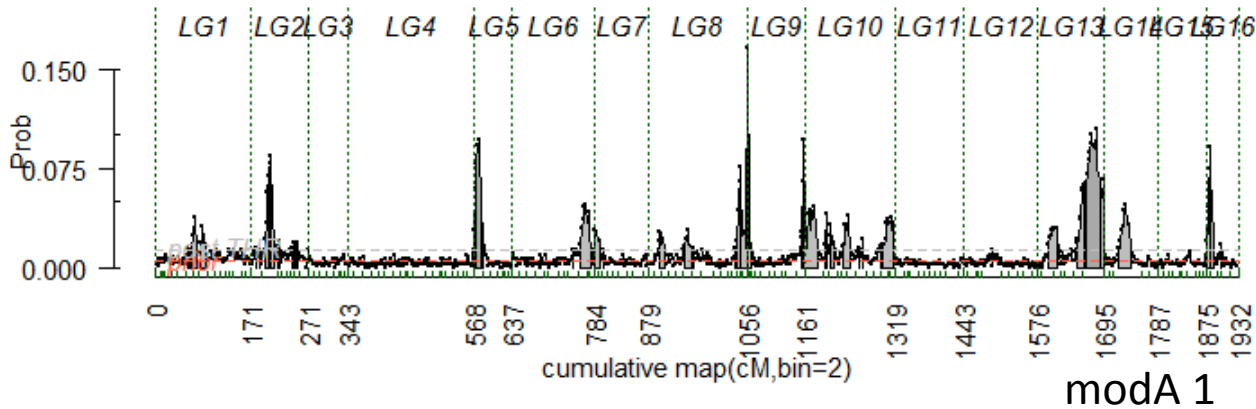


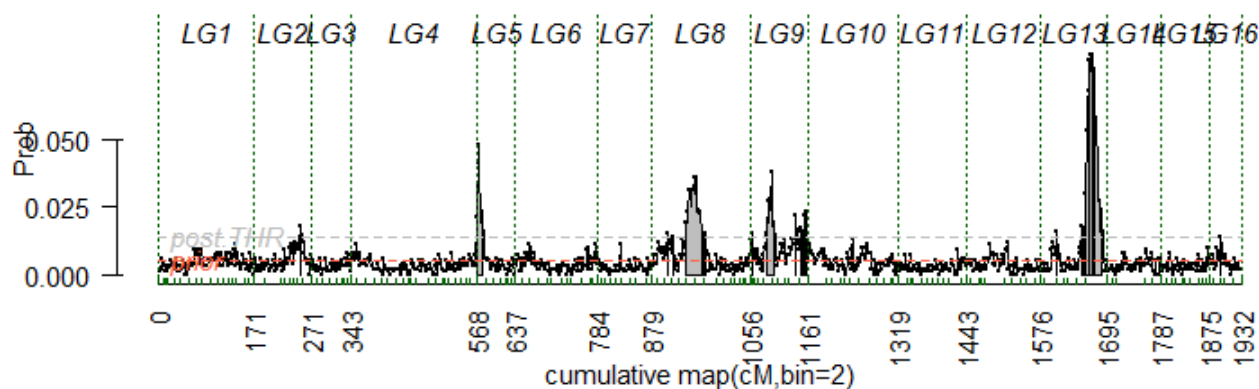
mod_C



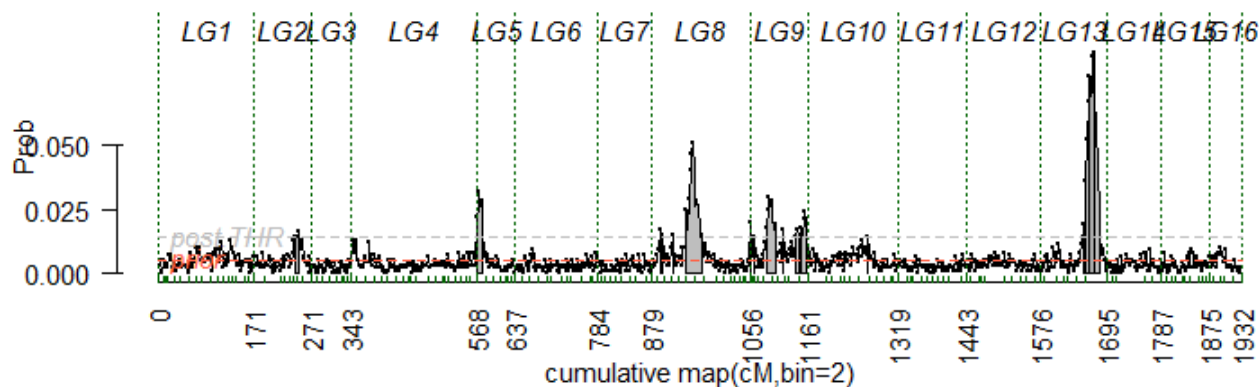
mod_D

Annexe IX : Sortie FlexQTL™, probabilité à *posteriori* des positions des QTL selon leur graine de départ (1 et 2) et selon leur modèle de prédiction (modA, modB, modC, modD).

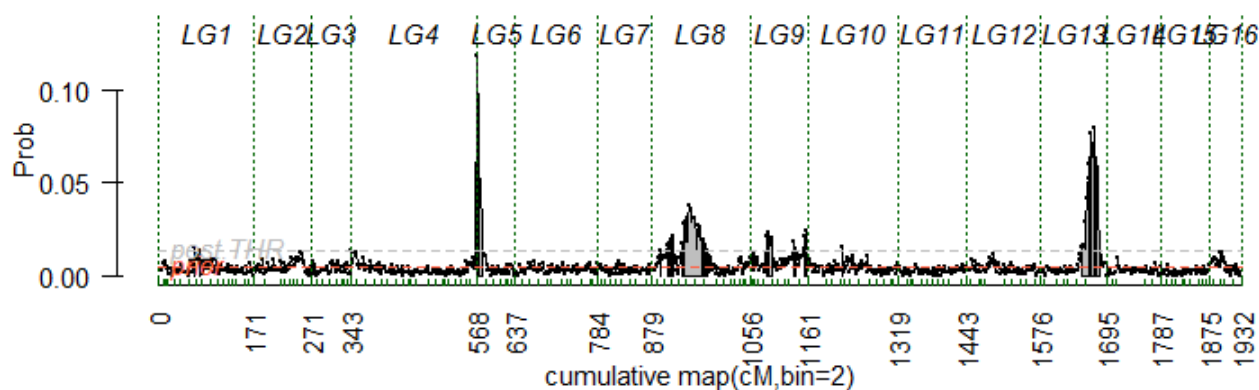




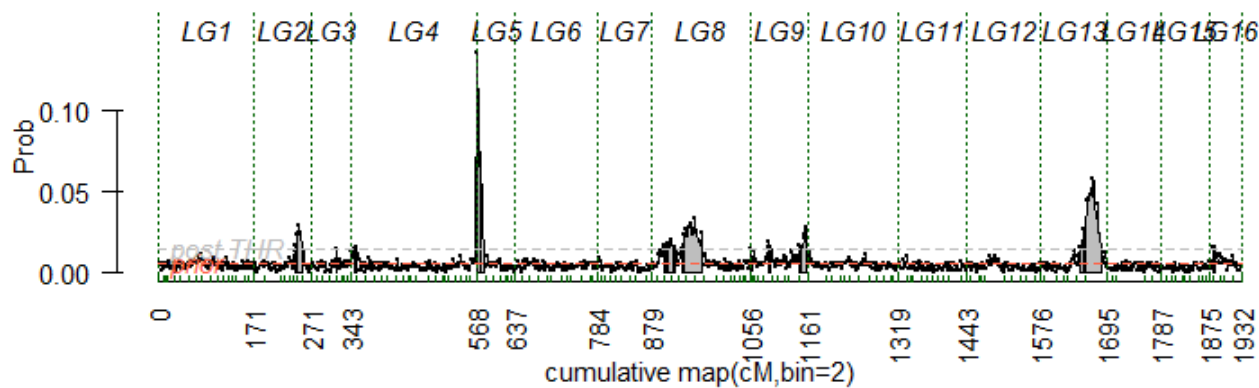
modC 1





modC 2



modD 1



modD 2

 	Diplôme et Mention : Master Biologie, Agrosociétés Parcours : Amélioration, Production et Valorisation du Végétal Option : Génétique, Génomique et Amélioration des Plantes Responsable d'option : Mélanie Jubault
Auteur(s) : Aurélie Daval	Organisme d'accueil : CIRAD
Date de naissance* : 24 mai 1993	Adresse : Avenue Agropolis
Nb pages : 25 Annexe(s) : 9	34398 Montpellier Cedex 5
Année de soutenance : 2018	France Maître de stage : Sébastien Tisné
Titre français : Analyse des bases génétiques de la tolérance à la pourriture basale du stipe chez le palmier à huile (<i>Elaeis guineensis</i> Jacq.)	
Titre anglais : Genetic analyses of basal stem rot tolerance in oil palm (<i>Elaeis guineensis</i> Jacq.)	
Résumé (1600 caractères maximum) : Le basidiomycète <i>Ganoderma boninense</i> est responsable de la pourriture basale du stipe, maladie majeure du palmier à huile (<i>Elaeis guineensis</i> Jacq.) pouvant causer jusqu'à 50 % de perte au champ. L'amélioration génétique du palmier à huile joue un rôle majeur dans la durabilité de la production d'huile de palme, premier apport mondial d'oléagineux. 335 palmiers d'un même groupe hétérotique ont été génotypés avec 199 marqueurs SSR. Ces palmiers font partie de 48 familles interconnectées issues de 7 fondateurs de deux origines génétiques distinctes. Cette étude exploite 10 ans de données de phénotypage issues de tests précoces standardisés d'inoculation de la maladie en pépinière. Ces données comptent le nombre de jeunes palmiers atteints par le <i>Ganoderma</i> durant la cinétique de la maladie. Un modèle linéaire mixte généralisé et l'aire sous la courbe de progression de la maladie ont permis de prédire les valeurs génétiques des palmiers d'intérêt à partir des données de pépinières. La mise en évidence de QTL de résistance aux <i>Ganoderma</i> dans cette population en pedigree a été faite selon une méthode Bayésienne multi-QTL. L'analyse a montré trois QTL de résistance au <i>Ganoderma</i> . La prédiction des allèles au QTL a permis d'identifier 3 fondateurs des 2 origines génétiques possédant au moins un allèle favorable. L'étude d'une population de sélection pour l'analyse des bases génétiques de la résistance au <i>Ganoderma</i> a permis d'identifier les allèles en ségrégations dans la population et les géniteurs porteurs des allèles favorables, et d'avancer vers une sélection assistée par marqueurs.	
Abstract (1600 caractères maximum) : The basidiomycete <i>Ganoderma boninense</i> causes basal stem rot disease on oil palm (<i>Elaeis guineensis</i> Jacq.) plantings. An average of 50 percent yield losses from basal stem rot occurs in mature stands. Oil palm genetic improvement is involved in durability and sustainability of the oil palm production, which is the first source of vegetal oil. 335 oil palms coming from the same heterotic group was genotyped with 199 SSR markers. These palms are organized in 48 interconnected families coming from 7 founders from 2 distinct genetic origins. This study analyzed 10 years of phenotyping data of nursery stage plants, concerning artificial and standardized inoculation of seedling with <i>Ganoderma</i> . This data counts the number of young oil palms infected by the <i>Ganoderma</i> during the disease kinetics. A generalized linear mixed model and the area under the disease progress curve were used to predict the breeding values of the oil palms of interest coming from the nursery data. <i>Ganoderma</i> resistance QTL in this pedigree population were mapped with a multi-QTL Bayesian approach. The analysis showed 3 <i>Ganoderma</i> resistance QTL. The QTL allele's prediction identified 3 founders of the 2 genetic origins with at least one favorable allele. The study of a breeding population for the determination of <i>Ganoderma</i> genetic resistance, showed alleles segregations in a pedigreed population with founders bearing favorable alleles and paves the way to a marker assisted selection.	
Mots-clés : Pedigree, Modèle linéaire mixte généralisé, Résistance aux maladies, approche Bayésienne, QTL	
Key Words: Pedigree, Generalized linear mixed model, diseases resistance, Bayesian approach, QTL	

* Élément qui permet d'enregistrer les notices auteurs dans le catalogue des bibliothèques universitaires