



**Master Biologie, Géosciences, Agronomie, Environnement
Spécialité : Biologie & Evolution des Plantes**

Parcours : M2 Biologie Fonctionnelle des Plantes

Année Universitaire 2008-2009

**Caractérisation du Déséquilibre de Liaison
chez le palmier à huile
(*Elaeis guineensis* Jacq.)**

Clélia SOLER

CIRAD UPR 28

Soutenu le

18 Septembre 2009

CIRAD-DIST
Unité bibliothèque
Lavalette



000100589

RESUME

Pour identifier des locus associés à des caractères agronomiques par génétique d'association, il est utile, au préalable, d'étudier l'étendue du Déséquilibre de Liaison (DL) entre deux locus afin de s'assurer que deux marqueurs, ou un marqueur et le caractère phénotypique soient bien associées.

Dans cette étude, menée sur l'espèce *Elaeis guineensis* Jacq. 23 marqueurs SSR microsatellites ont été choisis sur les groupes de liaison (GL) GL8 et GL15 qui représente des segments respectivement de 135.2 et 80.9 cM de long. La caractérisation du DL a été entreprise sur 3 origines améliorées de palmier à huile : La Mé (Côte d'Ivoire), Yangambi (République Démocratique du Congo) et Deli (Malaisie).

La légitimité de la population et sa structure ont été analysés avant de caractériser le DL. La structure du matériel étudié est divisé en 3 origines. Ce résultat est identique à celui mise en évidence lors de l'étude de la diversité génétique entrepris par Cochard (2008).

L'étendue du DL varie suivant les GL et les origines (entre 9 et 42 cM).

Les données récoltés lors de cette étude permettent de définir le nombre de marqueurs nécessaires à la recherche de QTL dans le cadre d'une approche genomescan. Ce qui permettra par la suite d'identifier des régions d'intérêt par une étude de génétique d'association.

ABSTRACT

To identify loci linked to agronomic trait through genetic association, Linkage disequilibrium study between 2 loci is useful in order to confirm that 2 markers, or a marker and the phenotype are really linked.

In this study, led with the specie *Elaeis guineensis* Jacq. , 23 SSR markers were selected on linkage groups (LG) LG8 and LG15 representing respectively segments of 135,2 and 80,9 cM length. Characterization of the LD was led for 3 oil palm ameliorated origins: La Mé (Côte d'Ivoire), Yangambi (République Démocratique du Congo) and Deli (Malaisie).

Population legitimacy and structure were analyzed before LD characterization. The materiel structure is divided in 3 origins. This result is identical to those found during the genetic diversity led by Cochard (2008).

LD pattern varies with LG and origins (between 9 and 42 cM).

Data resulting from this study allows us to define the amount of markers required to the QTL research, for a genomescan approach. Which will allows us to identify interesting regions through a genetic association study.

21 SEP. 2009

Remerciements

Je remercie tout d'abord l'UMR DAP (Développement et Amélioration des Plantes) et l'UPR 28 (Génétique Palmier) du CIRAD pour l'avoir accueillie dans le cadre de ce stage.

Je remercie également mon encadrant de stage, Benoît COCHARD pour son accueil chaleureux, les connaissances qu'il m'a apporté, et le temps passé à l'élaboration et à la correction de mon mémoire.

Je tiens à remercier Virginies Pomies, Virginie Riou pour l'aide précieuse qu'elles m'ont apportées pendant toute la durée de mon stage.

Je souhaite adresser mes vifs remerciements à Philippe Cubry et Jean Louis Noyer pour leur soutien, leur disponibilité et leurs conseils.

Merci à toutes les personnes que j'ai eu la chance de connaître dans ce laboratoire pour l'ambiance sympathique qui a permis de rendre mon stage encore plus agréable.

Enfin, je tiens à remercier par avance les membres de mon jury pour toute l'attention qu'ils apporteront à mon rapport de stage.

ABREVIATIONS

ADN	-----	Acide DésoxyriboNucléique
ADNc	-----	Acide DésoxyriboNucléique complémentaire
dNTPs	-----	désoxyNucléotide TriphosPhaté
ddNTP	-----	di-désoxyNucléotide TriphosPhaté
PCR	-----	Poly Chain Reaction
PPi	-----	PyroPhosphate
QTL	-----	Quantitative Trait Locus
DL	-----	Déséquilibre de Liaison
cM	-----	centi-Morgan
pb	-----	paire de bases
RFLP	-----	Restriction Fragment Length Polymorphism
GL	-----	groupe de liaison
AFLP	-----	Aamplified Fragment-Length Polymorphism

SOMMAIRE

1. Introduction

- 1.1. Taxonomie et distribution géographique page 1
- 1.2. Caractéristiques biologique et botanique de la plante page 1
- 1.3. Amélioration du palmier à huile page 3
- 1.4. Génétique d'association et déséquilibre de liaison page 5
- 1.5. Contexte et objectif de l'étude page 6

2. Matériels et Méthodes

- 2.1. Matériel végétal page 8
- 2.2. Analyse moléculaire page 9
- 2.3. Analyse du DL page 10

3. Resultats

- 3.1. Caractéristiques des marqueurs SSR page 12
- 3.2. Vérification de la légitimité page 12
- 3.3. Structuration des origines page 13
- 3.4. Déséquilibre de Liaison page 13

4. Discussion

- 4.1. Généralité page 16
- 4.2. Préalable à l'étude de déséquilibre de liaison page 16
- 4.3. Le Déséquilibre de Liaison: page 17

1 Introduction

1.1 Taxonomie et distribution géographique

Les *Elaeis* sont des monocotylédones qui appartiennent à l'ordre des *Arécales* (Cronquist, 1981) à la famille des *Arecaceae* (*Palmae*), tribu *Cocoeae* et sous-tribu *Elaeidinae*. Le genre *Elaeis* comprend deux espèces :

- *Elaeis guineensis* Jacq. ($2n = 32$) (photo 1)
- *Elaeis oleifera* (Kunth) Cortés (photo 2)

Le nom *Elaeis guineensis* fut attribué au palmier à huile depuis la Martinique en 1763, par le botaniste Néerlandais Jacquin. L'espèce est originaire du golfe de Guinée (Jacquin, 1763) et était déjà présente à l'époque du Miocène (Ergo, 1997). L'aire naturelle du palmier à huile s'étend du Sénégal à l'Angola (figure 1) de manière continue sur une bande de 200 à 300 kilomètres (Schad, 1914).

Dans son milieu naturel le palmier à huile peut vivre jusqu'à 150 ans. Les conditions pédoclimatiques naturelles favorables au palmier à huile sont celles des régions tropicales humides de basse altitude entre 0 et 300 m au dessus du niveau de la mer. Ces régions présentent des sols profonds meubles, drainés et de type argilo-sableux. Le palmier à huile a besoin d'une pluviométrie annuelle d'au moins 1800 mm, sans déficit hydrique chronique (Henry, 1958) et de 1800 heures d'ensoleillement annuel (Ferwerda, 1977). Les températures optimales se situent entre 24 et 28°C.

1.2 Caractéristiques biologique et botanique

1.2.1 Appareil végétatif : Le système racinaire fasciculé est caractéristique d'une monocotylédone. Ses racines forment un faisceau radial dont l'étendue moyenne avoisine la hauteur de l'appareil aérien. La partie aérienne de la plante adulte est constituée d'un stipe surmonté d'une couronne foliaire de 40 à 50 palmes déployées. Le stipe de forme tronconique à la base et sans ramification est caractérisé par des entre-nœuds courts correspondant à l'attache des palmes. L'élongation est assurée par un méristème apical unique protégé au cœur du bouquet foliaire. L'accroissement en hauteur, variable selon le génotype et les conditions environnementales, est de l'ordre de 25 à 75 cm par an, il diminue fortement avec l'âge de la plante. Le stipe mesure généralement entre 12 et 18 mètres pour des individus de 25 ans en palmeraies cultivées, mais il peut atteindre jusqu'à 25 - 30 mètres chez des palmiers très âgés des populations naturelles. Outre un rôle de soutien grâce à des fibres internes



Photo 1 : *Elaeis oleifera* (Kunth)



Photo 2 : *Elaeis guineensis* Jacq



Photo 3 : Inflorescence femelle en anthèse



Photo 4 : Inflorescence mâle en anthèse

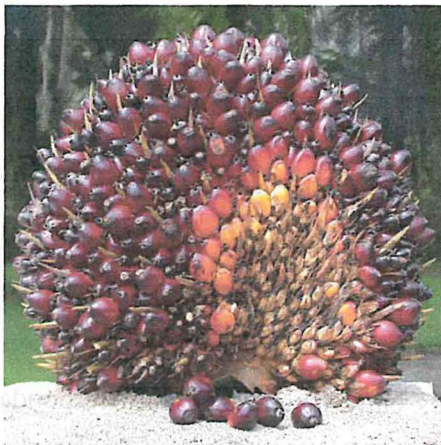
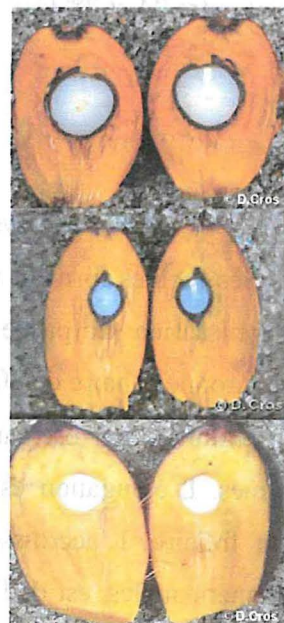


Photo 5 : Régime mature



Pisiféra

Tenera

Dura

Photo 6 : Différents type de fruit

ligneuses, le stipe assure le transport et le stockage des éléments nutritifs. La couronne foliaire est surmontée de jeunes feuilles en flèche, non déployées, qui se succèdent les unes aux autres. En condition normale, un palmier adulte déploie une feuille tous les 15 jours environ. Une période de 18 à 24 mois sépare le stade flèche du début de la sénescence de la feuille. Les feuilles pennées comportent un pétiole de section sensiblement triangulaire et mesurent environ 1,5 m.

1.2.2 Appareil reproducteur : *E. guineensis* se caractérise par des cycles alternatifs irréguliers de production de fleurs mâles ou femelles sur le même pied. Une inflorescence est émise à l'aisselle de chaque feuille (photo 3 &4), à partir de la 2ème année après la plantation et tout au long de la vie de la plante. L'absence de recouvrement de la maturité sexuelle d'une inflorescence à une autre implique une allogamie stricte. La pollinisation des fleurs femelles est anémophile et entomophile. La plante produit dès l'âge de trois ans (en conditions favorables) des régimes volumineux et compacts portant entre 200 et 4000 fruits sessiles de forme oblongue, de 2 à 5 cm de long et de 3 à 30 g (photo 5). Le poids des régimes augmente avec l'âge.

1.2.3 Fruits : Le fruit du palmier à huile est une drupe. Il se compose d'une pulpe huileuse (mésocarpe), d'un noyau (endocarpe), dénommé coque, et d'une amande également huileuse.

L'épaisseur de la coque chez la drupe est gouvernée par le gène Shell (Sh) qui comporte deux allèles codominants (Sh^+ et Sh^- ; Beirnaert et Vanderweyen, 1941). On distingue trois types : (photo 6)

- le type *Dura* : épaisseur de coque >2 mm Sh^+ / Sh^+
- le type *Pisifera* : absence de coque Sh^- / Sh^-
- le type *Tenera* : épaisseur de coque <2mm Sh^+ / Sh^-

Les individus de type *Dura* peuvent être utilisés comme géniteurs femelles en croisement avec des fleurs mâles d'individu de type *Pisifera* (qui à l'état naturel ont un avortement quasi systématique), afin de produire des individus hybrides *Tenera*.

1.3 Huiles et importance économique :

Le palmier à huile est le seul oléagineux produisant deux huiles différentes, l'huile de palme et l'huile de palmiste. L'huile de palme en forte teneur dans la pulpe du fruit (50-55%) est un produit d'élimination métabolique. C'est une huile assez riche en acide gras saturé (45%), dont 40 à 50% d'acide palmitique (C16 :0) et est riche en carotènes.

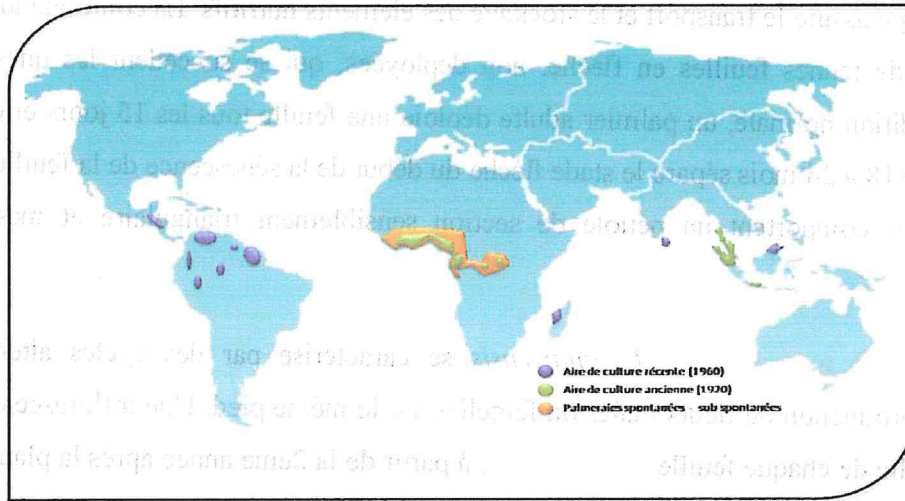


Figure 1 : Aire de répartition du palmier à huile (Selon jaquemard 1995)

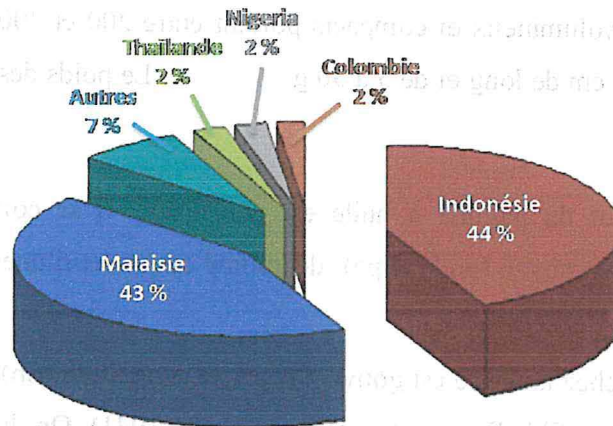


Figure 2 : Principaux pays producteurs d'huile de palme

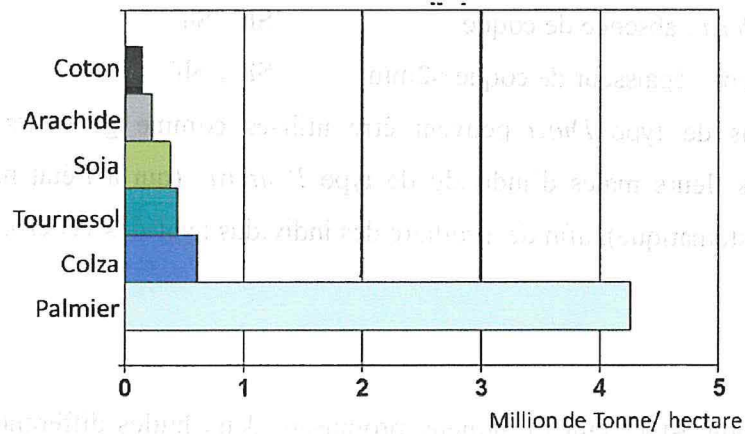


Figure 3: Rendement des différentes huiles mondiales

L'huile de palmiste extraite de l'amande est un produit de réserve. C'est une huile plus riche en acides gras saturés ($\pm 78\%$) que l'huile de palme (Gascon et Wuidart, 1975).

L'Indonésie et la Malaisie sont les principaux producteurs mondiaux (figure 2). L'huile est principalement utilisée dans les pays en voie de développement.

Depuis une dizaine d'année la production d'huile de palme connaît une croissance spectaculaire. D'après les statistiques de l'USDA (United States Department of Agriculture), la production d'huile de palme dans le monde est passée de 19 millions de tonnes en 1999 à plus de 40 millions de tonnes en 2008. Le palmier à huile possède le rendement le plus productif (figure 3) et le coût de production le plus faible de tous les oléagineux. Mais ce développement est réalisé au détriment des forêts asiatiques et de la faune y habitant, notamment les orangs outang.

Il est donc impératif que la culture du palmier à huile s'inscrive dans une dynamique d'agriculture durable. L'amélioration des pratiques culturales et la sélection variétale (résistance aux maladies, vitesse de croissance diminuée...) permettront de s'inscrire dans cette approche.

1.4 Amélioration du palmier à huile :

1.4.1 Historique : L'amélioration du palmier à huile a commencé à se développer dans les années 20, avec pour objectif l'augmentation de la teneur en huile.

En 1941, Beinaert et Venderweyen firent la découverte de l'hérédité de l'épaisseur de la coque, amenant les bases de la sélection actuelle du palmier à huile : les croisements *Dura* x *Pisifera* conduise à une production en huile de palme de 20% supérieure à celle des variétés traditionnelles *Dura*.

Dans les années 1950, l'IRHO (l'Institut de Recherche pour les Huiles et Oléagineux) à organisé le premier échange mondial de matériel végétal, connu sous le nom d'Expérience Internationale, entre cinq stations de recherches du palmier à huile. Ce programme de recherche a montré un phénomène d'hétérosis pour les croisements entre les populations africaines et les populations Deli (Asie). En se basant sur ces résultats l'IRHO (puis le CIRAD) a adapté le schéma de sélection récurrente réciproque dès 1957.

Ce schéma (figure 4) exploite le phénomène d'hétérosis obtenu par le croisement hybride entre des géniteurs issus des deux groupes améliorés l'un par rapport à l'autre de populations pour les composantes de rendement (Meunier et Gascon, 1972) :

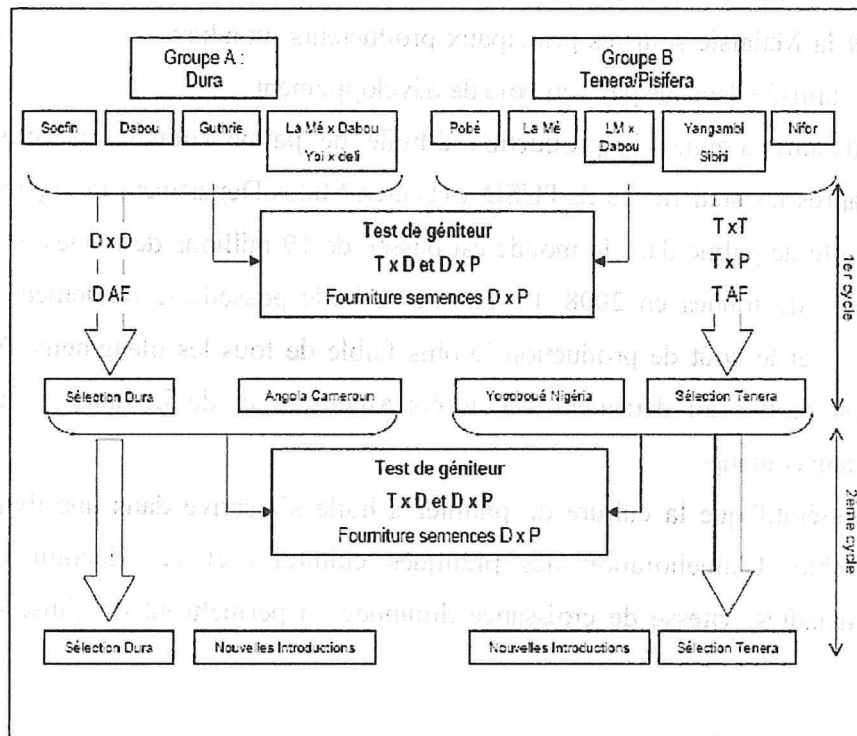


Figure 4: Schéma de Sélection Récurrente Réciproque adopté dans le programme d'amélioration du palmier à huile (d'après Meunier et Gascon, 1972) (D=Dura, T=Tenera, P=Pisifera, AF=AutoFécondation)

- un groupe A de populations Deli : géniteur femelle Dura, produisant un petit nombre de gros régime

- un groupe B de populations Africaines : géniteur mâle Tenera ou Pisifera, produisant un grand nombre de petit régime.

Deux cycles de sélection récurrente ont été achevés et chacun a apporté 12 à 18% d'amélioration de la productivité (Gascon *et al.* 1988, Cochard *et al.* 1993).

A l'heure actuelle, en raison des contextes économiques et environnementaux, les objectifs d'amélioration du palmier se sont diversifiés. En plus de l'augmentation du rendement en huile (Jacquemard et Baudouin, 1987), d'autres critères sont apparus :

- concernant la résistance génétique aux maladies :

- recherche de matériel végétal tolérant à la Fusariose, maladie majeure en Afrique due au champignon *Fusarium oxysporum f. sp. Elaeidis* (Renard *et al.* 1972)

- recherche de matériel végétal tolérant à la Pourriture du Cœur, présente en Amérique Latine (Mariau *et al.*, 1992)

- recherche de matériel végétal tolérant à le Ganoderma, présent en Asie de sud-est (de Franqueville *et al.*, 2001).

- déterminant la durée de vie économique de la plantation :

- faible croissance en hauteur du stipe (Jacquemard, 1979).

1.4.2 Diversité génétique : Pour comprendre la complémentarité entre les origines africaines et l'origine Deli et pour diversifier la base génétique utilisée dans les programmes de sélection, la structure des ressources génétiques déjà impliquées dans la sélection du palmier à huile ainsi que les populations subsponsanées doit être étudiée. Des études ont été menées à l'aide d'isoenzymes (Ghesquière, 1983), puis avec des marqueurs moléculaires (Maizura, 1999, Hayati *et al.* 2003; Cochard *et al.*, 2009). Ces marqueurs permettent d'identifier du polymorphisme entre différents individus. Ils permettent ainsi de calculer des distances génétiques entre individus ou entre groupes d'individus. Ils peuvent aussi être utilisés comme une assistance à l'amélioration, par exemple pour maximiser l'hétérozygotie résiduelle, pour éviter la dépression de consanguinité ou pour identifier des géniteurs ayant une forte homozygotie .

Dans les populations naturelles une forte diversité génétique est attendue dans le centre d'origine, cette diversité tend à diminuer au fur et à mesure de l'éloignement du centre d'origine (Maxted *et al.*, 1997). Une région englobant le Cameroun et le Nigeria pourrait donc

correspondre au centre d'origine du palmier à huile (Cochard, 2008). La population Deli *Dura* possède une faible diversité génétique, l'absence d'allèles rares dans cette population s'explique par la base génétique étroite dont elle est issue. Certains allèles ont pu être éliminés par la sélection ou n'étaient pas présent chez les fondateurs (Hayati *et al.*, 2003). On parle donc d'effet fondateur ou de « bottleneck » (goulot d'étranglement) dû à l'établissement d'une nouvelle population à partir d'un faible nombre d'individus

Les résultats de Cochard (2008) et Cochard *et al.* (2009) révèlent deux groupes originaux qui reflètent la discontinuité des espèces africaines dans la région de Dahomey-Gap : d'une part l'Afrique de l'ouest (Groupe I) et d'autre part le « Bénin-Nigeria-Cameroun-Congo-Angola » (Groupe II). Le groupe Deli (Groupe III) qui est le résultat d'une sélection humaine organisée dérive du groupe II. Comme conséquence pour l'amélioration, la structuration de l'échantillon suggère de réaliser des croisements de type (Deli x Yangambi) x La Mé pour exploiter à la fois la complémentarité entre les groupes I et II et bénéficier des qualités agronomiques de l'origine Deli.

1.4.3 carte génétique et détection de QTL : La première carte génétique du palmier à huile à été construite à l'aide de marqueurs AFLP (Mayes *et al.*, 1997). Celle-ci était incomplète, avec 24 groupes de liaison pour 16 paires de chromosomes homologues. Elle représentait 860 cM de longueur. Une carte génétique saturée à ensuite été construite à l'aide de marqueurs AFLP et microsatellites à partir du croisement LM2T X DA10D (Billotte *et al.*, 2005). Elle contient 944 marqueurs (255 microsatellites, 688 AFLP et l'allèle Sh⁺) répartis sur 16 groupes de liaison (. Elle présente une longueur de 1730 cM avec une densité moyenne d'un marqueur tous les 1.8cM. La recherche de QTL par cartographie génétique multi parentale connectée a permis de détecter 42 QTL pour les principales composantes de la production en huile du palmier à huile (Billotte *et al.*, 2004).

1.4.4 De la cartographie génétique à l'étude du DL: La détection de QTL par cartographie génétique présente plusieurs inconvénients :

- La variation génétique dans ces populations est limitée : chez les diploïdes seulement quatre allèles maximum peuvent ségréger, de plus, en l'absence de polymorphisme allélique entre les parents, les QTL ne peuvent pas être détectés.

- le matériel d'étude n'est pas représentatif du matériel élite utilisé en cours de sélection (Janninck *et al.*, 2001)

Principales mesure du Déséquilibre de Liaison

Première mesure D : (Lewontin en 1964, revue dans Jorde, 2000) dépend de la fréquence allélique. Cet estimateur mesure la différence, pour une paire de locus bialléliques, entre la fréquence observée de deux allèles et la fréquence attendue de cet haplotype si les allèles s'associaient de façon aléatoire.

Soient deux locus polymorphes :

- le locus A avec deux allèles A et a en fréquence respectives P_A et Q_a ;
- le locus B avec deux allèles B et b en fréquence respectives P_B et Q_b .

Il existe alors quatre combinaisons gamétiques : AB, Ab, aB et ab de fréquences respectives P_{AB} , P_{Ab} , P_{aB} , P_{ab} .

Le déséquilibre de liaison D entre les locus A et B s'écrit
 $D = P_{AB} - P_A \times P_B$

Une valeur nulle de D n'indique que les allèles aux deux locus s'associent au hasard dans la population et toute valeur différente de zéro met en évidence un déséquilibre entre ceux-ci. La mesure D est comprise entre -0.25 et 0.25. Le D

Deuxième mesure D' : déterminé en divisant D par le maximum absolu de D. $D' = 1$ si deux allèles ne sont pas séparés par une recombinaison durant l'histoire.

D' , où $D'_{ij} = D_{ij} / D_{max}$, (Lewontin, 1998), qui permet une normalisation de D. Avec :

$$D_{ij} = x_{ij} - p_i p_j$$

Et $D_{max} = \min[p_i q_j, (1-p_i)(1-q_j)]$ si $D_{ij} < 0$
 Ou $D_{max} = \min[p_i(1-q_j), (1-p_i)q_j]$ si $D_{ij} > 0$

Troisième mesure r^2 ou Δ^2 : Elle est complémentaire à D' et est devenue la mesure pour quantifier et comparer le déséquilibre de liaison. Cette mesure représente la corrélation des allèles aux deux sites et est formé en divisant g par le produit des quatre fréquences alléliques aux deux locus : $r^2 = 1$ si les marqueurs ne sont pas séparé par une recombinaison et ont la même fréquence allélique.

$$r = \frac{D}{\sqrt{P_A P_a P_B P_b}} \quad \text{Ou} \quad r^2 = \frac{D^2}{P_A P_a P_B P_b}$$

Ces mesures sont adaptés aux cas où les phases haplotypes sont connues. Dans le cas contraire, lorsque on ne dispose que de données génotypiques, le calcul du déséquilibre de liaison est plus compliqué.

D'après Lewontin, 1964, Hedrick, 1987, Ardlie et al, 2002)

- résolution faible : les intervalles de confiance de la localisation de QTL varient entre 10 et 20 cM (Davarsi *et al*, 1993) et la résolution est déterminée par la taille de la descendance en ségrégation.

Ces différentes limites ont amené à s'intéresser à des stratégies différentes, utilisées en génétique humaine et animale : la cartographie par Déséquilibre de Liaison ou la cartographie d'association. Cette dernière repose sur les variations produites par l'évolution, c'est-à-dire les recombinaisons historiques (Yu and Buckler, 2006).

Le déséquilibre de liaison (DL) aussi appelé déséquilibre gamétique, se définit comme l'association non aléatoire de deux allèles situés à des locus différents. Il représente l'écart à la situation d'équilibre de liaison pour laquelle les allèles aux différents locus considérés ségrégent de façon indépendante dans la population. Le DL décrit donc une association statistique entre deux locus (Ardlie *et al*, 2002). Deux locus situés sur des chromosomes différents et donc non liés physiquement peuvent se trouver en déséquilibre. Le déséquilibre de liaison est le reflet de la dérive génétique, des mutations, des goulets d'étranglements, des migrations, des effets d'épistasie et de l'histoire démographique des populations.

(encadré 1 &2).

1.5 Contexte et objectif de l'étude :

Pour identifier des locus associés à des caractères agronomiques par génétique d'association, il est utile, au préalable, d'étudier l'étendue du Déséquilibre de Liaison (DL) entre deux locus afin de s'assurer que deux marqueurs, ou un marqueur et le caractère phénotypique soient bien associés.

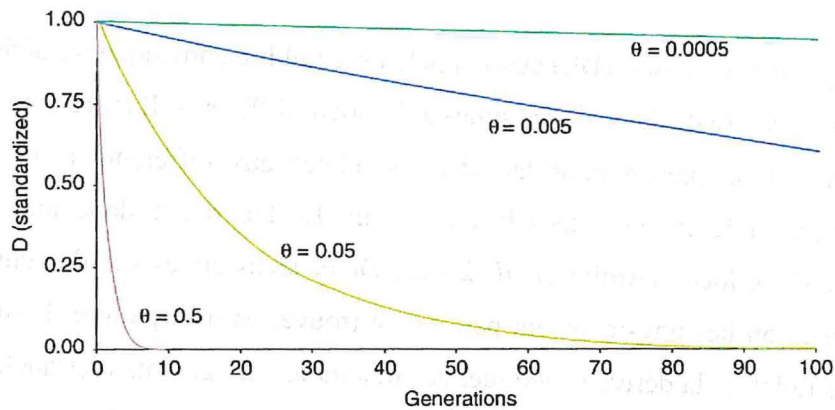
Une étude du DL local, préliminaire, sur un ensemble de locus très proches du Groupe de Liaison 8 (GL8) a été réalisée dans le cadre d'un stage de Master Pro en « Génétique et gestion de la biodiversité » (Rekima, 2006). Cette étude du DL local a été précisée et complétée par une étude du DL pangénomique (Cochard, 2008).

Cette étude du DL a été réalisée sur 3 origines améliorées, représentant le pool du matériel sélectionné par le CIRAD. Au total, 74 marqueurs SSR ont été retenus pour caractériser le DL, à raison d'environ 2 couples de locus par GL et deux segments ciblés des GL8 et GL15. L'estimation du DL a été réalisée à partir de données haplotypiques. Celles-ci ont été obtenues grâce à la connaissance plus ou moins complète des généalogies. Ces études ont montré que l'étendue du DL varie selon les origines, de 5 à 24 cM et en fonction des GL.

Facteurs influençant le Déséquilibre de Liaison

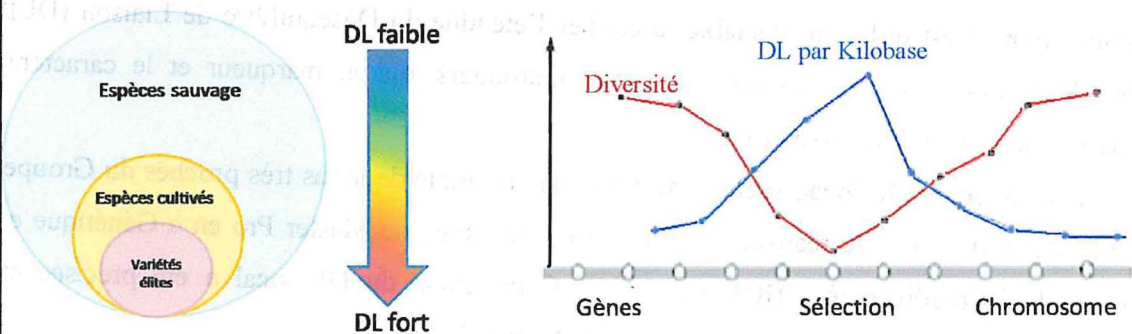
a. Diminution du Déséquilibre de Liaison au cours du temps :

En situation de croisements aléatoires, en l'absence de mutations, et de sélections, la valeur du déséquilibre de liaison, D , dans les génération successives décroît en fonction des taux de recombinaison. Plus le taux de recombinaison, r , est élevé, plus le Déséquilibre de Liaison décroît rapidement.



b. Variation du Déséquilibre de Liaison selon le type de population

Le type de population va influencer le déséquilibre de liaison. Les goulots d'étranglement au niveau des variétés élites, la pression de sélection, la structure de la population, sont des causes majeurs de variation du Déséquilibre de Liaison. De même, le système de reproduction de la population, peut faire décroître rapidement le Déséquilibre de Liaison (plantes allogames).



Représentation schématique des effets de la sélection sur le Déséquilibre de Liaison et la diversité génétique

L'existence d'un DL important a permis la réalisation d'une étude d'association entre quelques marqueurs du GL8 et le caractère de la hauteur. La présence de QTL précédemment détecté par cartographie multiparentale a été confirmée. Cette étude a confirmé la possibilité de détecter des associations, des QTL sur des fonds génétiques plus importants que ceux utilisés dans la détection de QTL (Cochard, 2008).

L'échantillonnage des marqueurs sur les GL8 et GL15 représentaient des segments respectivement de 135,2 et 80,9 cM. Suite aux études précédentes, il a paru utile de les compléter en réalisant un nouvel échantillonnage de façon répartie le long des GL, avec une distance entre marqueur de l'ordre de celle de l'étendue du DL préalablement détecté. Ceci permettra de mieux préciser l'étendue du DL pour ces 2 GL.

Cela permettra de définir le nombre de marqueurs nécessaires à la recherche de QTL dans les 3 populations, dans le cadre d'une approche genomescan.

Cette méthode permettra l'analyse du génome entier du palmier à huile avec un ensemble de marqueurs microsatellites dont le pas sera défini par l'étendue du DL. Elle permettra également l'identification de régions d'intérêt par une étude de génétique d'association.

2 Matériels et méthodes :

2.1 Matériel végétal :

L'échantillonnage de notre étude était composé de 327 individus, dont 6 ont été estimés à partir de 6 individus de leur autofécondation. L'ADN de deux cent cinquante quatre échantillons d'études précédentes (Rekima 2006 ; Cochard 2008) était disponible en quantité suffisante pour notre étude. Nous avons extrait 67 échantillons supplémentaires. Nous avons utilisé un lot de témoins définis pour toutes les études de marquage moléculaire. Du fait que les extractions étaient anciennes et pour vérifier la qualité de la dilution, nous avons testé la qualité des ADN des 327 individus à l'aide de l'amorce mCnCIR38. Cette amorce est utilisée en routine pour valider la qualité des ADN.

2.2 Analyse moléculaire :

2.2.1 Extraction d'ADN génomique : Les prélèvements des folioles de la plus jeune feuille ouverte (la flèche) ont été réalisés pour chaque palmier, sur deux principaux sites ; à Pobè (CRAPP, INRAB Bénin) et la Mé (CNRA, Côte d'Ivoire). Les échantillons ont été conditionnés dans des glacières. Le temps de transport variait de 3 à 5 jours. A réception, les échantillons foliaires ont été dénervurés, puis lyophilisés sous vide durant 4 à 7 jours à -40 °C. Après obtention d'une poudre fine par broyage, l'ADN génomique a été extrait. Deux kits commerciaux ont été utilisés indifféremment (DNeasy Plant 96, Allemagne) ou Nucleospin 96 plants (Marechey-Nagel, Allemagne).

Les feuilles sont délayées et lysées dans un mélange de tampon AP1 (90µl), de Rnase A (225µl, 100 mg/ml) et d'un réactif DX (agent anti-moussant, 225 µl), puis culotées par centrifugation (5min, 6000 tr/min), Les protéines et les polysaccharides sont précipités avec le tampon AP2 (130µl), après incubation pendant 10 min à -20°C, ils sont culotés par centrifugation (5min, 6000 tr/min), le tampon AP3/E permet la fixation de l'ADN sur la micro-colonne échangeuse d'ions. L'ADN est élué par ajout d'un tampon AE (100µl) et centrifugation (2min, 6000tr/min), cette dernière étape est répétée une fois (annexe 1).

2.2.2 Dosage de l'ADN génomique : Les concentrations d'ADN génomique ont été estimées à l'aide de deux techniques :

- par fluorimétrie : (Fluoroskan Ascent[®], Thermo Fisher Scientific[®], USA) : un volume de 2 µl de chaque extrait d'ADN génomique est prélevé et ajouté à 200 µl de mélange réactionnel (27 µl d'H₂O, 3 ml TNE 10X sigma, 6 µl HOESCHT bis benzidine 0,1 µg/ml). Une gamme étalon est réalisée à partir d'ADN de thymus de veau.

- Par absorbance de lumière : (Tecan, Infinite[®] R 200 NanoQuant, Suisse) un volume de 2 µl de chaque extrait d'ADN génomique est disposé sur une microplaque NanoQuante Plate[™]. L'absorbance est mesurée par un laser quad4 Monochromators[™] qui utilise 4 filtres monochromes simultanément. Les résultats sont sous formes de donnée Excel. Les échantillons sont conservés à une concentration de 20 ng/µl ou 5 ng/µl à une température de - 80 °C.

2.2.3 Amplification PCR : Les SSR ont été génotypés selon la méthode décrite par Billotte *et al* (2005). Pour chaque couple d'amorces, une séquence 5'-CACGACGTTGTAAAACGAC-3' de 19 nucléotides a été ajoutée sur l'extrémité 5' de l'amorce Forward. Cette extension est nommée queue M13. Sa séquence est identique à celle de l'amorce universelle M13 Forward marqué en 5' par un fluorochrome de type IR 700 ou IR 800. L'amorce universelle M13 Forward va s'hybrider à la séquence complémentaire de la queue M13 permettant ainsi de visualiser les produits d'amplification grâce au signal fluorescent émis (Roy *et al.* 1996).

Les amplifications PCR (polymerase Chain Reaction) ont été réalisées sur différents thermocycleurs, MJ Recherche PTC-100[™] et Mastercycler (Eppendorf).

Les réactions d'amplifications par PCR sont effectuées avec 5 µl de solution d'ADN à 5 ng/µl dans un volume réactionnel total de 10 µl. Le mélange utilisé contient 1 µl de Tampon PCR10X [10 mM Tris-HCl (pH8), 100 mM KCl, gélatine 0.5 % (w/v)], 1 µl de dNTP à 2 mM, de 0.12 µl de MgCl₂ à 50 mM, 0,08 µl de l'amorce portant la queue M13 (10 µM), 0,1 µl pour l'autre amorce (10 µM), 0.1 µl de l'amorce M13 fluorescente à 700 IR ou 800 IR, 0,5 µl de Taq (2U), et est complété par 2,1 µl d'eau Merck.

Le programme d'amplification est le suivant : première phase de dénaturation pendant 5 min à 94°C suivie de 35 cycles, constitué chacun d'une phase de dénaturation de 30 secondes à 94°C, puis d'une phase d'hybridation pendant 1 min 15s à 52°C, d'une élongation pendant 1 min 30 s à 72°C et enfin une élongation finale de 8 min à 72°C.

La PCR terminée, les produits d'amplifications sont conservés à - 20°C..

2.2.4 Migration des produits d'amplification : Le réarrangement des produits d'amplification dans des plaques 96 puits se fait à la main. 2µl du produit d'amplification en 700 et 3µl du produit d'amplification en 800 sont prélevés. Ajout de 5µl de bleu de charge contenant les bandes 'desmiling' utilisées pour évaluer la régularité du gel. La migration est effectuée sur un gel polyacrylamide à 6,5 % d'une épaisseur de 2mm (20 ml Polyacrylamide 6,5%, 175µl APS, 25 µl TEMED). Le gel est coulé entre 2 plaques de verre d'épaisseur 25 mm. Cet ensemble est monté sur un analyseur Li-Cor. Le tampon de migration est le TBE 1 X (Tris-Borate-EDTA). Le gel est ensuite préchauffé à 48 °C pendant 30 min.

Après dénaturation (3min à 94°C), 1µl de chaque échantillon est déposé par capillarité à l'aide d'un peigne à 100 dents sur le gel d'acrylamide, accompagné d'un marqueur de taille.

Les échantillons amplifiés sont analysés selon les conditions suivantes : 1300V, 35mA, 40W, 48°C. Les migrations durent de 1h à 2h selon le poids moléculaire des allèles. La détection s'effectue en temps réel grâce à l'excitation par deux lasers du fluorochrome fixé à l'amorce M13(annexe 2).

2.2.5 Lecture des gels : L'image du gel est directement numérisée et prête pour l'analyse. La lecture des gels été faite à l'aide du logiciel SAGA ^{GT}® (Li-Cor, Lincoln, USA). Les allèles sont identifiés selon leur taille en paire de base.

2.2.6 Choix des microsatellites : Une carte génétique consensus a été établie à partir de 4 croisements d'un factoriel 2x2 à l'aide du logiciel Carthagene (de Givry *and al*, 2005). Cette carte génétique comprend 253 marqueurs SSR, 1 marqueur AFLP et le locus Sh. Cette carte est saturée et comprend 16 GL. Elle mesure 1731 cM avec une densité moyenne de marquage de 7 cM (annexe3). Les marqueurs ont été choisis sur les GL8 et GL15. Trente six marqueurs ont été choisis sur cette carte, dont 16 ont été préalablement utilisés sur cet échantillon. Au cours de ce stage, nous avons génotypé les individus à l'aide de 23 marqueurs SSR. Les tailles, en paires de base (pb), et la distribution approximative des allèles pour 17 marqueurs étaient connues. Pour 6 autres, nous avons dû définir les tailles en pb sur le lot des 6 témoins avant de définir le multiplexage des marqueurs pour les migrations.

2.3 Analyse du DL :

2.3.1 Vérification de la légitimité : La vérification de la légitimité et/ou des erreurs de lecture a été réalisée en utilisant le logiciel PedAgree (<http://pedagree.free.fr> ; Couillaud 2002) en se basant sur la connaissance partielle de la généalogie des 3 origines. Ce logiciel contrôle

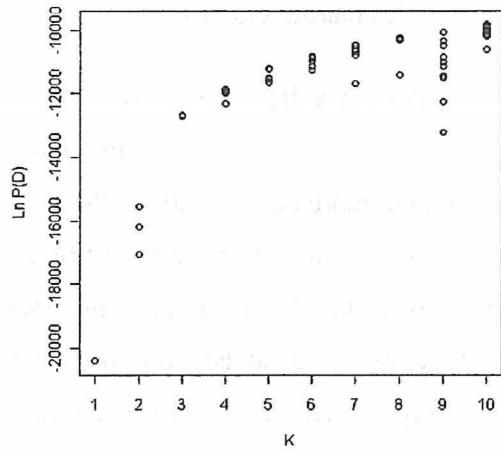
les incompatibilités Mendélienne à différents niveaux. Les invraisemblances entre les parents et leurs descendance sont détectées au sein de chaque famille nucléaire d'un même pedigree. Les incompatibilités entre les familles nucléaires sont ensuite recherchées. Le logiciel prend en compte les « mariages multiples » et la consanguinité. Il prend également en compte les parents non génotypés et peut faire une vérification de la légitimité sans connaissance des parents. De plus, le logiciel précise quels sont les individus et les marqueurs incompatibles.

2.3.2 Effet de la structure : La structure de notre échantillon a été analysée dans notre étude à l'aide du logiciel Structure 2.3.1 (<http://pritch.bsd.uchicago.edu/software> ; Pritchard, 2000) L'analyse a été effectuée sans connaissance a priori du nombre de populations (K). Les paramètres de l'analyse choisis ont été un « burn-in » de 10 000 et un nombre de répétition de 100 000. Nous avons fait varier K de 1 à 10. L'estimation du nombre de populations (K) constituant notre échantillon a été précisée en utilisant la méthode ad hoc proposée par Evanno *et al* (2005). L'étude de la structure a été réalisée séparément pour les locus du GL8 et du GL15.

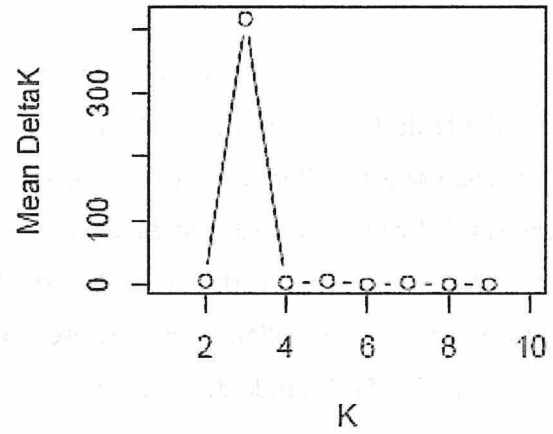
2.3.3 Estimation du Déséquilibre de Liaison : Nous avons estimé le DL à l'aide de la mesure du r^2 . Nous avons utilisé l'extension multiallélique du r^2 (Weir, 1996). Cette mesure permet de comparer facilement le DL. Elle représente la corrélation entre les allèles entre deux locus. Nous avons calculé un r^2 sur des données génotypiques, sans connaissance des haplotypes, à l'aide du logiciel Tassel (<http://www.maizegenetics.net> ; Yu *et al*, 2006).

2.3.4 Seuil de signification du déséquilibre de liaison : L'estimation du seuil de signification de r^2 a été réalisée en adaptant celle proposée par Breseghello et Sorrels (2006). Nous avons déterminé un seuil à partir du 95^e centile déterminé sur la distribution normalisée des DL estimés entre marqueurs non synténiques. En l'occurrence, nous avons choisi 5 marqueurs du GL8 et 5 marqueurs du GL15.

2.3.5 Estimation de l'étendue du DL : Nous avons réalisé un ajustement non linéaire du type $y=pr1 \exp(pr2-x)+pr3$ à l'aide du logiciel XLSTAT (www.xlstat.com) sur les mesures du DL réalisées pour chaque GL. Cet ajustement diffère de celui proposé par Breseghello et Sorrels (2006), où ils proposent un ajustement non paramétrique avec un polynôme de 2nd degré. L'intersection du seuil et de la courbe d'ajustement permet de définir l'étendue du DL.



Calcul du maximum de vraisemblance



Calcul du K maximum

Figure 5 : résultat de l'analyse de la structure de l'échantillon

3 Résultats :

3.1 Caractéristiques des marqueurs SSR :

Après le génotypage de notre échantillon à l'aide de nos 23 marqueurs SSR, puis lecture des gels de migration avec l'aide du logiciel SAGA^{GT}, les marqueurs mEgCIR3593, mEgCIR3718 et mEgCIR3732 présentaient un signal trop faible pour analyser les profils de migration. Les locus mEgCIR3728 et mCnCIRG9 n'ont pu être amplifiés.

La présence systématique de données de génotypage manquantes pour 34 individus a conduit à leur élimination de notre étude.

L'analyse des données génotypiques des 36 marqueurs a permis de mettre en évidence des locus monomorphes pour certaines origines. Pour l'origine La Mé, 2 marqueurs SSR se sont révélés monomorphes (mEgCIR3727 du GL15 et mEgCIR0836 du GL8). Deux autres marqueurs (mEgCIR2860 du GL15 et mEgCIR2440 du GL8) sont monomorphes pour l'origine Yangambi. L'origine Deli, très consanguine, présente le plus de marqueurs monomorphes : 9 (mEgCIR2590 du GL15 et mEgCIR3563, mEgCIR0910, mEgCIR3808, mEgCIR0778, mEgCIR2291, mEgCIR0774, mEgCIR1740 et mEgCIR0163 du GL8).

Les locus monomorphes ont été ensuite retirés de l'étude du DL, pour l'origine concernée. Les locus présentant trop de données manquantes ont aussi été éliminés de l'analyse.

En conséquence, l'étude du DL a été réalisée avec 12 marqueurs pour chaque GL pour l'origine Deli, 19 marqueurs pour le GL8 et 12 pour le GL15 pour l'origine Yangambi et 20 marqueurs pour le GL8 et 14 pour le GL15 pour l'origine La Mé

3.2 Vérification de la légitimité :

L'utilisation du logiciel PedAgree (Couillault 2002) a permis de vérifier les erreurs de lectures. Des corrections ont pu être réalisées. Les erreurs de lecture systématiques pour 52 individus ont entraîné leur élimination de notre étude. Ces 52 individus, lors d'une analyse précédente réalisée à partir de 75 marqueurs SSR, se sont révélés légitimes (Cochard 2008). Aussi, dans ce cas ci, il conviendra d'approfondir la vérification des lectures, point que nous n'avons pu faire, faute de temps disponible. Ainsi, notre étude du DL a porté sur 268 individus, 111 de l'origine Deli, 61 de l'origine La Mé et 96 de l'origine Yangambi.

Groupe de Liaison 8

Groupe de Liaison 15

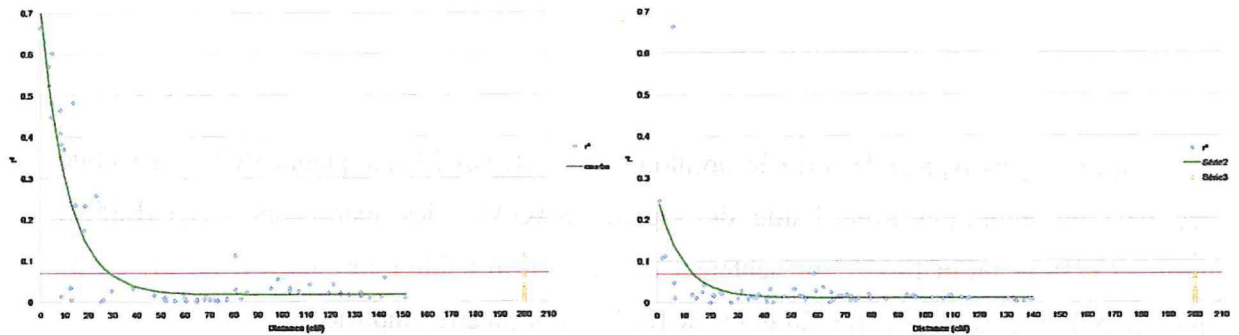


Figure 6 : Evolution du DL (r^2) en fonction de la distance génétique (cM). Origine Déli
 La Ligne horizontale rouge indique le 95^e centile de distribution des valeurs de r^2 des paires de locus non liés. La courbe est une régression non linéaire.

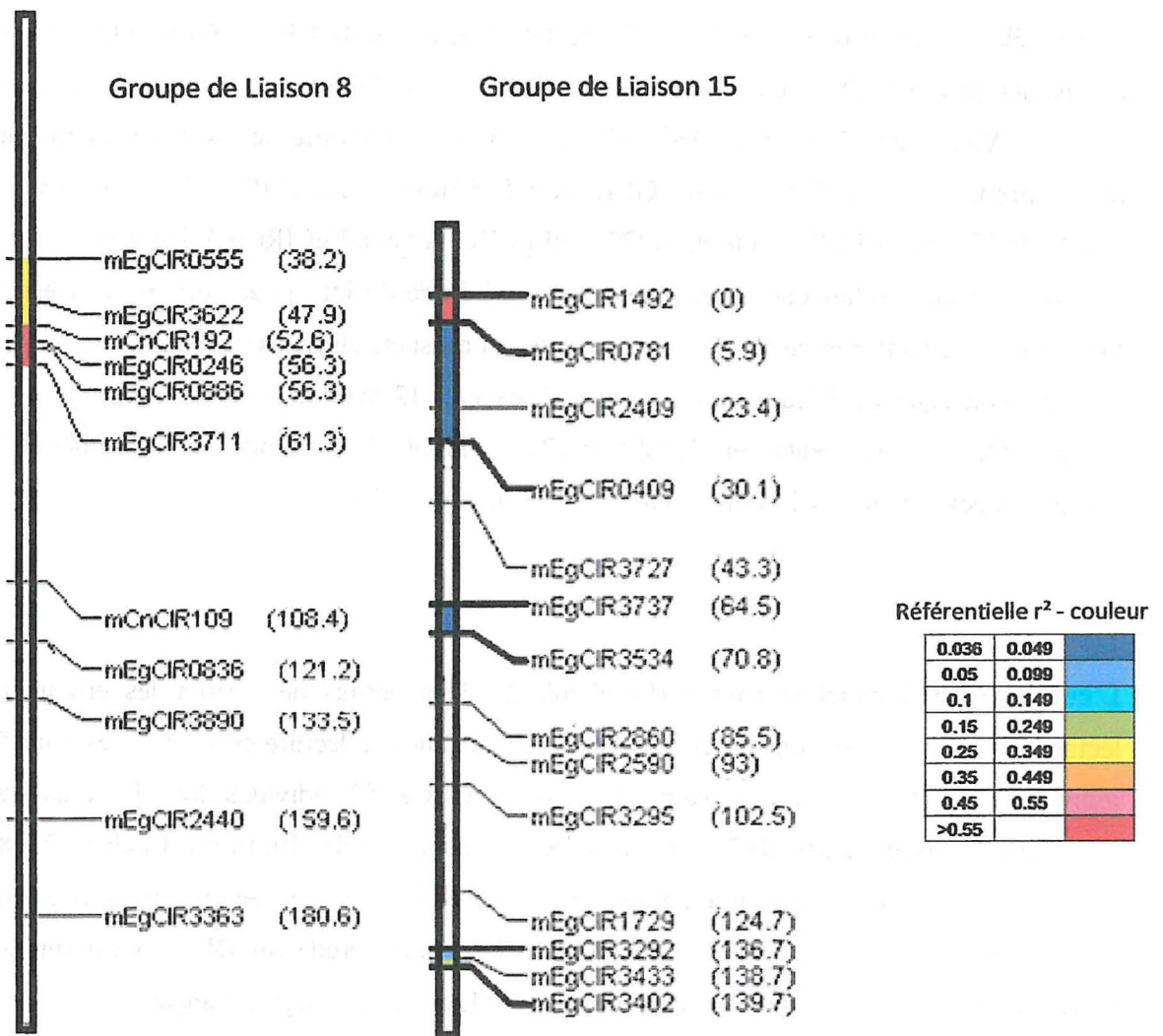


Figure 7: Représentation du DL (r^2) le long des groupes de liaison. Origine Déli

3.3 Structuration des origines :

L'étude de la structure de notre échantillon, réalisée à l'aide du logiciel Structure développée par Pritchard et al (2000) complétée par la méthode d'Evanno *et al.* (2005) permet de montrer, tant pour le GL8 que pour le GL15, que notre échantillon est structuré en 3 groupes d'appartenance (figure 5). Ces groupes d'appartenance correspondent aux 3 origines Deli, Yangambi et La Mé.

3.4 Déséquilibre de Liaison :

Le DL (r^2) a été calculé pour les trois origines et sur les deux groupes de liaison respectivement (annexe 4 &5). Le seuil de signification a été déterminé à partir du 95^e centile des valeurs de r^2 non synténique entre les locus des deux GL.

3.4.1 Population Deli : (figure6 &7 et annexe 4).

La moyenne de r^2 dans les échantillons provenant de la population Deli est de 0,072. Le 95^e centile est de 0,069. Le croisement entre ce seuil et la courbe d'ajustement non linéaire permet d'estimer l'étendue du DL.

Groupe de Liaison 8 : L'étendue du DL est de 29 cM. L'évolution du DL le long du segment du groupe de Liaison 8 (GL), présente un fort r^2 ($>0,15 - >0,55$) sur une longueur de 23,1 cM entre les marqueurs mEgCIR0555 et mEgCIR3711. De plus, entre les marqueurs mCnCIR192 et mEgCIR3711, le r^2 est supérieur à 0,55.

Groupe de Liaison 15 : L'étendue de DL est de 14cM. On peut distinguer trois segments le long du GL possédant un DL plus ou moins fort :

- Au début du GL, étalée sur 23,4 cM, entre les marqueurs mEgCIR1492 et mEgCIR 2409 : r^2 supérieur à 0,45

- Au milieu du GL, 41,1 cM plus loin que la première région, entre mEgCIR3737 et mEgCIR3534 d'une étendue de 6,3 cM avec un r^2 entre 0,036 et 0,050.

- A la fin du GL, sur les 3 derniers cM entre mEgCIR 3292 et mEgCIR 3402 (r^2 entre 0,1 et 0,25).

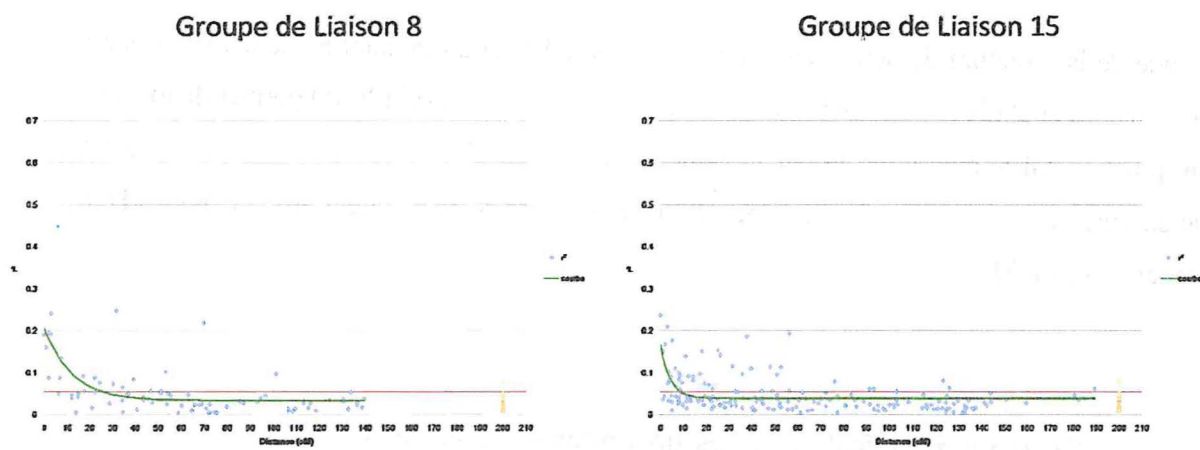


Figure 8 : Evolution du DL (r^2) en fonction de la distance génétique (cM). Origine La Mé
 La Ligne horizontale rouge indique le 95^e centile de distribution des valeurs de r^2 des paires de locus non liés. La courbe est une régression non linéaire.

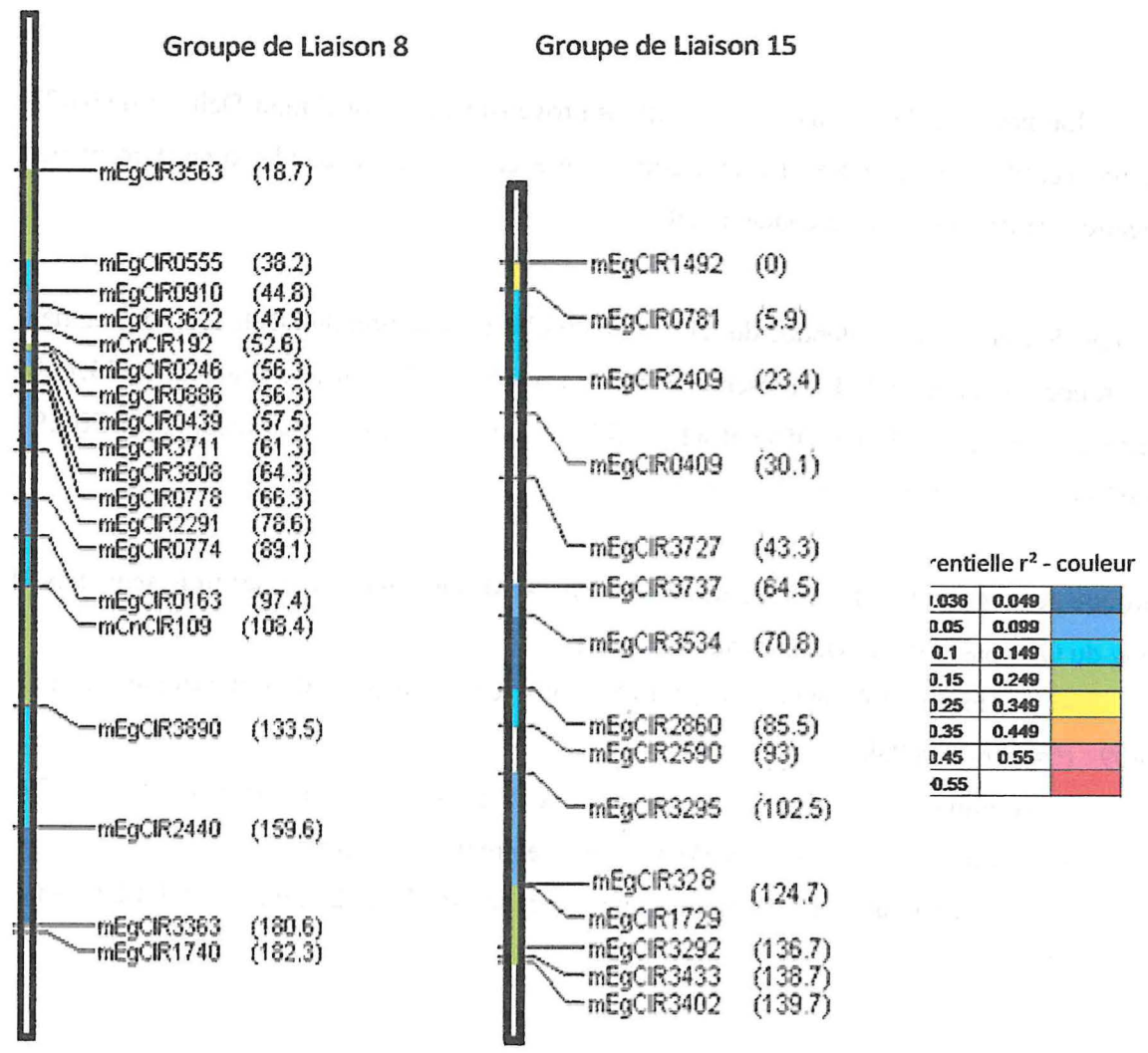


Figure 9 : Représentation du DL (r^2) le long des groupes de liaison. Origine La Mé

3.4.2 Population La Mé : (figure8 &9 et annexe 4).

La moyenne de r^2 dans les échantillons provenant de la population La Mé est de 0,050. et le 95^e centille de 0,055.

Groupe de Liaison 8 : L'étendue du DL est de 10 cM. On peut observer un r^2 variable allant jusqu'à 0.25 presque tout le long du GL. Trois interruptions entre deux marqueurs apparaissent :

- entre mCnCIR192 et mEgCIR0886 (intervalle de 3,7 cM)
- entre mEgCIR3808 et mEgCIR0778 (intervalle de 5 cM)
- entre mEgCIR2291 et mEgCIR0774 (intervalle de 10,5 cM)

Groupe de Liaison 15 : L'étendue de DL est de 29 cM. L'étendue du DL est de la même disposition que le GL8, présent sur une bonne partie du GL, avec une interruption entre les marqueurs mEgCIR409 et mEgCIR3737 de 34,4 cM, et une autre entre les deux marqueurs mEgCIR2590 et mEgCIR3295 de 9,5 cM.

3.4.3 Population Yangambi : (figure10 &11 et annexe 5).

La moyenne de r^2 des couples de marqueur Yangambi est de 0.030 et le 95^e centille de 0.036.

Groupe de Liaison 8 : Le GL8 possède une étendue de seulement 9 cM. L'évolution du DL le long de ce groupe de liaison est difficilement interprétable au vu du nombre de données présentes. Quatre petites zones possèdent un DL supérieur à 0.036 mais ne dépassant pas 0.099 (sauf entre mEgCIR3363 et mEgCIR1740 : $r^2= 0,15$) :

- entre mEgCIR 555 et mEgCIR 910 (intervalle de 6,6 cM)
 - entre mCnCIR 192 et mEgCIR 246 (intervalle de 3,7 cM)
 - entre mEgCIR 3711 et mEgCIR 778 (intervalle de 5 cM)
- entre mEgCIR 3366 et mEgCIR 1996 (intervalle de 8,5 cM)

Groupe de Liaison 15 : Par rapport au seuil calculé (0,036), l'ampleur du DL est de 42 cM. Très peu de données sont présentes au dessus du seuil. Cependant, il y a une zone de 70,8 cM

Groupe de Liaison 8

Groupe de Liaison 15

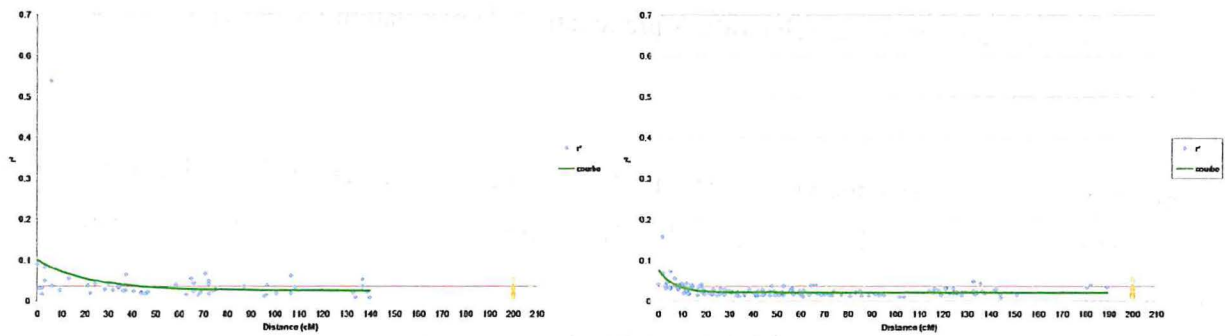


Figure 10 : Evolution du DL (r^2) en fonction de la distance génétique (cM). Origine Yangambi
 La Ligne horizontale rouge indique le 95^e centile de distribution des valeurs de r^2 des paires de locus non liés. La courbe est une régression non linéaire.

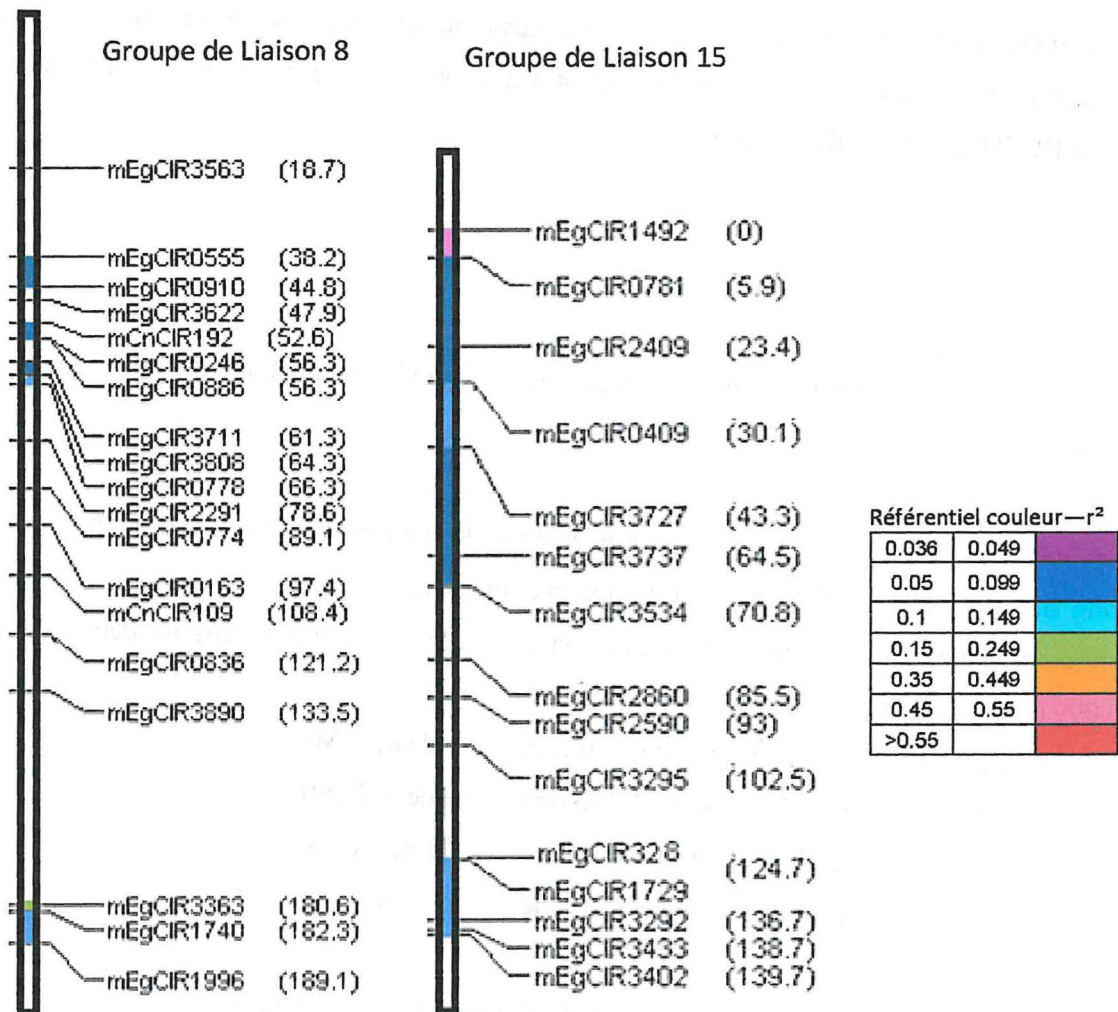


Figure 11 : Représentation du DL (r^2) le long des groupes de liaison. Origine Yangambi

au début du GL présentant des valeurs allant jusqu'à 0,099. Avec, une exception, entre le premier et le deuxième marqueur du GL qui possède un r^2 de 0,53.

4 Discussion :

4.1 Généralités :

Les marqueurs mEgCIR3593, mEgCIR3718, mEgCIR0409 et mEgCIR3732 présentait un signal trop faible pour pouvoir être analysés. Cela peut être dû au programme PCR « NBSAT52 ». Ce programme a été conçu pour être le plus compatible possible avec toutes les amorces disponibles. Il n'est donc pas optimisé (température, stringence, nombre de cycles) pour chaque amorce. L'absence de signal peut aussi être due aux amorces fluorescentes, qui étaient soit trop anciennes soit contaminées. Ces 4 amorces n'ont pas été repassées, faute de temps.

Deux cent cinquante quatre échantillons utilisés lors de cette analyse provenaient d'un lot extrait en 2006. Malgré une conservation à -20°C, leur ADN a pu être dégradé, rendant impossible l'amplification et le génotypage. Pour vérifier cette hypothèse, il faudrait repasser ces échantillons sur un gel d'agarose. Certains échantillons étaient des géniteurs de 1^{ère} génération.

4.2 Préalable à l'étude de déséquilibre de liaison

4.2.1 Vérification des erreurs de lecture

Notre étude du déséquilibre de liaison a été réalisée sur du matériel végétal constitué de différentes origines, chacune en cours d'amélioration. Les croisements étant du type plein frères, la présence d'erreurs de lecture peut être vérifiée si la généalogie est confirmée. La généalogie ayant été confirmée précédemment, pour 73 locus SSR (Cochard, 2008), nous avons pu révéler les erreurs de lectures, éventuellement les corriger, ou en cas de doute, ôter les individus de l'étude.

4.2.2 Etude de la structure

La structure étant un des principaux facteurs influençant le déséquilibre de liaison, nous avons vérifié celle-ci en utilisant notre jeu de marqueurs. Les deux jeux de marqueurs liés, l'un situé sur le GL8, l'autre sur le GL15, ont permis d'obtenir les mêmes résultats que l'étude réalisée sur des locus indépendants (16 locus, 1 locus par groupe de liaison) (Cochard, 2008).

Cette étude, réalisée sur du matériel en cours d'amélioration (1^{ère} et 2^{ème} génération) confirme la structuration génétique mise en évidence sur un ensemble de 9 origines géographiques (Cochard *et al*, 2009). Nous avons une séparation systématique entre les deux origines africaines ; le groupe I est composé des origines ivoiriennes, l'origine Yangambi appartient bien au groupe II (groupe de l'Afrique Centrale plus le Nigeria et le Bénin). Le 3^{ème} groupe met en évidence la différenciation des origines Deli par rapports aux origines africaines, reflet de l'impact de 4 à 5 générations de sélection dans un système très consanguin sur la structure de la diversité génétique.

Les deux études, de vérification des erreurs de lecture et de la structure, montrent que les SSR sont des marqueurs moléculaires très adaptés à ces études, bien qu'ils soient considérés comme hautement variables.

4.3 Le Déséquilibre de Liaison:

4.3.1 Etendue du DL

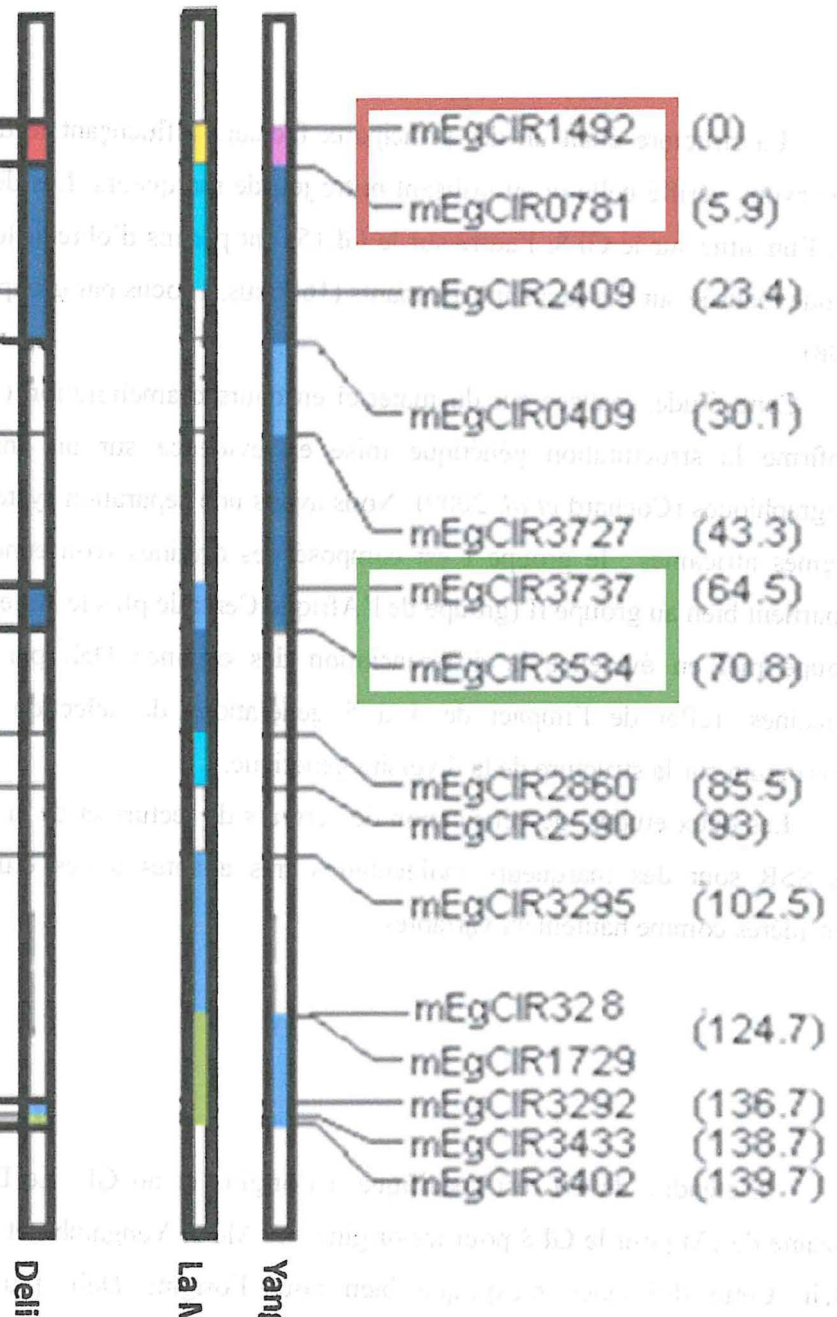
L'étendue du DL est spécifique à l'origine et au GL. Le DL est de l'ordre d'une dizaine de cM pour le GL8 pour les origines La Mé et Yangambi, et de 29 cM pour l'origine Deli. Cette différence s'explique bien pour l'origine Deli, plus consanguine que les précédentes et donc susceptible de présenter davantage de DL.

Cependant, cette explication n'est pas suffisante, puisque pour le GL15, les résultats sont opposés. L'étendue du DL est inférieure pour l'origine Deli. Ce constat remet en cause la considération suivante : l'autofécondation n'augmente pas systématiquement l'étendue du DL.

L'étendue du DL étant dans tous les cas relativement élevée, cela signifie que nous pouvons mener une étude d'association de type genomescan. Ainsi, il est possible de choisir un marqueur environ tous les 9 cM (étendue du DL la plus faible), pour faire une détection de QTL. Au total, cela représente 170 marqueurs à génotyper, sachant que la longueur moyenne des cartes génétiques varie autour de 1600 cM (Billotte *et al*. 2005)

QTL de production:
nombre et poids
moyen des régimes

QTL gouvernant
l'appareil végétatif :
Hauteur du stipe



Référentiel couleur—r²

0.036	0.049	Red
0.05	0.099	Blue
0.1	0.149	Cyan
0.15	0.249	Green
0.35	0.449	Orange
0.45	0.55	Pink
>0.55		Red

Figure 12 : Relation entre DL (r²) et QTL

Ce résultat permet aussi de pouvoir réaliser une sélection de type genomewide, en recherchant et sélectionnant sur l'ensemble des QTL mis en évidence.

4.3.2 Variation du DL au sein des GL

Hormis des segments spécifiques à chacun des GL, les marqueurs avaient été choisis afin de disposer d'un marqueur tous les 10 à 15 cM. Les nombreux marqueurs monomorphes ou absents ont conduit à une distribution irrégulière des marqueurs le long des deux GL. Le DL est rarement conservé pour le même GL, selon les origines.

Il y a cependant deux exceptions qui concernent le GL15. Pour les 3 origines, la paire mEgCIR1492 – mEgCIR0781 distante de 5,9 cM présente systématiquement un $r^2 > 0,35$, ainsi que la paire mEgCIR3737 – mEgCIR3534 distante de 6.3 cM qui présente un DL faible (r^2 entre 0,05 -0,15)

Lorsqu'on compare ces deux zones avec les études précédentes menées sur l'identification de QTL de caractères végétatifs et de production (Billotte, 2004), on s'aperçoit que ces deux zones de DL conservées co-localisent avec deux zones de QTL significatif (Figure12).

La paire mEgCIR1492-mEgCIR0781 co-localise avec un QTL de production (nombre et poids moyen des régimes) et la paire mEgCIR3737-mEgCIR3534 avec un QTL gouvernant l'appareil végétatif : la hauteur du stipe.

De la même manière, les zones sans QTL significatif correspondent aux zones sans DL de notre étude (Figure 12)

La méthodologie utilisée dans les études précédentes pour échantillonner des marqueurs sur l'espèce *Elaeis guineensis* Jacq. est donc validée. Il y a bien une relation entre DL et QTL. Ces résultats rendent donc possible une étude d'association de type genomescan.

ANNEXES

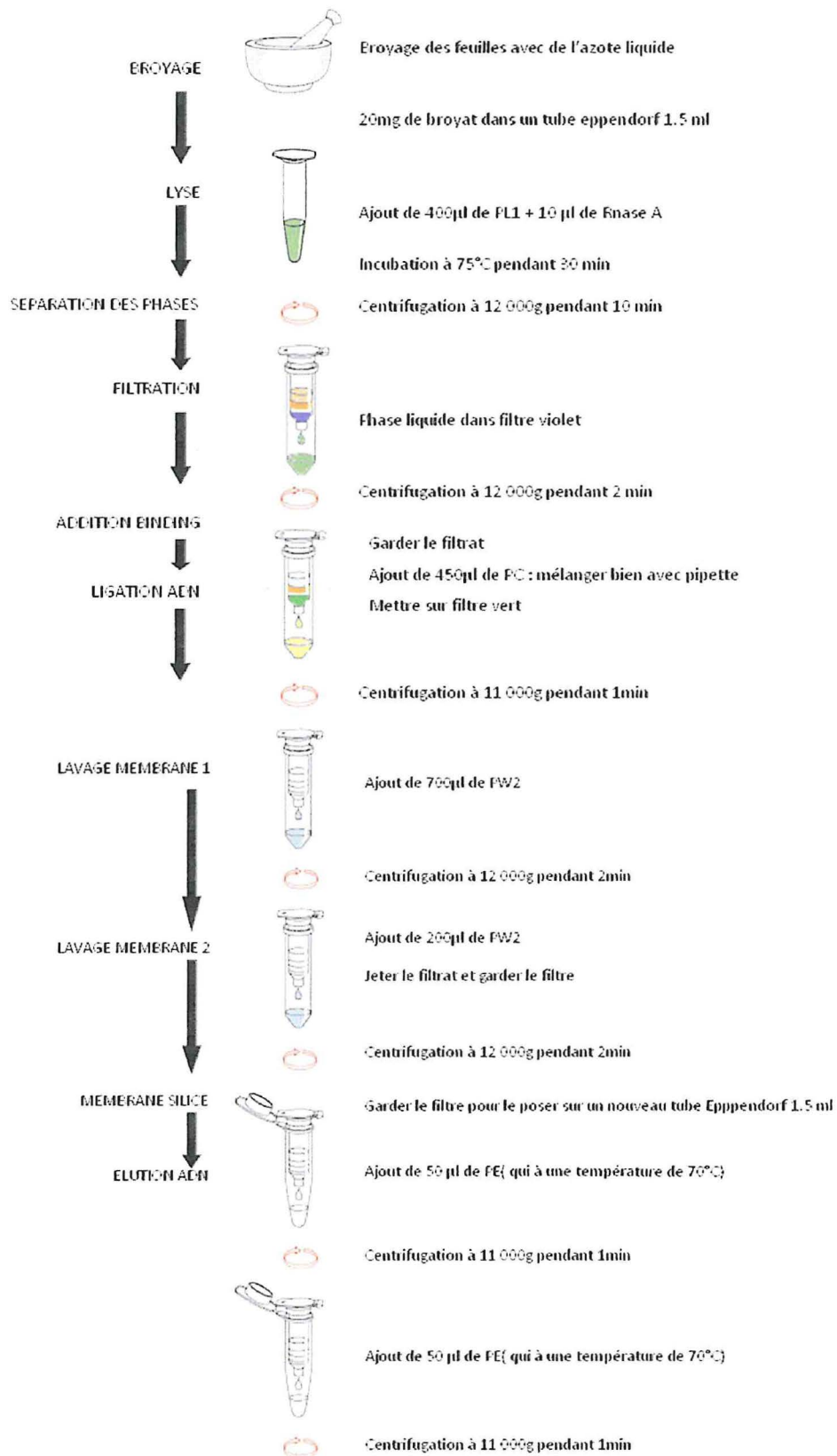
Annexe 1 : Protocole extraction ADN kit NucleoSpin

Annexe 2 : protocole séquençage LI-Cor[®]

Annexe 3 : Carte consensus utilisé pour l'échantillonnage

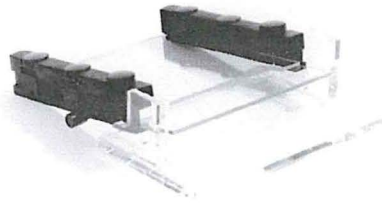
Annexe 4 : représentation graphique du DL (r^2) entre des paires de locus non liés .Origine Déli et La Mé

Annexe 5 : représentation graphique du DL (r^2) entre des paires de locus non liés. Origine Yangambi

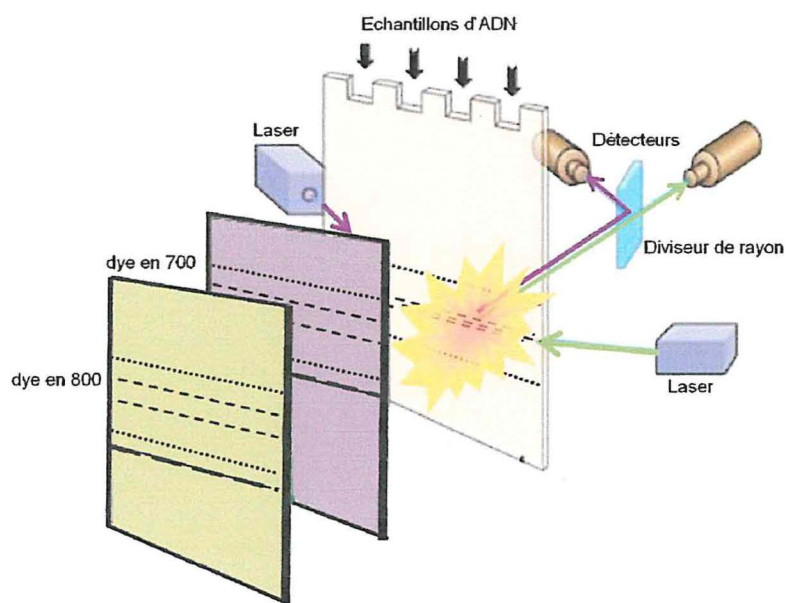


Annexe 1 : Protocole extraction ADN kit NucleoSpin

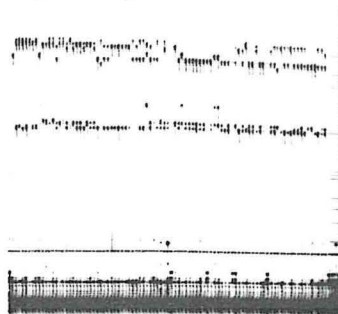
Etape 1 : préparation des plaques



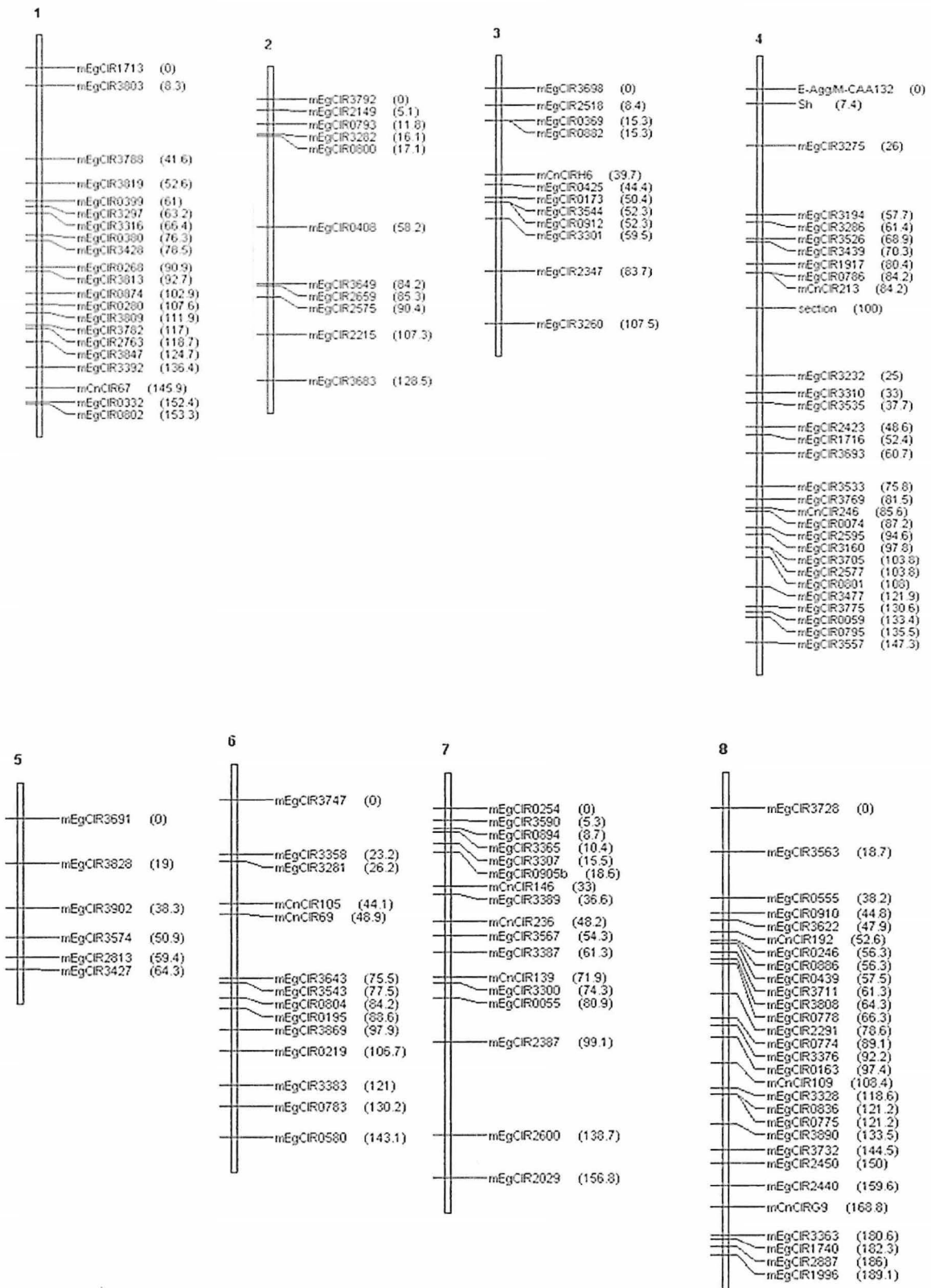
Etape 2 : Séquençage par électrophorèse



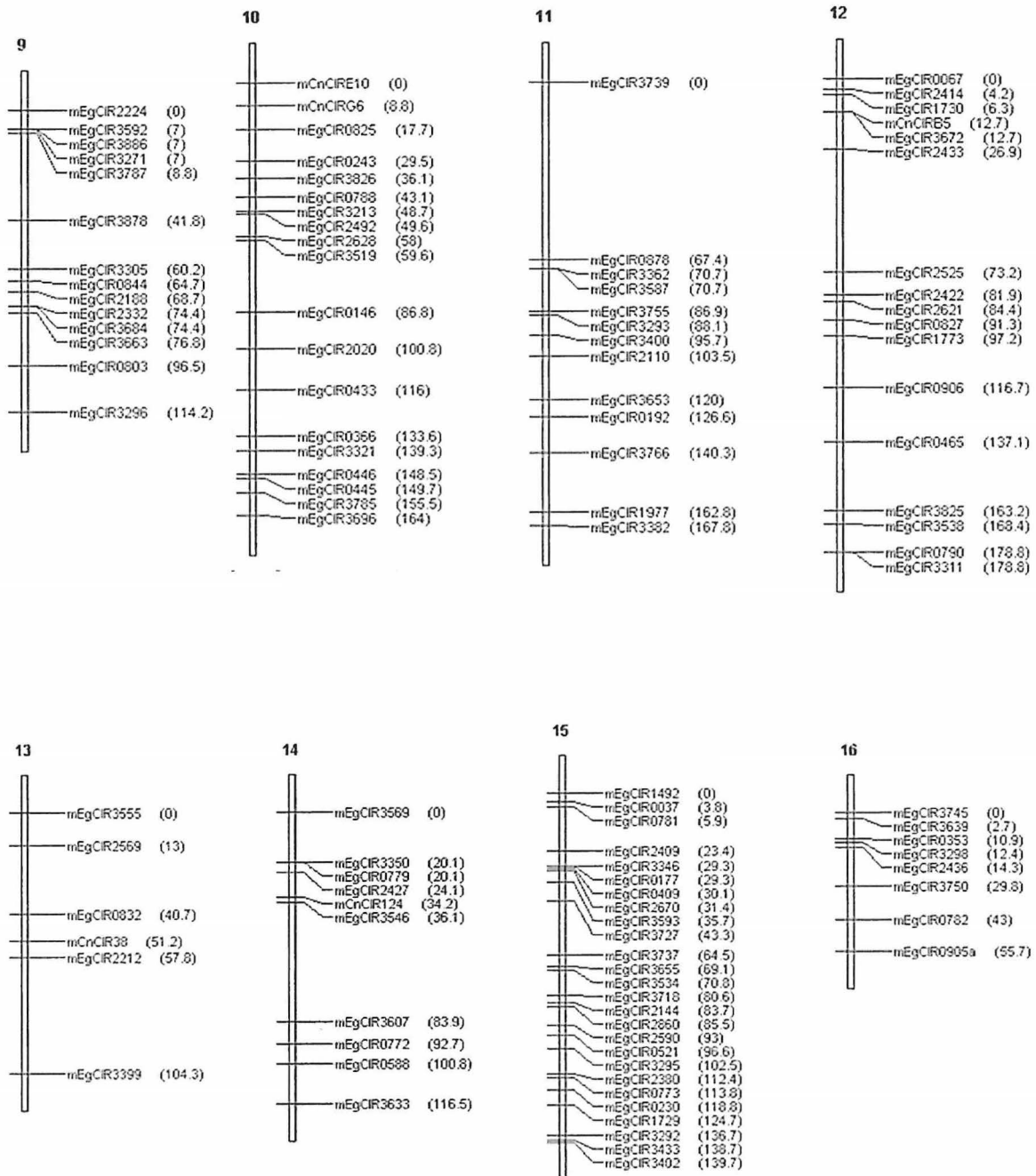
Etape 3 : Récupération de l'image du gel sur le serveur



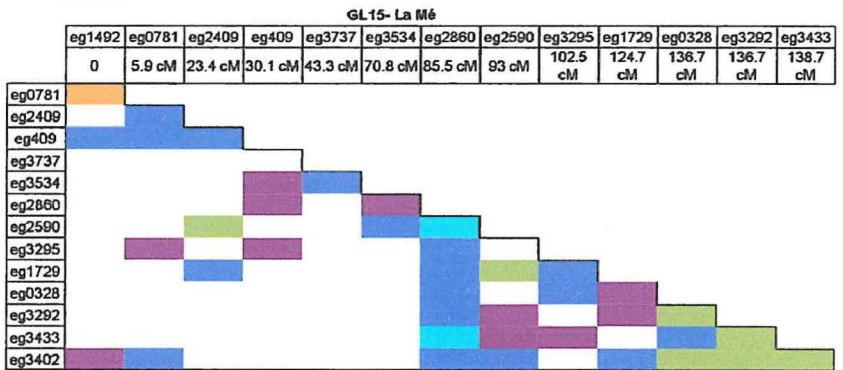
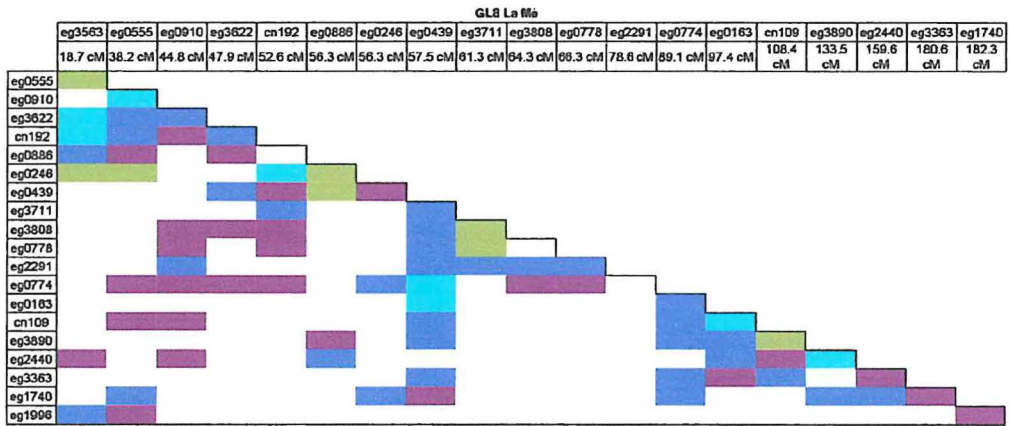
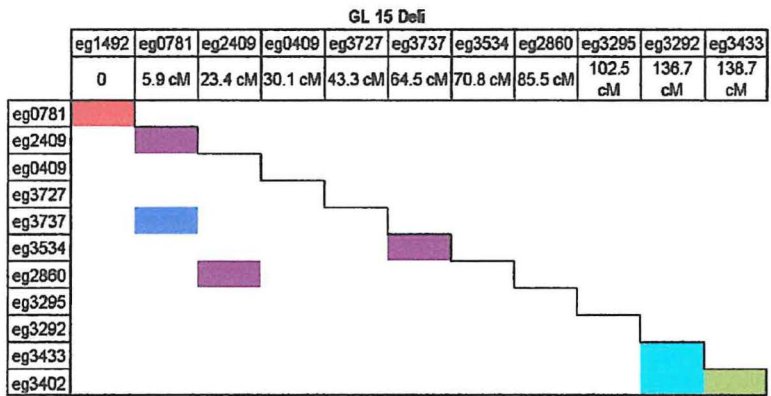
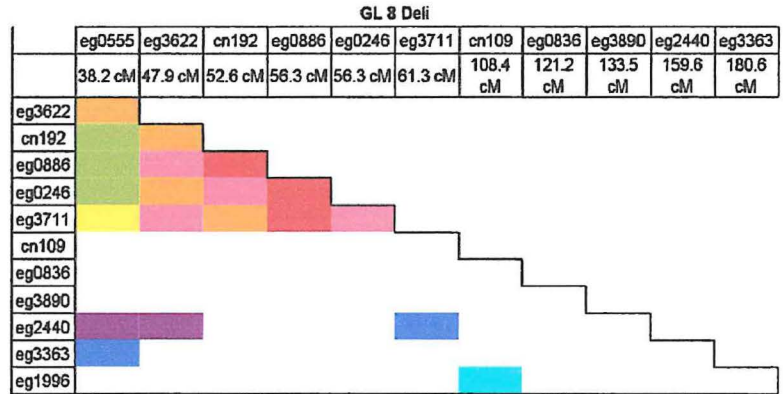
Annexe 2 : protocole séquençage LI-Cor[®]



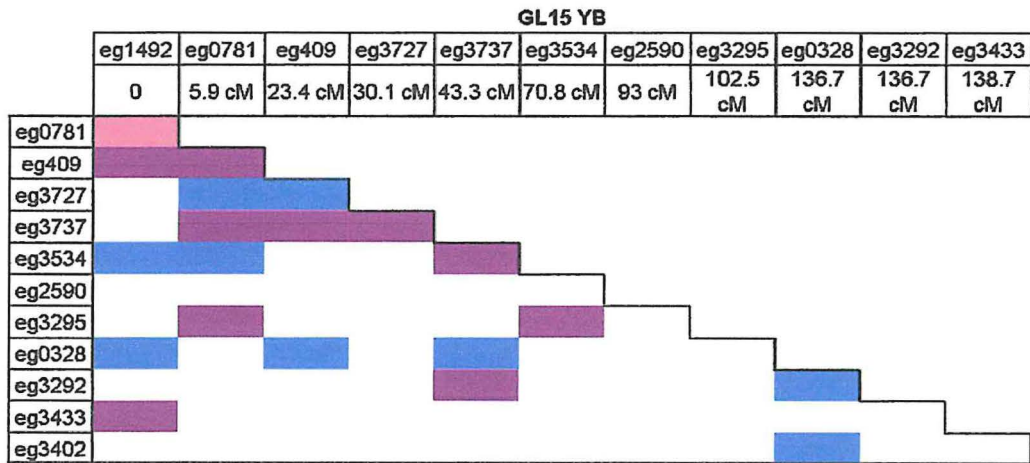
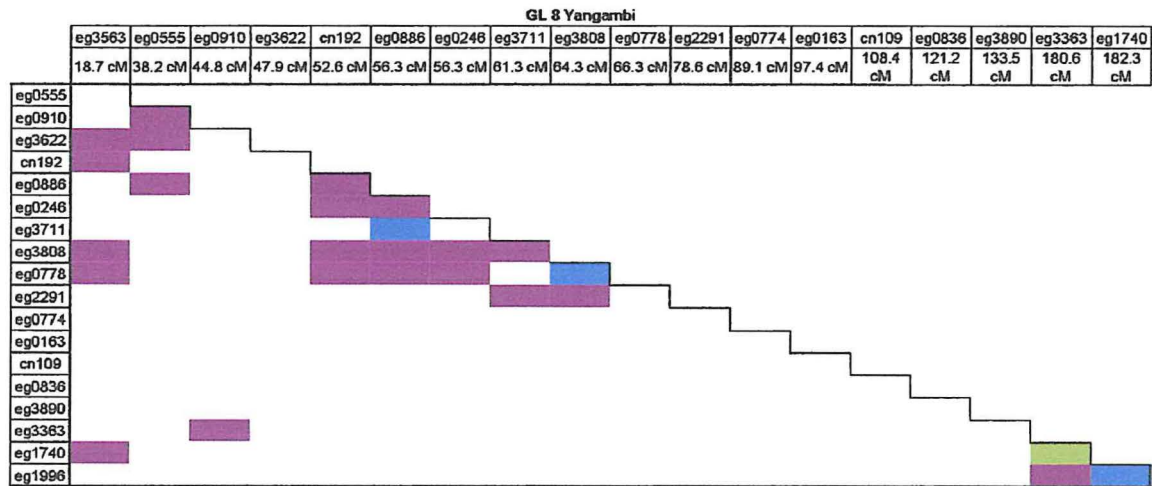
Annexe 3 : Carte consensus utilisé pour l'échantillonnage



Annexe 3 suite: Carte consensus utilisé pour l'échantillonnage



**Annexe 4 : représentation graphique du DL (r^2) entre des paires de locus non liés .
Origine Déli et La Mé**



Annexe 5 : représentation graphique du DL (r^2) entre des paires de locus non liés. Origine Yangambi

Bibliographies

Ardlie KG, Kruglyak L, Seielstad M. (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics*. Vol 3 : 299- 309

Billotte N., Marseillac N et al. (2005). Microsatellite-based high density linkage map in oil palm. *Theor Appl Genet*. 110: 754–765

Bakker P, Yelensky R and al. (2005), Efficiency and power in genetic association studies, *NATURE GENETICS*, Vol. 37 No 11

Beirnaert A., Vanderweyen R. (1941). *Contributio à l'étude génétique et biométrique des variétés d'Elaeis guineensis Jacq.*. Publication de l'Institut National pour l'étude Agronomique du Congo Belge. Série Scientifique n° 27

Billotte N, Marseillac N, Brottier P, Noyer JL, Jacquemoud-Collet JP, Moreau C, Couvreur T, Chevalier MH, Pintaud JC et Risterucci AM. (2004). Nuclear microsatellite markers for the date palm (*Phoenix dactylifera* L.): characterization and utility across the genus *Phoenix* and in other palm genera. *Molecular Ecology Notes* 4, 256 –258

Castilho A, Vershinnin A, Heslop-Harrison J.S (2000), Repetitive DNA and the Chromosomes in the Genome of Oil Palm (*Elaeis guineensis*), *Annals of Botany*, 85: 837±844

Castilho A, Vershinnin A, Heslop-Harrison J.S (2000), Repetitive DNA and the Chromosomes in the Genome of Oil Palm (*Elaeis guineensis*), *Annals of Botany*, 85: 837±844

Ching A, Caldwell K and al. (2002), SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines, *BMC Genetics*, 03:19:00

Cochard B., Adon B., Rekima S., Billotte N., Desmier De Chenon R., Koutou A., Nouy B., Omoré A., Razak P., Glaszmann J.C., Noyer J.L.. (2009), Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding, *Tree genetics and genomes*, 5 (3) : 493-504.

Cochard B., Amblard P., Durand-Gasselin T (2005), Oil palm genetic improvement and sustainable development, *OCL. Oléagineux corps gras lipides*, *OCL. Oléagineux corps gras lipides*

Cordell H.J, Clayton D (2005), Genetic association studies, *lancet* , 366: 1121–31

Darvasi A., Weinreb A., Minke V., Weller JI et Soller M., (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetic* 134 : 943-951

Durand-Gasselin T., Cochard B (2005), Oil palm seed distribution, *OCL. Oléagineux corps gras lipides*, 12 (2) : 148-153

Durand-Gasselin T., Cochard B et al. (1999), Stratégie d'amélioration du palmier à huile, *Plantation, Recherche, Développement*, Sept-Oct 344-358

Evanno G, Regnaut S, Goudet J, (2005), Detecting the number of clusters of individuals using the software STRUCTURE : a simulation study, *Molecular Ecology*, 14,, 2611–2620

Excoffier L. and Slatkin M (1995), Maximum-Likelihood Estimation of Molecular Haplotype Frequencies in a Diploid Population, *Mol. Biol. Evol*, 12(5):921-927.

Ferwenda J, (1977), Oil palm, in Alvin, P.DE.t, Kozlowski T.t. (Editeur. Eds), *Ecophysiology of tropical crops*. Academies Publisher, New York, 351-382

Garris A, McCouch S and Kresovich S (2003), Population Structure and Its Effect on Haplotype Diversity and Linkage Disequilibrium Surrounding the xa5 Locus of Rice (*Oryza sativa* L.), *Genetics*, 165: 759–769

Gascon et Wuidart (1975), Study of the composition of *Elaeis guineensis* Jacq. oil. Possibilities for improvement. *AGRIS Record*.v. 30(10) p. 401-408

Gascon JP, Le Guen V., Nouy B., Asmady et Kanga F., (1988), Résultats d'essais de second cycle de sélection récurrence réciproque chez le palmier à huile *Elaeis guineensis* Jacq. *Oléagineux*, 43 (1), p 1-5

Ghesquière M,(1983). Contribution à l'étude de la variabilité génétique du palmier à huile (*Elaeis guineensis* Jacq.) : le polymorphisme enzymatique. Thèse. Université de Paris -Sud, pp 116

Goldstein D (2001), islands of linkage disequilibrium, *Nature* , 29(2):109-111.

Gupta P, Rustgi S and Kulwal P (2005), Linkage disequilibrium and association studies in higher plants: Present status and future prospects, *Plant Molecular Biology*, 57:461–485

Hartley C. W. S. (1988). The Oil palm. 3rd edition. Tropical agriculture Series. Ed. Longman. Essex. 761 p

Hayati · R. Wickneswari · I. Maizura · N. Rajanaidu, (1999) Genetic diversity of oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: implications for improvement and conservation of genetic resources. *Theor Appl Genet* 108:1274–1284

Jaquemard JC, Baudouin LO, (1987) Contribution à l'étude de la croissance du palmier à huile: présentation d'un modèle descriptif. *Oléagineux*. vol. 42, no10, pp. 343-351

Jacquemard JC, (1979), Contribution à l'étude du palmier à huile en Afrique occidentale française. Ed. Emile Larose. Paris, 98 p.

Jomphe. V, (2006) Comparaison de la puissance de test de déséquilibre de liaison dans les études génétiques , Thèse faculté des sciences et des génie: Université Laval, Quebec,

Jorde L.B. Linkage disequilibrium and the search for complex disease genes. *Genome research*, 10, 1435-14444

Jouannica. S, Argout. X, Lechauve. F (2005), Analysis of expressed sequence tags from oil palm (*Elaeis guineensis*), Federation of European Biochemical Societies. ICA4-CT-2001-10066.

Mackay I and Powell W, (2007) Methods for linkage disequilibrium mapping in crops, *TRENDS in Plant Science*, Vol.12 No.2

Meunier. J, Gascon JP,(1972). Le schéma général d'amélioration du palmier à huile à l'IRHO. *Oléagineux*, vol.27, pp: 1-12.

Mueller JC. Linkage disequilibrium for different scales. *Briefings in bioinformatics*. Vol 5. no4. 355–364.

Price. Z, Schulman. A-H and Mayes. S (2003), Development of new marker methods—an example from oil palm, *Plant Genetic Resources*, 1(2-3); 103–113

Pritchard JK, and Przeworski M (2000). Linkage Disequilibrium in Humans: Models and Data. *Human Genetics*. 69, 1-14

Rafalski A (2002), Applications of single nucleotide polymorphisms in crop genetics, *Current Opinion in Plant Biology*, 5:94–100

Rafalski A and Morgante M (2004), Corn and humans: recombination and linkage disequilibrium in two genomes of similar size, *TRENDS in Genetics*, Vol.20 No.2

Remington D, Thornsberry J and al. (2001), Structure of linkage disequilibrium and phenotypic associations in the maize genome, *PNAS*, vol. 98 no. 20 11479–11484

Renard J, L, Gascon JP, and Bachy A. (1972), Recherches sur la fusariose du palmier à huile. *Oléagineux* 27, 581-591

Rohde W, Sniady V, Herrán A, Construction and Exploitation of High-Density DNA Marker and Physical Maps in the Perennial Tropical Oil Crops Coconut and Oil Palm: from Biotechnology towards Marker-Assisted, EC INCO-DC project, ICA4-CT-2001-10066.

Rostoks N, Ramsay. L, MacKenzie. K (2006), Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties, *PNAS*, 18656–18661 vol. 103 no. 49

S. Mayes, P. L. Jack, R. H. V. Corley, and D. F. Marshall (1997); Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). *Genome* 40(1): 116–122

Seguin M, Attard A et al. (2006), Analyse et prédiction des patrons de déséquilibre de liaison dans les collections de ressources génétiques de plantes pérennes ou annuelles, autogames ou allogames, *Les Actes du BRG*, 6 ; 57-74

Sherry A. Flint-Garcia (2003), Structure of linkage disequilibrium in plants, *Plant Biol.* 54:357–74

William Y. S. Wang, Bryan J. and al. (2005), Genome wide association studies : theoretical and practical concerns, *Nature genetics*, Vol. 6 110-118

Yu J and Buckler S. (2006) Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*. 17:155–160

Site internet :

www.tecan.com

www.licor.com

<http://pedegree.free.fr>

<http://pritch.bsd.uchicago.edu/software>

<http://www.maizegenetics.net>

www.xlstat.com