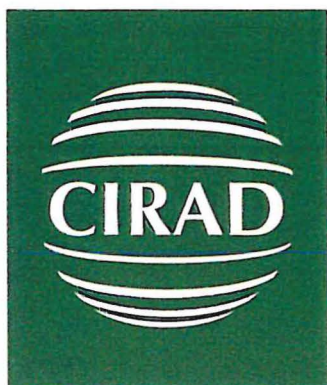


UNIVERSITE MONTPELLIER II SCIENCES ET TECHNIQUES DU LANGUEDOC
ECOLE NATIONALE SUPERIEURE AGRONOMIQUE - MONTPELLIER



Diplôme d'études approfondies (DEA)
Ressources Génétiques et Interactions biologiques
2005

CUBRY Philippe

Analyse de la diversité et évaluation du déséquilibre de
liaison chez quelques populations naturelles et cultivées
de caféiers *Coffea canephora*

Stage effectué dans le laboratoire Amélioration des plantes du Cirad campus de
Lavalette UMR 1096 PIA sous la direction de Magali Dufour

Soutenu le 1 juillet 2005.



Remerciements

Je tiens à remercier Magali Dufour pour m'avoir proposé ce sujet et Jean-Christophe Glazmann pour m'avoir permis de travailler au sein de son UMR.

Merci à Thierry Leroy de m'avoir accueilli dans son équipe et pour l'ensemble des discussions enrichissantes que nous avons pu avoir.

Merci à Hyacinthe Legnaté du CNRA de Côte d'Ivoire pour la fourniture de matériel végétal.

Merci à Pascal Musoli du CORI d'Ouganda pour le matériel végétal récolté dans les forêts ougandaises (projet INCO-DEV COWIDI contact ICA4-CT-2001-10006), et à Pauline Aluka du CORI d'Ouganda pour le matériel végétal cultivé en Ouganda (projet USDA-ICRAF, Agreement 58-4001-3-F157)

Merci à David Pot et Fabien de Bellis pour leur amitié et leurs conseils, à mes colocataires du bureau 111 qui m'ont permis de travailler dans une atmosphère très agréable.

Merci à Ronan Rivallan et Claire Billot, responsables de la plate-forme de génotypage de la génopole Montpellier-LR.

Merci à Xavier Perrier et Brigitte Courtois pour leurs lumières sur l'utilisation de certains logiciels.

Merci à l'ensemble des personnels du bâtiment 3 « amélioration des plantes » du CIRAD.

Et à tous ceux que j'oublie...

A Camille...

I.	REVUE BIBLIOGRAPHIQUE	1
I.1	INTRODUCTION.....	1
I.2	HISTORIQUE DE LA CULTURE ET DE L'AMELIORATION DES CAFEIERS	1
I.3	L'ESPECE COFFEA CANEPHORA.....	1
I.3.1	<i>Généralités.....</i>	<i>1</i>
I.3.2	<i>Diversité génétique.....</i>	<i>2</i>
I.3.3	<i>La cartographie génétique de Coffea canephora et les marqueurs microsatellites</i>	<i>2</i>
I.4	LE DESEQUILIBRE DE LIAISON (DL).....	3
I.4.1	<i>Définition</i>	<i>3</i>
I.4.2	<i>Expression du déséquilibre de liaison.....</i>	<i>3</i>
I.4.3	<i>Facteurs susceptibles d'affecter le déséquilibre de liaison.....</i>	<i>3</i>
I.4.4	<i>Vers les études d'association, une perspective de cartographie fine de QTL.</i>	<i>4</i>
II.	MATERIELS ET METHODES.....	5
II.1	MATERIEL VEGETAL	5
II.2	MARQUEURS MICROSATELLITES	6
II.3	METHODES	6
II.3.1	<i>Extraction d'ADN et dosage</i>	<i>6</i>
II.3.2	<i>PCR pour migration sur séquenceurs automatiques Li-Cor®.....</i>	<i>6</i>
II.3.3	<i>Migration sur séquenceur automatique Li-Cor® et lecture des gels.....</i>	<i>6</i>
II.4	ANALYSES DES DONNEES.....	7
II.4.1	<i>Analyses statistiques préliminaires</i>	<i>7</i>
II.4.2	<i>Diversité</i>	<i>7</i>
II.4.3	<i>Structure des populations.....</i>	<i>7</i>
II.4.4	<i>Etude du déséquilibre de liaison.....</i>	<i>8</i>
III.	RESULTATS	8
III.1	PURIFICATIONS ET EXTRACTIONS D'ADN.....	8
III.2	LECTURE DES GELS	9
III.3	DESCRIPTION DU JEU DE DONNEES	9
III.4	ANALYSE DE DIVERSITE A L'AIDE DU LOGICIEL DARWIN 5.0.....	9
III.5	CALCUL DES F_{ST}	10
III.6	RECONSTRUCTION DE POPULATIONS THEORIQUES A L'AIDE DE STRUCTURE	10
III.7	ETUDE DE DESEQUILIBRE DE LIAISON	11
III.7.1	<i>Test d'équilibre génotypique</i>	<i>11</i>
III.7.2	<i>Evolution du déséquilibre de liaison intragroupe de liaison avec la distance génétique.....</i>	<i>11</i>
IV.	DISCUSSION.....	12
IV.1	ETUDE DE DIVERSITE	12
IV.2	DESEQUILIBRE DE LIAISON	13
V.	CONCLUSIONS ET PERSPECTIVES	14

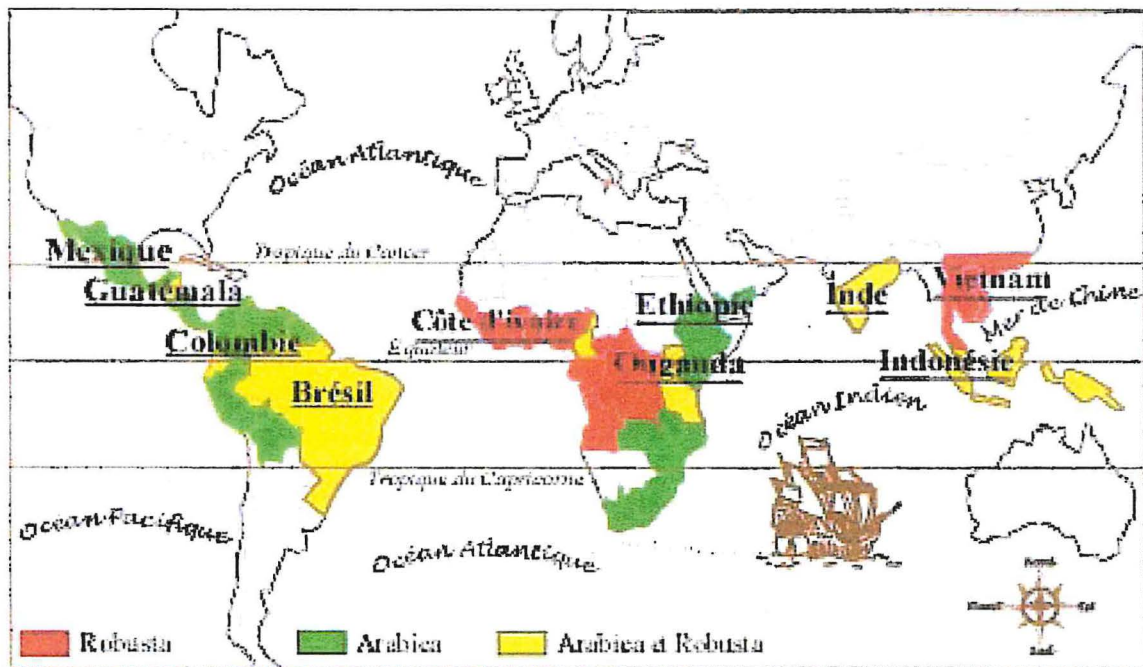


Figure 1 : Répartition de la production actuelle des deux principales espèces de caféiers cultivés, *Coffea arabica* (en vert) et *Coffea canephora* (en rouge). Les zones en jaune présentent les pays cultivant les deux espèces. Source : Merica.

I. Revue bibliographique

I.1 Introduction

Le café au sens large revêt une grande importance économique au niveau mondial, occupant la deuxième place en terme de valeur d'échange juste derrière le pétrole. Le café Robusta, produit par l'espèce *Coffea canephora* Pierre qui nous intéresse dans cette étude représente environ 30 à 35% de la production mondiale. Le café Arabica, produit par *Coffea arabica* représente le reste de cette production. Le genre *Coffea*, qui regroupe une centaine d'espèces, appartient à la famille botanique des Rubiacées qui compte environ 500 genres et plus de 6000 espèces. Le caféier est une plante pérenne, mise au champ plus de vingt ans.

I.2 Historique de la culture et de l'amélioration des caféiers

Le café n'est consommé en Europe que depuis le XVII^e siècle. A cette époque, seule l'espèce *C. arabica* était connue, principalement cultivée et commercialisée par les Arabes. Une expansion importante de cette culture s'est produite à partir de cette époque jusqu'au XX^e siècle. Celle-ci s'est réalisée sur la base d'un faible nombre de géniteurs et donc sur une base génétique très réduite, et s'est rapidement accompagnée d'attaques parasitaires très importantes.

Pour faire face à ces attaques et à la faible aptitude à la culture en basse altitude, une dizaine d'autres espèces ont fait l'objet de tentatives de mise en culture au XIX^e (Charrier et Eskes, 1998). Parmi elles on trouve *C. liberica*, *C. congensis*, *C. stenophylla* et *C. canephora*. Cette dernière s'est imposée par sa vigueur et ses résistances, notamment à la rouille orangée, véritable fléau pour la culture de *C. arabica*. Cette espèce est cultivée dans les zones intertropicales de basses altitudes, d'Afrique, d'Asie et d'Amérique, notamment au Brésil (**figure 1**).

Les caféiers sont tous diploïdes ($2n=2X=22$) et quasiment tous strictement allogames à l'exception notable de *Coffea arabica* qui est un allotétraploïde ($2n=4X=44$) autocompatible. Toutes les espèces diploïdes du genre dériveraient d'un même génome ancestral (Charrier et Eskes, 1998), *C. arabica* résulterait de l'association entre deux génomes diploïdes issus de deux espèces, dont *C. canephora*.

L'amélioration variétale du genre *Coffea* et en particulier de l'espèce *C. canephora* a été initiée sur la base d'hybridations intra et inter spécifiques. Cependant, l'obtention de nouvelles variétés clonales ou hybrides améliorées est rendue lente par la longueur du cycle de la plante, et le gain génétique par unité de temps reste faible, un cycle de sélection durant au minimum 16 ans (Montagnon, 2000).

I.3 L'espèce *Coffea canephora*

I.3.1 Généralités

Le genre *Coffea* est originaire du continent africain, de Madagascar et des îles Mascareignes. L'espèce *Coffea canephora*, dont les premiers plants spontanés furent découverts en 1860 près du lac Victoria, sur le territoire de l'Ouganda actuel, fut pour la première fois décrite en 1895.

Groupes génétiques identifiés							
Auteurs	Isoenzymes			RFLP	Nomenclature adoptée		Correspondance géographique
	Berthaud, 1986	Montagnon <i>et al.</i> , 1992	Dussert <i>et al.</i> , 1999	Dussert <i>et al.</i> , 1999			
	Guinéen	Guinéen	1	D	Pool guinéen	Guinéen	Guinée et Côte d'Ivoire
Congolais		SG1	3	A	Pool congolais	SG1	Façade atlantique de l'Afrique Centrale
		SG2	2	E		SG2	Bassin du fleuve Congo
				B		B	Sud Centrafrique, Nord République Démocratique du Congo
				C		C	Sud-ouest Centrafrique, Nord-Ouest République Démocratique du Congo, Est Cameroun

Tableau 1 : Groupes de diversité obtenus pour les différentes études basées sur des marqueurs isoenzymatiques ou RFLP. D'après Montagnon, 2000.

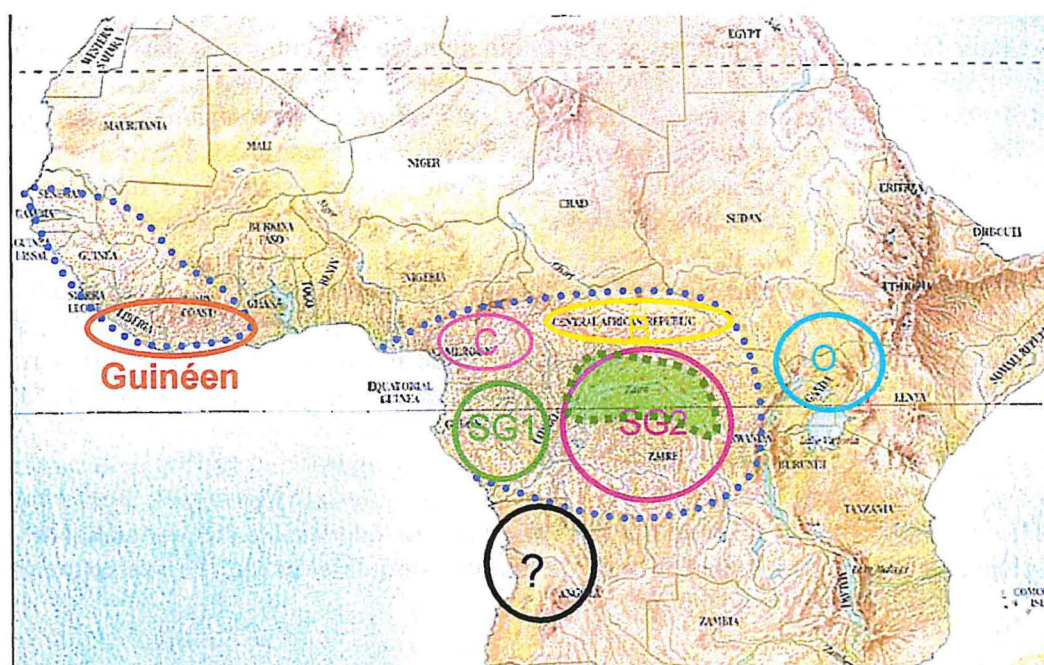


Figure 2 : Aire de répartition et groupes de diversité de *Coffea canephora*. Les populations Ougandaises (O) sont pour la première fois étudiées dans notre travail afin de définir leurs liaisons génétiques avec les groupes déjà identifiés. Les populations de caféiers sauvages supposées exister en Angola n'ont pour l'instant pas été prospectées. (modifié d'après Montagnon 2000). Les zones en bleu et vert sont respectivement les zones à saison sèche courte et sans saison sèche. Les forêts Ougandaises restent à caractériser pour ces critères mais elles semblent être relativement humides et d'altitude non négligeable (1500 m).

Cette espèce est diploïde ($2n=2X=22$) et strictement allogame par le fait d'un mécanisme d'autoincompatibilité gamétophytique (Berthaud, 1980). Du point de vue des ressources génétiques, différentes collections au champ sont présentes dans le monde, notamment en Côte d'Ivoire, en Ouganda, au Cameroun, à Madagascar ou au Brésil. La collection dont sont issues la plupart des populations que nous étudierons dans ce travail est celle de Côte d'Ivoire, regroupant des plantes provenant de différentes prospections de caféiers sauvages dans l'aire de répartition de l'espèce, ainsi que de variétés cultivées prospectées en plantation et des génotypes sélectionnés de divers centres de recherche africains. De plus des génotypes récemment collectés en Ouganda ont été incorporés à nos analyses.

I.3.2 Diversité génétique

La diversité génétique de *Coffea canephora* a été étudiée à l'aide de polymorphismes isoenzymatiques par Berthaud (1986) qui a décrit deux groupes principaux, les « guinéens » (G) composés principalement des populations sauvages de Côte d'Ivoire et de Guinée (Afrique de l'Ouest) et les « congolais » (C) comprenant les caféiers sauvages de République Centrafricaine, du Cameroun, du Congo (Afrique Centrale) et des génotypes cultivés issus d'autres centres de recherche. Plus récemment, Montagnon *et al* (1992) ont identifiés deux sous groupes (SG1 et SG2) au sein du groupe « congolais ». Les différents groupes alors identifiés possédaient des caractéristiques phénotypiques propres à chacun d'eux (**annexe 6**). Une étude menée sur 110 génotypes sauvages et cultivés pour 29 marqueurs isoenzymatiques a confirmé cette structuration (Dussert *et al*, 1999). Dans ce même travail, Dussert *et al* ont réalisé une étude complémentaire à l'aide de marqueurs RFLP. Deux nouveaux groupes ont ainsi pu être identifiés au sein du pool génétique congolais, les groupes B et C (**tableau 1 et figure 2**). Il est intéressant de noter que deux zones proches de celles étudiées n'ont pu faire l'objet de prospections de par le fait de difficultés d'accès. Nos individus issus de l'une de ces zones (Ouganda) constituent donc une nouveauté dans les études de diversité de l'espèce *Coffea canephora*.

La structuration entre guinéens et congolais a été la base des stratégies d'amélioration de l'espèce, en effet, Leroy *et al* (1994 et 1997) ont pu montrer que les meilleurs clones et descendances sélectionnés étaient en fait des hybrides entre les groupes guinéens et congolais. Un schéma de sélection récurrente réciproque (SRR) a été mis en place et a permis d'aboutir à l'identification de nouveaux hybrides intergroupes aussi homogènes et productifs que les clones vulgarisés (Montagnon, 2000).

I.3.3 La cartographie génétique de *Coffea canephora* et les marqueurs microsatellites

Les marqueurs utilisés dans notre étude sont des microsatellites ou « Simple Sequence Repeat » (SSR). Ces marqueurs sont composés de répétitions de motifs d'un faible nombre de bases (de 1 à 5) et dont le polymorphisme porte sur le nombre de ces répétitions. Ces marqueurs sont considérés comme neutres (Jarne et Lagoda, 1996), ont une hérédité mendélienne, sont codominants et présentent un polymorphisme important en règle générale. L'ensemble de ces caractéristiques en font donc des marqueurs de choix pour notre étude (nombreux allèles pour la description de la diversité génétique, détection d'un grand nombre de phénomènes démographiques et génétiques).

Une carte intraspécifique de *Coffea canephora* est en cours d'élaboration au CIRAD (Berline, 2002). Cette carte repose sur une descendance de type pseudo-backcross entre un hybride intergroupe 888 (Congolais SG1 x Guinéen) croisé avec un autre parent guinéen

Factor	Effect
Recombination rate	Higher recombination lowers LD
Mating systems: selfing species	High LD
Mating systems: outcrossing species	Low LD
Genetic isolation between lineages	Increases LD
Population subdivision	Increases LD
Population admixture	Increases LD
Natural and artificial selection	Locally increases LD
Population size	Small populations have more LD
Balancing selection	Increases LD
Mutation rate	High mutation rate decreases overall LD but LD around newly created mutated allele remains high until dissipated by recombination
Genomic rearrangements	Rearrangements suppress local recombination leading to LD increase in the vicinity
Stochastic effects (chance)	Increase or decrease in LD

Tableau 2 : Facteurs influençant le déséquilibre de liaison (DL). Rafalski et Morgante, 2004.

Species	Mating system	LD range ^a
Human	Outcrossing	
Nigerian		5 kb
European		80 kb
Cattle	Outcrossing ^b	10 cM
Drosophila	Outcrossing	< 1 kb
Maize	Outcrossing	
Diverse maize		1 kb
Diverse inbred lines		1.5 kb
Elite lines		> 100 kb
Arabidopsis	Selfing	250 kb
Sugarcane	Outcrossing/vegetative propagation	10 cM

^aThe LD value provided is estimated where $r^2 = 0.10$.

^bThere is an extreme sex bias in terms of number of breeding males versus breeding females.

Tableau 3 : Etendue moyenne du DL chez quelques espèces et populations. La valeur minimale retenue comme significative est $r^2 = 0,1$. Flint-Garcia *et al*, 2003.

02183 (**Annexe 2**). Cette descendance est donc très intéressante puisqu'elle est susceptible de capter des allèles issus des principaux groupes de diversité connus. Cette carte comporte actuellement 79 marqueurs répartis sur 17 groupes de liaison à LOD5 (15 à LOD3) et couvre une distance de 596 cM. Elle n'est donc pas saturée mais constitue une base de travail solide pour nos travaux d'évaluation du déséquilibre de liaison.

I.4 Le déséquilibre de liaison (DL)

I.4.1 Définition

Le déséquilibre de liaison, aussi appelé déséquilibre gamétique, se définit comme l'association non aléatoire de deux allèles à des loci polymorphes différents. Il se calcule comme un écart à l'indépendance statistique des allèles au niveau gamétique et peut s'exprimer de différentes manières. Le terme déséquilibre de liaison amène souvent une confusion avec la liaison physique des loci. En effet, un déséquilibre de liaison peut se retrouver même entre allèles non liés physiquement (c'est-à-dire non portés par le même chromosome ou espacés de plus de 50 cM) bien qu'une forte liaison physique (marqueurs proches sur le chromosome) se traduise en général par une forte valeur de DL.

I.4.2 Expression du déséquilibre de liaison

La première expression communément utilisée, D , a été proposée par Lewontin en 1964 (revue dans Jorde, 2000). Cet estimateur mesure la différence, pour une paire de locus bialléliques, entre la fréquence de co-occurrence observée de deux allèles et la fréquence attendue de cette co-occurrence en cas d'équilibre de liaison (les deux allèles ségrégent alors de manière indépendante). Ainsi, si l'on considère deux loci avec les allèles A et a au premier locus et B et b au second, on peut écrire :

$$D = G_{AB} - \pi_A \pi_B$$

Avec G_{AB} la fréquence observée de l'haplotype AB et $\pi_A \pi_B$ le produit des fréquences de A et B soit la fréquence attendue de cet haplotype sous l'hypothèse de l'équilibre de liaison entre les loci.

Cette mesure de D varie au plus de -1/4 à +1/4 selon la fréquence des allèles et est à la base de toutes les statistiques utilisées pour mesurer le DL, on considère qu'il existe un déséquilibre de liaison si D diffère significativement de 0.

I.4.3 Facteurs susceptibles d'affecter le déséquilibre de liaison

Etant donné que le DL est sensible aux fréquences alléliques et aux recombinaisons entre loci de par sa nature, la plupart des processus connus et observables en génétique des populations sont susceptibles de l'affecter (**tableau 2**). De nombreux mécanismes sont capables d'engendrer du DL (dérive génétique, mutations, structuration des populations, sélection naturelle). De plus certains facteurs, même s'ils ne créent pas du DL, sont susceptibles de l'influencer (balayage sélectif, consanguinité, régime de reproduction).

- La mutation crée le polymorphisme allélique et produit donc la matière de base que l'on retrouve à l'origine du DL (Flint-Garcia *et al*, 2003).
- La recombinaison est le principal facteur de diminution du DL intrachromosomique et va l'affecter de manière importante. Notamment la variation du taux de recombinaison le long du génome va pouvoir créer des blocs de forts DL séparés par des points chauds de recombinaison, c'est notamment le cas observable pour le maïs et certaines zones du génome humain (Rafalski et Morgante, 2004).

- La dérive génétique, dans les populations de faibles effectifs, va engendrer la perte constante d'allèles rares, faisant apparaître du déséquilibre.
- La sélection, qu'elle soit naturelle ou artificielle, aura tendance à réduire la diversité à un locus et à augmenter le DL dans la région adjacente (balayage sélectif ou autostop génétique).
- L'effet de l'histoire de la population ou de la structure des populations est potentiellement important. En effet les goulots d'étranglement ont un effet majeur sur le DL en l'augmentant de manière non négligeable (Rafalski et Morgante, 2004). L'ensemble de ces paramètres, ainsi que la sélection artificielle, seront particulièrement à considérer dans le cas de la comparaison entre nos populations qui présentent des histoires très différentes.
- Le mode de reproduction joue aussi un rôle important sur l'évolution du DL au cours du temps. Les nombreuses études menées sur *Arabidopsis*, le riz et le maïs ont en effet mis en évidence une différence importante de diminution du DL selon le régime de reproduction, les espèces allogames présentant une diminution plus rapide du DL comparativement aux autogames pour lesquelles la recombinaison est moins efficace.

I.4.4 Vers les études d'association, une perspective de cartographie fine de QTL.

Contrairement aux analyses de liaisons pour lesquelles les données familiales sont utilisées, les études de DL reposent des populations non issues de croisements contrôlés et s'affranchissent des données parentales en inférant les sources inconnues de création de DL (Flint-Garcia *et al*, 2003), permettant de s'intéresser directement aux populations naturelles. De nombreuses études ont ainsi été menées en génétique humaine afin de cartographier des QTL liés à des maladies (Jorde, 2000, Flint-Garcia *et al*, 2003). Plus récemment des études menées sur des plantes ont montré la viabilité d'une telle approche pour l'identification de QTL.

L'étude de DL permet d'évaluer la quantité de marqueurs nécessaires pour couvrir l'ensemble du génome en vue de réaliser de telles études d'association ainsi que d'élaborer des stratégies de cartographie se basant notamment sur le choix des populations à étudier. Notamment des populations à fort DL pourront être utilisées pour la recherche de QTL sur l'ensemble du génome grâce à un nombre relativement limité de marqueurs. Une cartographie plus fine de ces QTL sera possible par l'utilisation de populations à faible DL, la résolution de ce type d'approche étant supérieure à celle des études de cartographies basées sur des croisements contrôlés.

Les populations présentant un DL significatif à une distance importante pourront ainsi être incorporées aux schémas d'amélioration de l'espèce étant donné leur possibilité d'avoir conservé des blocs de forts DL pouvant contenir des gènes d'intérêt. De plus un autre avantage de cette approche basée sur des populations, et notamment sur des populations naturelles, est la possibilité de détecter un nombre d'allèles bien plus important que dans les études de cartographies classique.

En pratique ces études ont été menées chez de nombreuses espèces modèles telles que *Arabidopsis*, le riz, le maïs. C'est ainsi que le gène candidat *dwarf8* a pu être associé au temps de floraison chez le maïs (Thornsberry *et al*, 2001). D'autres études ont permis de cartographier précisément une variation importante du polymorphisme d'expression de *sugary1* comme étant liée à une seule mutation nucléotidique. Le DL a aussi été étudié chez la canne à sucre (Jannoo *et al*, 1999), chez *Arabidopsis* et chez le riz. Plus récemment une étude réalisée chez une espèce pérenne, le cacaoyer, a permis de mettre en évidence un DL important et persistant à plus de 10 cM et de mener une étude d'association dont la plupart des QTL détectés collocalisaient avec des QTL préalablement identifiés par des études de cartographies classiques (Pugh-Moreno, T., Thèse 2005).

Ces études d'associations permettant de relier des marqueurs moléculaires avec des caractères quantitatifs donnent la possibilité d'améliorer et d'accélérer les schémas de sélection. Ceci est particulièrement vrai pour les espèces pérennes avec l'évaluation précoce de la valeur d'un individu.

II. Matériels et méthodes

II.1 Matériel végétal

Les individus que nous avons retenus pour notre étude couvrent l'ensemble de la diversité de *C. canephora* décrite dans la littérature. Nous y avons ajouté des individus issus de prospections réalisées dans des forêts d'Ouganda qui n'ont pas encore été étudiés. Les populations naturelles de l'espèce *C. canephora* sont généralement isolées géographiquement et elles sont constituées d'un nombre d'individus assez faible (souvent inférieur à 100 individus) dont un nombre limité (de 2 à 30) participe à la reproduction. Nous avons choisi de limiter notre étude à une population par groupe de diversité à l'exception des guinéens pour lesquels nous avons deux populations. Ce choix s'est basé sur des difficultés d'obtention du matériel de Côte d'Ivoire et sur la limitation du nombre d'individus à génotyper.

A l'issue du choix des populations, nous avons un panel composé de 210 individus répartis de la manière suivante :

Parents de cartographie (02183 et 888) : le parent 02183 est issu d'une population guinéenne (Logbonou), le parent 888 est un hybride intergroupe (Congolais SG1, population A1 améliorée x Guinéen, population Kouilou prospecté en plantation) (**voir annexe 2**).

Clone 126 : 1 génotype hybride issus de prospection en plantation, l'un des meilleurs clones cultivé, ayant notamment servi à la confection d'une banque enrichie en microsatellites ainsi qu'à celle d'une banque BAC (Leroy *et al*, 2005, accepté TAG).

Congolais SG1 : 9 individus Niaouli, variété cultivée notamment au Bénin et au Togo, vraisemblablement originaire du Gabon.

Congolais SG2 : 25 individus Ineac, les individus de notre échantillon appartiennent à deux populations vraisemblablement composées de demi-frères, chacune de ces populations étant composée d'individus issus de graines prélevées sur un seul arbre mère (arbres n°2 et 7 du champ semencier de Duekoue, les deux arbres mères étant des génotypes sélectionnés dans un centre de recherche du Congo belge introduits en Côte d'Ivoire en 1935).

Congolais B : 39 individus (Libengue), population sauvage prospectée en République Centrafricaine.

Congolais C : 10 individus (Nana), variété cultivée à l'issue d'une mise en culture de populations naturelles locales, elle présente comme principale caractéristique de posséder une mutation récessive conférant un phénotype particulier (nanisme) à ses représentants.

Guinéens : 37 individus (Pélézi) et 32 individus (Mouniandougou), ces populations sont des populations sauvages de Côte d'Ivoire.

Ougandais : notre échantillon est composé de neuf témoins de diversité, 3 issus de plantations (Farm), trois génotypes identifiés comme la forme N'Ganda et trois identifiés comme la forme Erect, ces deux formes ayant été décrites dès 1935 par Thomas qui désigne la forme N'ganda comme spontanée et la forme Erect comme introduite par des cultivateurs européens sous forme de graines en provenance de Java au début du XX^e siècle.

46 individus sauvages issus de prospections récentes dans deux forêts (Kibale et Itwara, **figure 3**), avec quatre sites par provenance. Nous avons choisis de représenter chaque site de prospection par deux à quatre individus à l'exception de la population Ngogo Kibuguta pour laquelle nous avons 20 individus.

II.2 Marqueurs microsatellites

Les marqueurs ont été choisis à l'aide de la carte génétique disponible au CIRAD (**figure 4**). Trente-quatre marqueurs ont ainsi été retenus pour l'étude de diversité. Parmi eux vingt-cinq liés génétiquement pour des distances allant de 2 à 38 centimorgans ont été utilisés pour l'étude de déséquilibre de liaison (**tableau 4**), et quinze non liés sur la carte pour l'utilisation du logiciel *Structure* (Pritchard *et al*, 2001).

II.3 Méthodes

II.3.1 Extraction d'ADN et dosage

L'ADN de feuilles broyées a été extrait à l'aide d'un protocole donné en **annexe 1**. Ces ADN ainsi qu'un certain nombre extraits précédemment (entre 9 et 12 mois auparavant) montrant des résultats mitigés en PCR ont été purifiés à l'aide d'un kit de purification d'ADN génomique de la firme Promega (Wizard® Genomic DNA Purification Kit Cat.# A1125) reposant sur l'utilisation de deux solutions (Nuclei Lysis Solution et Protein Precipitation Solution).

Le dosage des extraits est réalisé sur gel d'agarose à 0,8%, les ADN sont ensuite dilués à 0,5 ng/μl et placés dans des racks. La qualité des ADN et les dilutions ont été testées par PCR (couples d'amorces des microsatellites 327 et DL010) et migration sur gel d'agarose à 0,8%.

II.3.2 PCR pour migration sur séquenceurs automatiques Li-Cor®

Les PCR réalisées ont été faites en simplex, c'est-à-dire qu'un seul couple d'amorce est utilisé par réaction de polymérisation. La particularité de nos PCR est qu'elles font appel à des amorces particulières. En effet une des deux amorces du couple utilisé porte une queue M13 sur laquelle viendra s'hybrider une amorce M13 portant un fluorochrome (700 ou 800). Cette stratégie d'amplification sert à rendre visualisable les produits PCR par migration sur séquenceur automatique Li-Cor® et nécessite une mise au point par rapport aux PCR « classiques », en particulier la concentration de l'amorce portant la queue M13 est inférieure à celle de la seconde amorce du couple (**voir annexes 2 et 4**).

II.3.3 Migration sur séquenceur automatique Li-Cor® et lecture des gels

La migration s'effectue sur des gels d'acrylamide à 6,5% dont le pouvoir séparateur est de l'ordre de la paire de base. Le séquenceur détecte le passage des produits PCR portant le fluorochrome à l'aide d'un laser. La détection se fait sur deux longueurs d'onde (700 et 800 nm), ce qui permet de multiplexer les PCR lors de la phase de migration et de pouvoir faire

migrer jusqu'à six marqueurs simultanément. Les données collectées par le séquenceur sont directement visibles sur un écran d'ordinateur et sont stockées sous la forme d'un fichier .TIF qui est ensuite importé sous SAGA^{GT} (LiCor Biosciences) afin d'être exploité. Ce logiciel permet de définir les différents allèles pour chaque marqueur et d'assigner un couple d'allèles à chaque individu. Cette approche permet par ailleurs d'améliorer la qualité de lecture des gels multiples en indiquant les tailles réelles des fragments reconnus. A l'aide de témoins présents sur les différents gels nous pouvons donc vérifier le calibrage des allèles ce qui permet de réduire considérablement le risque d'erreur de génotypage par rapport à une lecture manuelle des gels.

II.4 Analyses des données

II.4.1 Analyses statistiques préliminaires

Pour chaque locus, nous avons estimé le nombre d'allèles, les fréquences alléliques et génotypiques, l'hétérozygotie attendue (gene diversity : $\hat{D}_l = (1 - \sum_{u=1}^k p_{lu}^2)$ avec l le locus considéré, p_u la fréquence de l'allèle u) et l'hétérozygotie observée, ainsi que le « Polymorphism Information Content », mesure proche de l'hétérozygotie attendue, se calculant comme $PIC_l = 1 - \sum_{u=1}^k \tilde{p}_{lu}^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^k 2\tilde{p}_{lu} \tilde{p}_{lv}$ (Bolstein *et al*, 1980), cela pour l'échantillon global et les populations supposées, à l'aide du logiciel PowerMarker (Liu, 2004, v3.23).

II.4.2 Diversité

A partir des données génotypiques obtenues pour les marqueurs microsatellites, un tableau de dissimilarités est calculé à l'aide du logiciel DARwin 5.0 (Perrier, 2005) selon la formule :

$$D_{ij} = 1 - \frac{1}{p} \sum \frac{m}{\pi}$$

avec m le nombre d'allèles du marqueur k présents à la fois chez i et chez j et π la ploïdie. Ce tableau de dissimilarités est ensuite soumis à une Analyse Factorielle sur Tableau de Dissimilarités (AFTD) afin de regrouper les individus selon leurs ressemblances. Le pourcentage d'explication des données de chaque axe retenu pour la sortie graphique est indiqué. Cette méthode permet de décrire de manière visuelle la structuration de notre échantillon.

Un arbre génétique basé sur la matrice de dissimilarités obtenue lors de l'étape précédente est ensuite construit à l'aide de la méthode du Neighbour-Joining (NJ) afin de représenter les relations de parentés entre les individus et les groupes.

II.4.3 Structure des populations

Pritchard *et al* (2001) ont montré l'importance de la prise en compte de la structure des populations dans la réduction de la détection de faux positifs lors du calcul du déséquilibre de liaison. Nous avons l'avantage d'avoir des individus issus de populations identifiées et nous

nous intéresserons donc pour l'étude du DL principalement à ces populations. Cependant nous validerons cette structuration par les sorties AFTD de DARwin et des calculs de F_{st} dont la significativité sera testé par bootstrap sur 10000 permutations. Une analyse grâce au logiciel *Structure* (Pritchard, 2001), sur la base de marqueurs non liés, qui affecte chaque individu à une population théorique à l'aide d'un algorithme d'agglomération Bayésienne sera réalisée. Nous reconstruirons ainsi des populations théoriques que nous comparerons aux populations identifiées.

II.4.4 Etude du déséquilibre de liaison

Le DL sera calculé à l'intérieur des populations possédant un effectif supérieur à 15 individus.

La significativité d'un déséquilibre sera testé par un test d'équilibre génotypique entre marqueurs, en effet nos données ne nous permettent pas de nous intéresser directement à un déséquilibre gamétique entre allèles liés physiquement du fait de l'absence d'information sur la phase des doubles hétérozygotes. Ce test, implémenté sous GDA (Lewis et Zaykin, 2001), reconstruit des tableaux de contingence génotypique et teste l'hypothèse d'indépendance des génotypes à l'aide d'un test exact de Fisher de 10000 permutations. Ce test est particulièrement intéressant dans nos données puisqu'il ne nécessite pas que les marqueurs étudiés soient à l'équilibre de Hardy-Weinberg (EHW) et peut donc être utilisé avec l'ensemble des marqueurs polymorphes.

Plusieurs méthodes de calcul du DL existent et sont pour la plupart adaptées à des données haplotypiques ou génotypique à phase connue. Nous nous intéresserons dans notre travail à une mesure génotypique du DL. Nous grouperons pour cela les allèles rares en une seule classe afin de retrouver un cas simple de calcul de DL entre loci bialléliques. Afin de pouvoir comparer entre population le déclin du DL en fonction de la distance, la mesure composite (Delta) du DL proposée par Weir (1996) implémentée sous GDA sera calculée :

$$\Delta_{AB} = p_{AB} + p_{A/B} - 2p_A p_B = D_{AB} + D_{A/B}$$

avec p_{AB} fréquence des allèles A et B en couplage (liés physiquement) ; $p_{A/B}$ fréquence des allèles A et B en répulsion

Cette mesure ne nécessite pas de faire l'hypothèse d'EHW et permet donc de travailler sur l'ensemble des loci. Elle est par ailleurs la seule qui soit directement adaptée aux données génotypiques.

III. Résultats

III.1 Purifications et extractions d'ADN

Un total de 287 ADN préalablement extraits (entre 9 et 12 mois auparavant) ont été purifiés et dosés sur gels d'agarose, pour des concentrations finales évaluées de 0 à 400 ng/ μ l. En plus de ces individus, 135 nouvelles extractions suivies d'une phase de purification et de dosage ont été réalisées. Les ADN ont été dilués à une concentration estimée de 0,5 ng/ μ l et mis en rack. Leur qualité a été testée par une PCR avec le couple d'amorce du microsatellite 327. Celui-ci a permis d'amplifier l'ensemble des individus à l'exception de ceux issus du

Plan d'AFTD N° 2 Axes : 1 / 2 Population globale

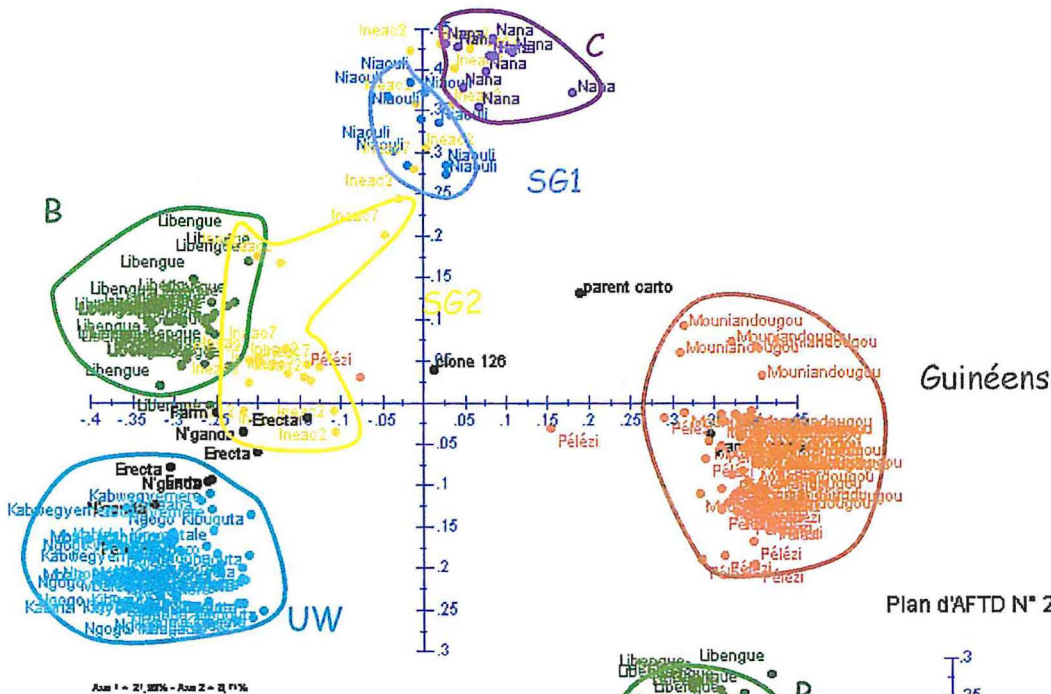


Figure 5 : AFTD sur la population totale, axes 1 et 2

Plan d'AFTD N° 2 Axes : 1 / 3 Population globale

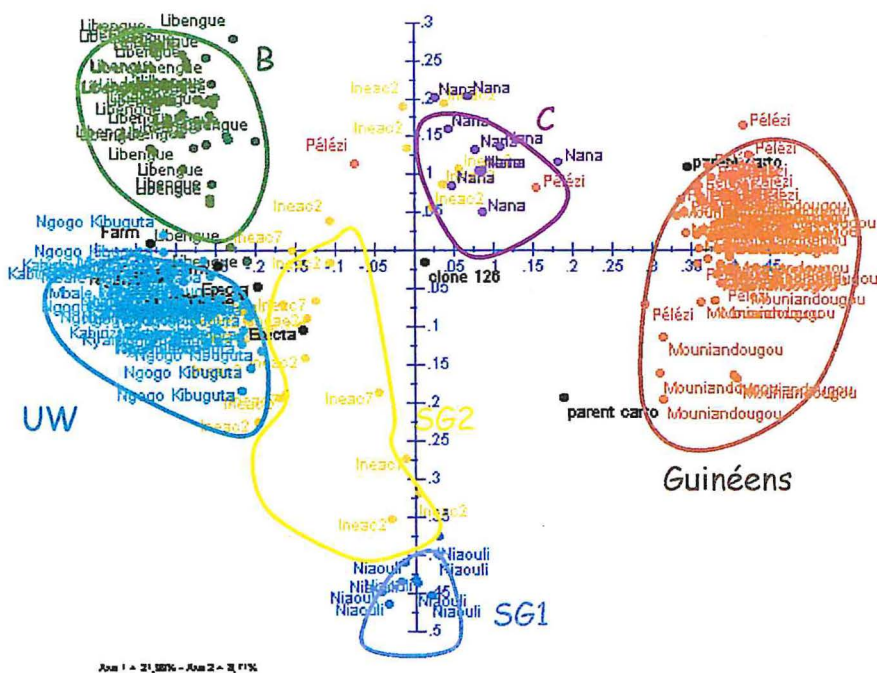


Figure 6 : AFTD sur la population totale, axes 1 et 3

Plan d'AFTD N° 1 Axes : 1 / 2 Guinéens

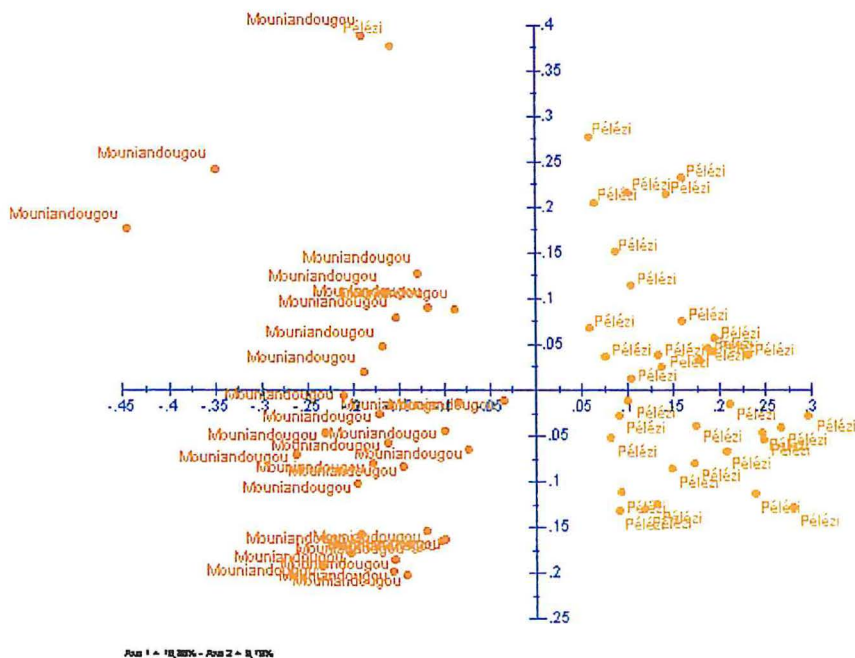


Figure 7 : AFTD sur les 2 populations guinéennes, axes 1 et 2

groupe C. Une nouvelle PCR a été réalisée à l'aide du couple d'amorces de DL010 ayant pour résultat une amplification sur l'ensemble des individus. Le microsatellite 327 présente donc un allèle nul chez les individus de la population Nana.

III.2 Lecture des gels

La lecture s'est faite à l'aide du logiciel SAGA^{GT}, ce qui nous a permis d'exporter une matrice de résultat brut que nous avons ensuite mise en forme pour l'analyse des données par divers logiciels à l'aide du tableur Microsoft Excel®.

III.3 Description du jeu de données

L'ensemble des statistiques descriptives a été estimé pour la population globale ainsi que pour les groupes et populations définies a priori, pour chacun des trente-quatre marqueurs génotypés. Notre échantillon présente un nombre moyen de données manquantes 10%, ce qui est relativement faible. Ce nombre de données manquantes varie de manière importante selon les marqueurs et selon les populations.

Le nombre moyen d'allèles par marqueur est de 12,5 démontrant une importante diversité au sein de notre panel, d'autant que les indices « classiques » de diversité que sont l'hétérozygotie attendue et le PIC possèdent des valeurs importantes, respectivement 0,73 et 0,71 en moyenne. Ces résultats sont supérieurs à ceux obtenus par Poncet *et al*, 2004, pour *C. canephora* et *C. pseudozanguebariae*.

III.4 Analyse de diversité à l'aide du logiciel DARwin 5.0

L'analyse à l'aide de DARwin 5.0 a permis de reconstruire à partir de nos données une matrice des dissimilarités (assimilable à une matrice des distances génétiques) sur laquelle plusieurs analyses factorielles ont été menées.

Une première analyse incorpore l'ensemble des 210 individus pour les trente-quatre marqueurs génotypés, les deux premiers axes de l'AFTD expliquent 21,99% et 8,71% de la variabilité (**figure 5**). Cette analyse montre une nette différenciation entre les grands groupes précédemment décrits (congolais versus guinéens) ainsi que la particularité des caféiers prospectés en Ouganda. On retrouve par ailleurs nettement les groupes B, SG1 et C des congolais. En revanche, les individus classés a priori dans le groupe SG2 se séparent en deux groupes, l'un homogène et bien différencié des autres sur la représentation, le second s'intégrant au groupe C. Comme attendu, le parent guinéen 02183 se place parmi les guinéens alors que l'hybride intergroupe 888 (congolais SG1 x guinéen) se place entre ces deux groupes. Le clone 126 est un hybride particulier qui se retrouve au centre de notre représentation. Afin d'affiner notre analyse, nous avons utilisé les axes 1 et 3 pour une nouvelle représentation. L'axe 3 explique 5,75% de la variabilité et permet de séparer de manière nette les groupes SG1 et C qui étaient plus proches sur la représentation précédente (**figure 6**). Les deux populations guinéennes que nous avons retenues semblent se mélanger dans cette analyse. Afin d'affiner l'analyse intergroupes, deux autres AFTD ont été réalisées, l'une pour les deux populations guinéennes, la seconde sur l'ensemble des populations originaire d'Afrique Centrale et d'Afrique de l'Est.

L'analyse portant sur les guinéens permet de séparer nettement les deux populations (**figure 7**), bien que certains individus se comporte de manière particulière. Les deux premiers

Plan d'AFTD N° 1 Axes : 1 / 2 Congolais plus Ougandais

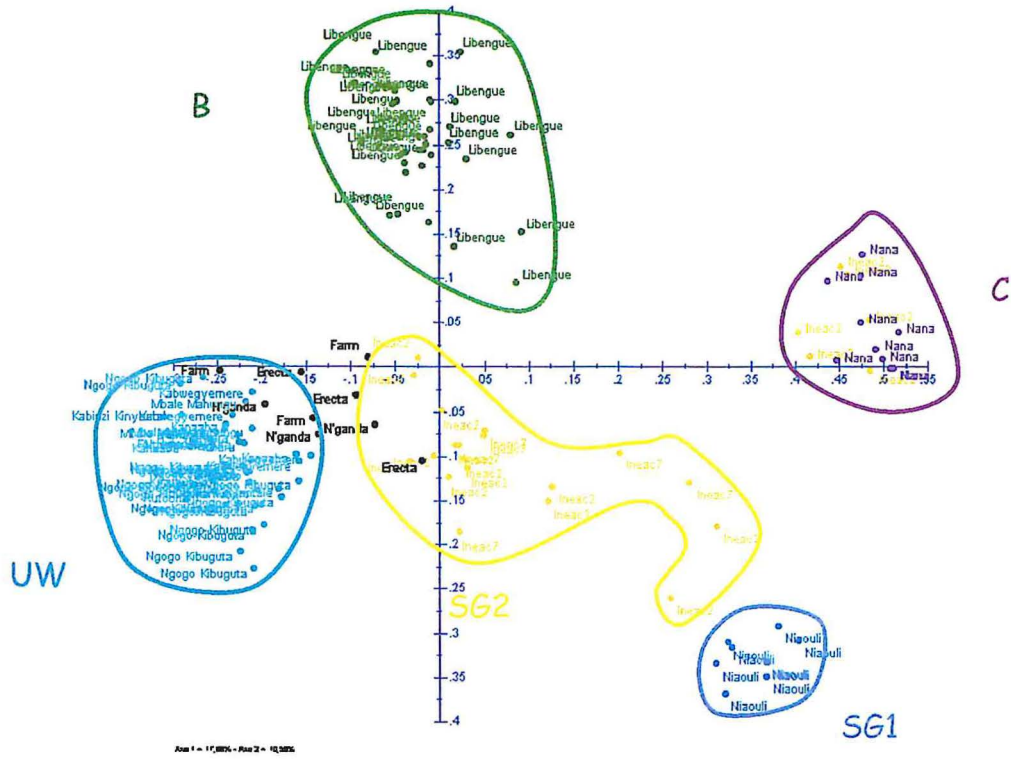


Figure 8 : AFTD sur les populations congolaises et ougandaises, axes 1 et 2

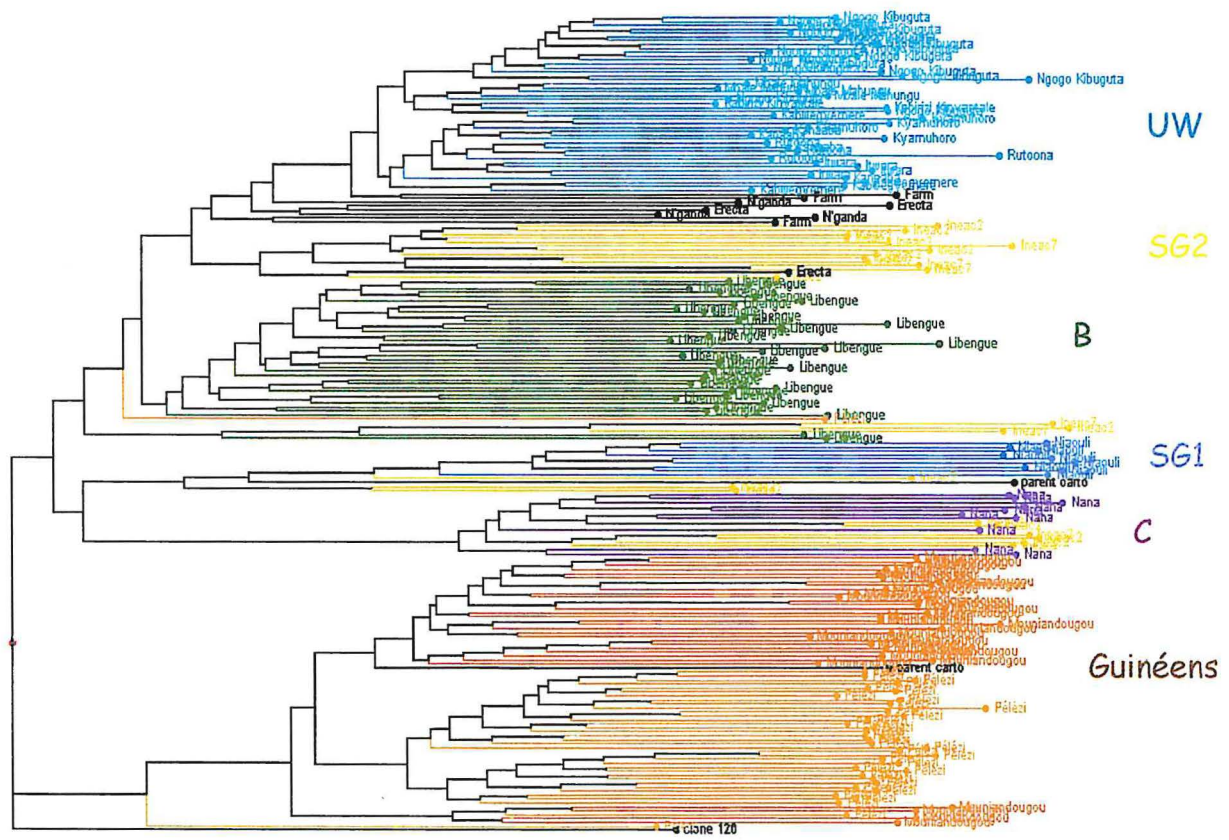


Figure 9 : arbre NJ sur la population totale

axes, utilisés pour cette représentation, expliquent 16,88% et 9,73% de la variabilité. La prise en compte de l'axe 3 (7,28% d'inertie) n'apporte pas d'information supplémentaire.

L'analyse Congolais-Ougandais permet d'affiner la perception des sous-groupes Congolais précédemment décrits (B, C et SG1) (**figure 8**). Un autre résultat de cette analyse est la particularité des populations d'Ouganda qui constituent un nouveau groupe de diversité. Globalement ce nouveau groupe se retrouve assez proche des groupes B et SG2, les groupes C et SG1 se retrouvant logiquement proche génétiquement ce qui concorde de manière intéressante avec la répartition géographique de ces différents groupes. Les deux premiers axes principaux expliquent 17,96% et 10,56% de la variabilité et la prise en compte du troisième axe n'apporte pas d'information supplémentaire.

L'arbre génétique construit nous permet d'évaluer les relations de parenté entre les différents groupes ainsi que leur éloignement (**figure 9**). En complément des résultats obtenus par Dussert *et al* (1999) à l'aide de marqueurs RFLP, nous pouvons observer que les guinéens sont plus proches des groupes C et SG1. Les groupes SG2, B et Ougandais sont plus éloignés. Si nous nous intéressons en parallèle à la répartition géographique de ces groupes nous pouvons observer que ce gradient de différenciation génétique concorde avec la localisation géographique de nos groupes.

III.5 Calcul des F_{st}

Le F_{st} global, calculé entre populations, a une valeur significative de 0,42. Cette valeur élevée reflète une structuration très importante de notre échantillon, ce qui confirme les résultats précédents. Le F_{st} a été estimé de la même manière pour le groupe guinéen à 0,23 et pour les congolais à 0,32. La valeur plus faible des guinéens montre qu'il s'agit d'un groupe particulier plus homogène que la population totale ou que les congolais. Ce résultat confirme encore une fois les précédents résultats obtenus par l'AFTD et l'arbre NJ.

Une estimation du F_{st} pour les ougandais a aussi été réalisée, celle-ci possède valeur de 0,20, montrant là aussi une plus grande homogénéité du groupe, d'autant que le calcul inclut en plus des populations sauvages les témoins issus de prospections en plantation.

Ces études de F_{st} sont à rapprocher de celles réalisées précédemment pour dix marqueurs microsatellites (Avia, 2004) sur des populations congolaises et guinéennes qui présentaient des F_{st} de 0,34 pour l'ensemble et de 0,37 et 0,19 respectivement pour les congolais et les guinéens, soit des valeurs proches et non significativement différentes de celles que nous obtenons.

III.6 Reconstruction de populations théoriques à l'aide de Structure

Afin de compléter nos précédents résultats, nous avons soumis une matrice de 15 marqueurs non liés sur la carte génétique à une analyse par le logiciel *Structure*. La vraisemblance des données semble être maximisée pour $K=9$ (avec K le nombre de populations supposé). Cette structuration des individus en 9 populations semble cohérente avec les résultats précédemment obtenus notamment grâce à DARwin.

Nous retrouvons ainsi de manière claire les groupes B, SG1 et C de même que les deux populations guinéennes. Le groupe SG2 forme comme pour les résultats précédents un groupe homogène malgré quelques individus qui sont assignés de manière forte au groupe C. Pour le groupe ougandais, *Structure* subdivise ces caféiers en deux groupes. Ce résultat doit cependant être pris avec précaution du fait la composition de l'échantillonnage (**annexe 7**).

Population	Effectif	Données manquantes	Marqueurs polymorphes
SG1	10	25%	21
SG2	15	22%	23
B	36	11,5%	24
C	16	9,1%	24
Mouniandougou	32	12%	20
Pélézi	35	9%	21
Ougandais sauvages	46	7%	21

Tableau 5 : Description des populations considérées après l'analyse par *Structure*. La population SG1 possédant très peu de représentants et un nombre important de données manquantes est écartée de l'étude de DL.

Population	Nombre de valeurs de P-Value calculées	Nombre de valeurs inférieures à 20%	Nombre de valeurs inférieures à 5%	Pourcentage de valeurs inférieures à 20% localisées intragroupe de liaison
B	506	85 (16,8%)	25 (4,9%)	22,4%
SG2	462	74 (16,01%)	37 (8%)	18,9%
C	506	60 (11,9%)	24 (4,7%)	16,7%
Mouniandougou	306	45 (14,7%)	18 (5,9%)	22,2%
Pélézi	380	64 (16,8%)	35 (9,21%)	31,25%
UW	342	59 (17,3%)	30 (8,8%)	20,3%

Tableau 6 : Tableau de comparaison entre populations du nombre et de la localisation des valeurs significatives de P-Value. (UW = Ougandais sauvages)

Globalement nos groupes d'origine sont homogènes à l'exception de quelques individus (**annexe 7**). Ces derniers sont soit réassignés à un groupe existant soit écartés pour la suite des analyses afin de minimiser le risque de faux positifs liés à la structure dans les calculs de déséquilibre de liaison. La décision de réassigner ou non un individu à un groupe s'est faite sur la base de la probabilité d'appartenance à ce groupe donné par structure. Nous avons choisi de ne réassigner que les individus renvoyant une valeur de 1 soit la valeur maximale et de conserver une unique population pour les ougandais sauvages.

Après réassignation, les effectifs de nos populations ont donc changés et se répartissent de la manière décrite dans le tableau 5 (les statistiques sur les marqueurs ont été estimées sur les 25 marqueurs de l'étude de déséquilibre de liaison).

Cette approche nous permet aussi de connaître ou de vérifier la structure génétique de nos témoins. Ainsi l'hybride intergroupe 888 voit son génome partitionné entre le groupe SG1 et les groupes guinéens ce qui correspond bien au croisement effectué pour son obtention. Le clone 126 est quant à lui vraisemblablement une mosaïque de différents génomes, son origine semblant être à rechercher entre les groupes B, SG2 et guinéens selon les attributions renvoyées par *Structure*.

III.7 Etude de déséquilibre de liaison

III.7.1 Test d'équilibre génotypique

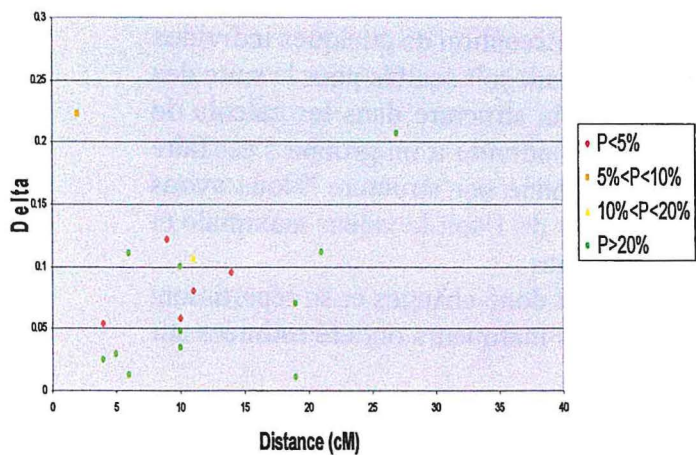
La présence d'un déséquilibre significatif a été testée sur l'ensemble des populations excepté la population SG1 pour laquelle le nombre d'individus était faible et le nombre de données manquantes important. Ce test est représenté graphiquement par la construction de matrice de contingence donnant les intervalles de P-Value renvoyée pour chaque couple de marqueurs polymorphes par le test exact de Fisher (**Annexe 3**). Afin de pouvoir comparer nos populations entre elles nous avons calculé pour chacune une proportion de valeurs de déséquilibres significatifs à 5% par rapport à l'ensemble des P-Value calculées (**tableau 6**).

Nous avons aussi calculé la proportion de déséquilibres significatifs à 5% localisée intragroupe de liaison. Nous pouvons sur la base de ces résultats grouper nos populations en deux groupes, celles présentant un nombre plus important de déséquilibres comme les populations SG2, Pélézi et UW par rapport à celles ayant de plus faibles effectifs de déséquilibres telles que C et Mouniandougou.

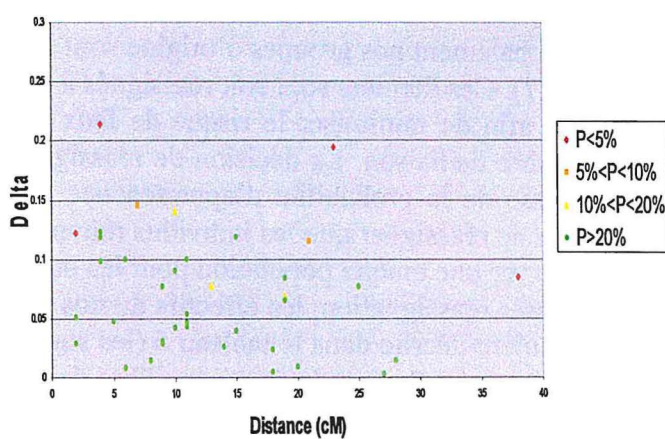
III.7.2 Evolution du déséquilibre de liaison intragroupe de liaison avec la distance génétique

Afin de comparer entre elles nos populations nous avons choisi de représenter graphiquement l'évolution du Delta (mesure composite de Weir) en fonction de la distance génétique en se plaçant à l'intérieur des groupes de liaison (**figure 10**). Nous avons représenté sur le même graphique les valeurs de delta obtenues pour chaque classe de P-Value. Nous nous intéresserons ici principalement aux deux premières classes qui représentent un seuil de significativité de 10%. Le déséquilibre présent dans chaque population décroît assez rapidement avec la distance. Pour SG2, on constate une absence de déséquilibre significatif à longue distance au sein des groupes de liaison. La population C, composé des caféiers de la Nana, présente peu de déséquilibres significatifs, ceux-ci possédant néanmoins une valeur assez élevée et persistants à une distance importante. Les individus de la population B, seule

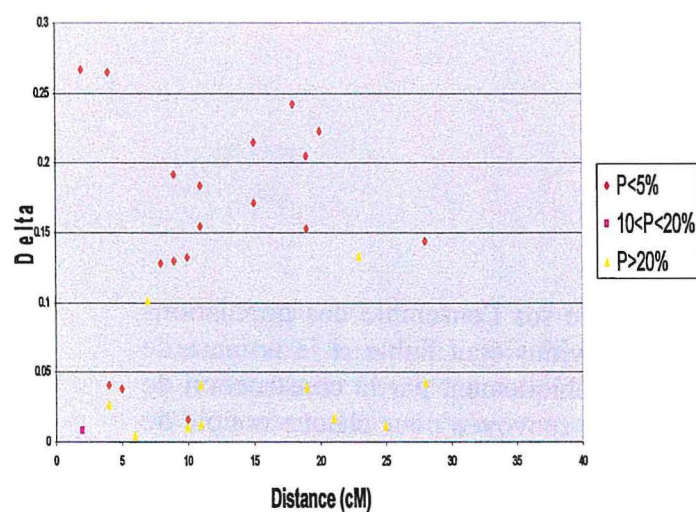
DL SG2



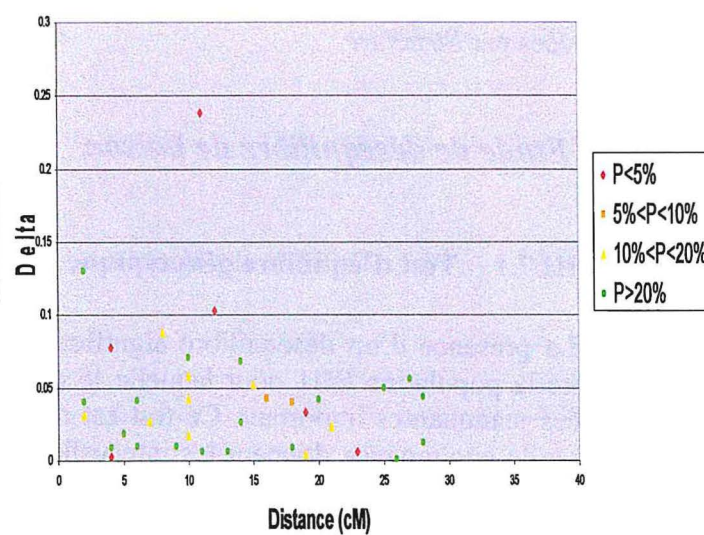
DL C



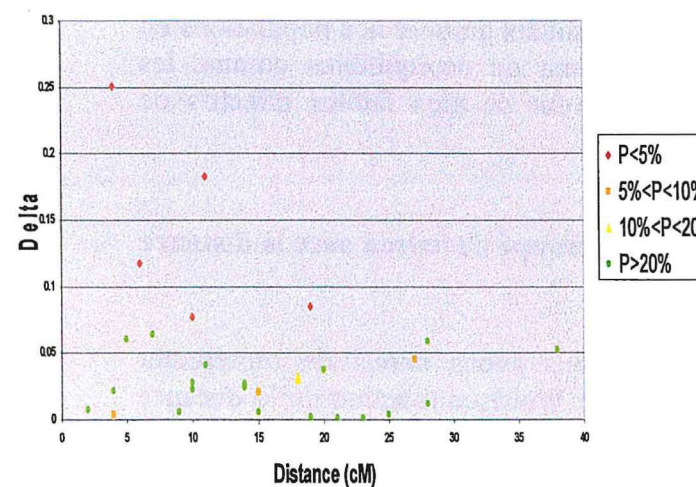
DL Pélési



DL B



DL UW



DL Mouniandougou

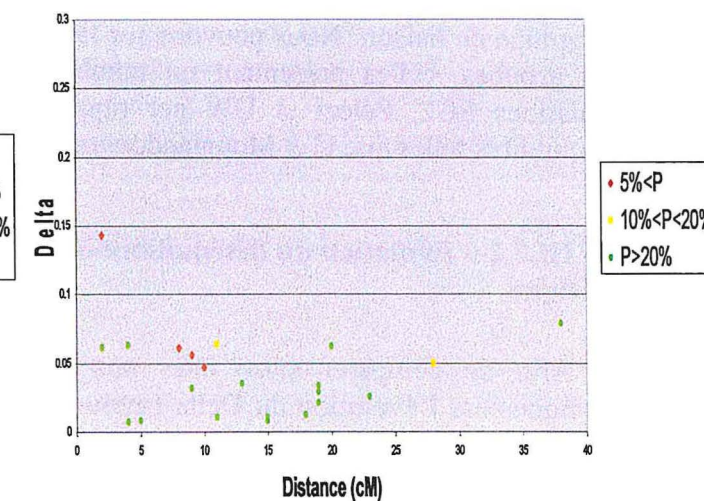


Figure 10 : évolution du Delta en fonction de la distance au sein de quelques populations de caféiers. En rouge les déséquilibres significatifs à 5%, en orangée ceux compris entre 5 et 10 %, en jaune ceux compris entre 10 et 20%, en vert supérieurs à 20%.

population véritablement sauvage de notre échantillon du groupe congolais présentent des DL faibles et peu significatifs. Ceci en opposition aux valeurs plus élevées et plus persistantes des déséquilibres de la population ougandaise sauvage qui possède moins de diversité que la population B.

Enfin si nous nous intéressons aux populations guinéennes, le déséquilibre obtenu pour la population Mouniandougou est faible, à rapprocher de celui de la population B. En revanche les déséquilibres obtenus pour Pélézi sont beaucoup plus importants, fortement localisés à l'intérieur des groupes de liaison au vu des résultats précédemment exposés et persistants à longue distance, au-delà de 20 centimorgans.

IV. Discussion

IV.1 Etude de diversité

Notre étude nous a permis de confirmer la structuration en groupe de l'espèce *C. canephora* obtenue lors de précédentes études (Dussert *et al*, 1999, Montagnon, 2000). La structuration génétique et les relations existant entre les divers groupes reflètent la structuration géographique de notre échantillon. Des études parallèles menées sur des gènes impliqués dans le métabolisme des sucres (Sucrose Synthase, Sucrose Phosphate Synthase et Cell Wall Invertase) ont confirmé nos résultats en mettant en évidence les homologies d'haplotypes entre les différents groupes, les guinéens ayant des haplotypes communs avec les groupes C et SG1 mais jamais avec les groupes B et SG2 (Bouchet, Master 2 2005). Nos résultats mettent en évidence un nouveau groupe de diversité, originaire des forêts d'Ouganda, proches génétiquement des groupes congolais précédemment décrits. La spécificité du groupe ougandais est très intéressante dans une perspective d'amélioration variétale. Ce groupe, très éloigné génétiquement des guinéens pourra être incorporé aux schémas de sélection afin de tester sa valeur en croisement. Ce nouveau groupe semble relativement proche des populations B et SG2, les plus proches géographiquement. Le groupe guinéen reste quant à lui fortement distinct de l'ensemble des autres populations, il semble donc isolé à la fois géographiquement et génétiquement. Ceci pourrait s'expliquer par le fait que lors du dernier grand épisode de glaciation les deux groupes ait été séparés dans des zones refuges différentes (Adams et Faure, 1997).

Les études de *Fst* confirment bien à la fois la forte structuration au sein de l'espèce et du groupe congolais. On retrouve pour les génotypes d'Ouganda une valeur proche de celle obtenue pour les guinéens.

La prise en compte des résultats du logiciel *Structure* pour reconstruire des populations théoriques nous a permis de confirmer la composition de nos groupes et de réassigner certains individus tout en écartant certains douteux. Dans notre cas les résultats obtenus vérifient sans équivoque la structuration obtenue par DARwin.

Nous nous attarderons plus particulièrement sur le groupe SG2 pour lequel nous avons deux populations, Ineac 2 et 7, issues d'un centre de recherche du Congo belge. Les informations sur chacune de ces populations indiquaient que nous étions probablement en présence de deux populations de demi-frères issus de graines prélevées sur un arbre mère (un pour chaque population) issus du centre de recherche INEAC du Congo belge et introduits en Côte d'Ivoire au début du siècle. Ces génotypes améliorés sont donc attendus comme formant un groupe homogène d'individus pour la représentation en analyse factorielle. Or si cela semble vérifié pour les Ineac 7 et une partie des Ineac 2, une fraction de cette dernière population se retrouve totalement incorporée au groupe C, à la fois dans l'étude de diversité réalisée à l'aide d'une analyse factorielle et dans la reconstruction des populations par *Structure*. Les individus formant le groupe que nous avons identifié SG2 sur l'AFTD et sur

les résultats de *Structure* semblent se positionner de manière intermédiaire entre différents groupes. Ce résultat semble cohérent avec l'origine géographique de ces individus. Les individus incorporés au groupe C pourraient résulter d'un mélange de plantes lors d'un des nombreux transferts de ce matériel végétal car ceux-ci sont très éloignés du groupe SG2 supposé. Toutefois les contours du groupe SG2 supposé pourront être affinés en considérant l'ensemble des génotypes assignés à ce groupe.

A l'aide de *Structure*, nous avons pu reconstruire un génome hypothétique pour le clone 126. Ce génotype serait constitué d'un tiers de génome guinéen, un tiers de génome SG2 et un tiers de B. Il s'agit donc bien d'une structure très hybride. Ce génotype a été prospecté en plantation en Côte d'Ivoire et pourrait résulter d'une hybridation entre des guinéens locaux d'une part et des génomes introduits de la zone congolaise d'autre part. Ceci est particulièrement intéressant dans une perspective d'amélioration variétale. En effet ce clone est l'un des meilleurs connus en terme de qualité, ce qui pourrait amener les sélectionneurs à tenter de reconstruire ce génotype ou de produire de nouveaux hybrides sur la base du croisement intergroupe supposé avoir donné ce clone, à savoir un croisement trois voies de type (B x SG2) x Guinéen.

Enfin nous soulignerons que malgré la présence de matériel amélioré et le brassage génétique à la fois par hybridation et par mouvement de matériel sélectionnés entre les grands centre de sélections successifs (Java, Madagascar, Ouganda, Congo belge, Côte d'Ivoire,...) la structuration en groupe reste forte et visible, tendant à montrer que les flux de gènes entre compartiments sauvages et cultivés restent limités, ceci étant d'autant plus vrai qu'il s'agit d'une espèce très récemment domestiquée ayant subi peu de cycles de recombinaisons depuis la sortie de la forêt.

IV.2 Déséquilibre de liaison

Les populations sauvages congolaises (B) et cultivées issues de variété spontanée (C) ainsi que les Mouniandougou (G) présentent un nombre de valeurs de DL significatives assez limité sans doute du au fait que ces populations n'ont pas ou peu subi de pressions de sélection ou de goulot d'étranglement, et qu'elles semblent proches d'un équilibre entre recombinaison et dérive.

En revanche la population SG2 cultivée présente un déséquilibre important. Ce résultat est étonnant car nous sommes en présence de génotypes améliorés et l'on pourrait s'attendre à un déséquilibre plutôt faible grâce au brassage génétique important et donc d'un accroissement de la recombinaison efficace. Cependant le fait que nous soyons en présence de populations constituées, à priori, de demi-frères pourrait expliquer ce résultat. On observe d'autre part une absence de déséquilibre significatif à longue distance au sein des groupes de liaison. Cela peut-être un effet du mélange de génomes ayant mené à la constitution de cette population, ceci est en adéquation avec l'histoire supposée de cette population. Cependant le déséquilibre obtenu persiste tout de même jusqu'à environ 15 cM. Ceci est à mettre en regard du précédent résultat et semble conforter notre analyse, à savoir que le déséquilibre important observé est principalement dû au fait que nous sommes en présence de populations de demi-frères.

La population C possède peu de DL significatifs mais ceux-ci sont élevés et persistent à logue distance. Ce phénomène peut être du à la sélection de quelques génotypes seulement des populations sauvages lors de la mise en culture résultant en un goulot d'étranglement ayant augmenté le déséquilibre.

La population Pélézi présente un grand nombre de valeurs significatives à longue distance, contrairement à la population Mouniandougou. Ceci peut s'expliquer par la structure même de cette population. En effet, il s'agit d'une population établie récemment, en

conditions plutôt sèches, et construite de la manière suivante : un arbre mère unique est vraisemblablement responsable de l'établissement de la population, deux arbres un peu moins gros participent à la reproduction et enfin une cinquantaine de semenceaux sont dispersés à la base de ces trois arbres mères. Cette histoire même place Pélézi dans le cadre typique d'un goulot d'étranglement du à un effet fondateur. Cette population se détache donc de manière très nette de toutes les autres. D'autant que cette population utilisée dans les schémas de sélection s'est révélée à la fois très particulière et très intéressante en croisement avec des génotypes guinéens.

Les ougandais sauvages présente un DL moyen tant en nombre qu'en persistance. Ces populations sauvages semblent à l'équilibre et se positionnent comme intermédiaires entre d'autres types de population. Il semblerait que ce type de comportement en terme de structure de déséquilibre soit celui attendu pour des populations sauvages isolées telles que celles que nous avons étudiées.

La population Mouniandougou, bien qu'appartenant au même groupe de diversité que Pélézi présente un déséquilibre beaucoup plus restreint. Cette différence peut s'expliquer par plusieurs facteurs, tout d'abord cette population est moins soumise au stress hydrique que la population Pélézi, elle semble par ailleurs implantée depuis un nombre de cycle plus long et donc plus proche d'un équilibre que la population Pélézi. Enfin un grand nombre d'arbres participent à la reproduction (au moins trente) et de récents flux de gènes avec d'autres populations pourraient aussi expliquer les faibles valeurs observées.

Ces résultats confirment ceux précédemment obtenus (Avia, 2004) et permettent d'envisager des applications directes au DL observé. Il semble que dans notre cas les phénomènes démographiques sont responsables des plus grandes variations observées et sont sans doute à l'origine de la particularité importante de la population guinéenne Pélézi qui est la seule à présenter de très forts déséquilibres significatifs.

V. Conclusions et perspectives

Dans cette étude nous n'avons analysé qu'un échantillon des populations connues. Afin de prolonger notre étude et d'affiner la connaissance de la diversité de l'espèce *Coffea canephora*, il sera important d'augmenter le nombre d'individus analysés par groupe de diversité et d'incorporer des génotypes cultivés des différents groupes (Guinéens, Congolais, Ougandais). En parallèle un travail d'évaluation de la diversité et des flux de gènes au sein des différentes populations ougandaises est en cours de réalisation. Ce travail pourra permettre de mieux connaître ce nouveau groupe de diversité, de l'évaluer et de l'incorporer aux schémas d'amélioration variétale de l'espèce.

Les formes N'Ganda et Erect pourront aussi être caractérisées pour le déséquilibre de liaison, ainsi que les populations de ferme.

Il sera utile de développer d'autres marqueurs tels que des SNP et de multiplier le nombre de marqueurs microsatellites génotypés sur les populations d'intérêt et ce pour deux approches complémentaires. L'une concernera la recherche de QTL par des études d'association sur des populations adaptées et caractérisées, une autre consistera à identifier des zones du génome fortement soumises à sélection, qu'elle soit naturelle ou artificielle, et donc recelant potentiellement des gènes d'intérêt pour la qualité.

Le but de nos travaux était de mettre en évidence la structure globale de DL au sein de l'espèce *Coffea canephora* et d'utiliser au mieux les diverses populations dont nous disposons pour les différents types d'applications pratiques que nous envisageons :

- Tout d'abord le fort DL à forte distance de la population Pélézi semble l'identifier comme une population adaptée aux études d'association pour lesquelles un nombre limité de marqueurs, tous les 4 à 5 cM environ devrait suffire pour couvrir tout le génome avec des marqueurs qui seraient alors liés fortement avec des gènes d'intérêts. Une sélection assistée par marqueur faisant intervenir cette population serait alors très efficace.
- Les génotypes guinéens cultivés semblent être des hybrides ou bien un groupe hétérogène couvrant à peu près l'ensemble de la diversité des guinéens (Avia, 2004), ces guinéens cultivés pourront eux aussi être caractérisés afin de comparer la dynamique de DL entre les différents types de populations sauvages et les populations cultivées.
- Les populations sauvages présentant un déséquilibre intermédiaire pourront servir à la fois comme source de diversité et de nouveaux allèles pour la sélection ainsi que pour des études de cartographies plus fines.

Nos études montrent l'intérêt pour la sélection de *Coffea canephora* de divers groupes. Tout d'abord elles confirment la spécificité et l'intérêt en sélection des génotypes guinéens. Le groupe ougandais qui constitue un nouveau groupe de diversité qui pourra être évalué dans les schémas de sélection. La reconstitution du génome du clone 126 montre l'intérêt des croisements entre génotypes des différents groupes congolais.

Enfin les résultats obtenus pourront mener à la mise en place de la sélection assistée par marqueurs. Néanmoins cette dernière devra être réalisée sur la base d'un plus grand nombre de marqueurs localisés dans des régions où se trouvent des QTL. La mise en place d'une telle stratégie permettrait l'évaluation précoce des hybrides issus de la SRR.

Bibliographie:

1. Adams, J.M., Faure, H. et QEN members (1997) Review and Atlas of Paleovegetation: preliminary land ecosystem maps of the world since the Last Glacial Maximum. (TN, ed), Oak Ridge National Laboratory
2. Avia, K. (2004) Cartographie génétique de *Coffea canephora* et premières analyses du déséquilibre de liaison dans onze populations. In *DEA Génétique et Amélioration des Plantes*, Ecole Nationale Supérieure d'Agronomie de Rennes
3. Berline, M. (2002) Définition de marqueurs microsatellites polymorphes et cartographie génétique préliminaire d'une descendance intraspécifique de *Coffea canephora*. In *UFR SVT, DESS Bioingénierie*, 34, Université Paul Sabatier, Toulouse III
4. Berthaud, J. (1980) L'incompatibilité chez *Coffea canephora*: méthodes de test et déterminisme génétique. *Café Cacao Thé* 24, 267-274
5. Berthaud, J. (1986) *Les ressources génétiques pour l'amélioration des caféiers africains diploïdes: évaluation de la richesse génétiques des populations sylvestres et de ses mécanismes organisateurs, conséquences pour l'application.*
6. Bolstein, D., White, R.L., Skolnick, M. et Davis R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32, 314-331
7. Bouchet, S. (2005) Diversité nucléotidique de cinq gènes impliqués dans la voie de biosynthèse du saccharose chez *Coffea canephora*: évolution moléculaire et implication pour la mise en place de la sélection assistée par marqueurs pour la qualité organoleptique. In *Master 2 recherche, GAP*, Ecole Nationale Supérieure d'Agronomie de Rennes
8. Charrier, A., Eskes, A.B. (1998) Les caféiers. BRG http://www.brg.prd.fr//brg/ecrans/rgvPi_cafeiers.htm
9. Dussert, S., Lashermes, P., Anthony, F., Montagnon, C., Trouslot, P., Combes, B.C., Berthaud, J., Noirot, M., Hamon, S. (1999) Le caféier, *Coffea canephora*. In *Diversité génétique des plantes tropicales cultivées* (CIRAD, ed), Perla Hamon, Marc Séguin, Xavier Perrier et Jean-Christophe Glaszmann
10. Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S. (2003) Structure of linkage disequilibrium in plants. *Annual Reviews of Plant Biology* 54, 357-374
11. Janoo, J., Grivet, L., Dookun, A., D'Hont, A. (1999) Linkage disequilibrium among modern sugarcane cultivars. *Theor. Appl. Genet.* 99, 1053-1060
12. Jarne, P., Lagoda, P.J.L. (1996) Microsatellites, from molecules to populations and back. *Tree* 11, 424-429
13. Jorde, L.B. (2000) Linkage disequilibrium and the search for complex disease genes. *Genome Research* 10, 1435-1444

14. Leroy, T., Montagnon, C., Cilas, C., Yapo, A.B., Charmetant, P. et Eskes, A.B. (1997) Reciprocal recurrent selection applied to *Coffea canephora* Pierre. III. Genetic gains and results of first intergroup crosses. *Euphytica* 95, 347-354
15. Leroy, T., Montagnon, C., Cilas, C., Charrier, A. et Eskes, A.B. (1994) Reciprocal recurrent selection applied to *Coffea canephora* Pierre. II. Estimation of genetic parameters. *Euphytica* 74, 121-128
16. Leroy, T., Marraccini, P., Dufour, M., Montagnon, C., Lashermes, P., Sabau, X., Ferreira, L., Jourdan, I., Pot, D., Andrade, A., Glaszmann, J.C., Vieira, L. et Pietro Piffanelli (2005) Construction and characterization of a *Coffea canephora* BAC library: a valuable tool to study genome structure of sucrose biosynthetic genes. *Theor. Appl. Genet.* In press
17. Lewis, K.F., et al. (2001) The effectiveness of computational methods in haplotype prediction. *American Journal of Human Genetics* 69, 503-503
18. Liu, K., Muse, S. (2004) PowerMarker: new genetic data analysis software. Version 3.0. Free program distributed by the author over the internet from <http://www.powermarker.net>.
19. Montagnon, C. (2000) Optimisation des gains génétiques dans le schéma de sélection récurrente réciproque de *Coffea canephora* Pierre. In *Thèse de doctorat*, Ecole Nationale Supérieure d'Agronomie de Montpellier
20. Montagnon, C., Leroy, T. et Yapo, A.B. (1992) Diversité génotypique et phénotypique de quelques groupes de caféiers (*Coffea canephora* Pierre) en collection. Conséquences sur leur utilisation en sélection. *Café Cacao Thé* 36, 187-198
21. Perrier, X., Flori, A., Bonnot, F. (2003) Data analysis methods. In *Genetic diversity of cultivated tropical plants* (Hamon, P., Seguin, M., Perrier, X., Glaszmann, J. C., ed), 43 - 76, Enfield, Science Publishers
22. Poncet, V., Hamon, P., Minier, J., Carasco, C., Hamon, S., Noirot, M. (2004) SSR cross-amplification and variation within coffee trees. *Genome* 47, 1071-1081
23. Pritchard, J.K., et al. (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959
24. Pugh Moreno, T. (2005) Etude des déséquilibres de liaison dans une collection de cacaoyers (*Theobroma cacao* L.) appartenant au groupe Criollo/Trinitario et application au marquage génétique des caractères d'intérêt. In *Thèse de doctorat*, Ecole Nationale Supérieure d'Agronomie de Montpellier.
25. Rafalski, A., Morgante, M. (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in genetics* 20, 103-111
26. Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M., Bucler IV, E.S. (2001) Structures of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* 98, 11479-11989

27. Thomas, A.S. (1935) Types of Robusta coffee and their selection in Uganda. *The East African Agricultural Journal* Nov., 1935, 193-197
28. Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D., Buckler, E.S. (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics* 28, 286-289
29. Weir, B.S. (1996) *Genetic data analysis II*. Sinauer, Sunderland, MA

Protocole d'extraction d'ADN de café (Méthode MATAB)

I) Broyage des échantillons :

1. Nettoyer les feuilles à l'éthanol (EtOH 70 %) quand les conditions le permettent.
2. Broyer tout le matériel dans un mortier (refroidi préalablement au N₂) en présence d'azote liquide (N₂) et d'une pointe de spatule de PVPP (polyvinylpolypyrrolidone) jusqu'à l'obtention d'une fine poudre. Stocker cette poudre à -80°C dans un tube « Sarstedt » de 50 mL.
3. Nettoyer le mortier et le pilon à l'éthanol entre chaque échantillon. => répéter les étapes 1 à 3 autant de fois que nécessaire.

N.B. : Le broyage (étapes 1 à 3) peut-être effectué séparément de l'extraction. Broyez le matériel végétal frais dès son arrivée sur site. Les extractions peuvent ensuite être organisées suivant le temps disponible.

II) Lyse nucléaire :

Tampon d'extraction MATAB :

	Concentration finale	250 mL de tampon
Tris-HCl 1M pH 8.0	100 mM	25 mL
NaCl 5M	1.4 M	70 mL
EDTA 0.5M pH 8.0	20 mM	10 mL
MATAB = Mixed Alkyltrimethylammonium Bromide	2 %	5 g
PEG 6000 = PolyEthylene Glycol	1 %	2.5 g
Sulfite de Sodium (NaSO₄)	0.5 %	1.25 g
H₂O distillé		Compléter jusqu'à 250 mL

1. Préparer des tubes Sarstedt 15 mL nominatifs.
2. Ajouter 5 mL de tampon d'extraction et garder les tubes au bain-marie à 74°C.
3. Ajouter 600 mg de poudre et homogénéiser par retournement.
4. Laisser incuber 30 minutes au bain-marie sous agitation à 74°C.
5. Laisser refroidir 10 minutes à température ambiante.
6. Ajouter 1 volume (i.e. 5 mL) de Chloro IsoAmylAlcohol (CIAA).
7. Mélanger jusqu'à émulsion.
8. Centrifuger à 9000 g pendant 15 minutes.
9. Transférer la phase aqueuse dans un nouveau tube « Sarstedt » de 15 mL avec une pipette Pasteur en prenant soin de ne pas prélever le dépôt à l'interphase.
10. Vider le reste du tube dans le bidon solvant chloré (jeter le tube dans la poubelle blanche).

III) Précipitation de l'ADN :

1. Ajouter 0,7 volume d'isopropanol (propanol-2).
2. Agiter doucement par retournement.
3. Conserver à -20°C les tubes « Sarstedt » pendant au moins 20 minutes ou plus longtemps pour stockage.

N.B. : Les tubes peuvent être conservés à -20°C la nuit pour faciliter l'organisation.

4. Centrifuger à 8000 g pendant 20 minutes.
5. Enlever l'isopropanol.
6. Transférer le culot dans un tube Eppendorf 1,5 mL.
7. Laver le culot avec 500 µL d'éthanol 70 %.
8. Centrifuger 20 minutes à 20000 g.

N.B. : Programmer la centrifugeuse à 4°C en prévision de l'étape 15.

9. Enlever l'éthanol.
10. Sécher le culot avec la pompe à vide (SpeedVac). Vérifier visuellement pour l'obtention d'un bon résultat.
11. Réhydrater le culot avec 600 µL de Nuclei Lysis Solution.
12. Laisser incuber au bain-marie 20 minutes à 65°C (homogénéiser jusqu'à complète re-suspension de l'ADN).
13. Refroidir les échantillons immédiatement à 4°C.
14. Ajouter 200 µL de Protein Precipitation Solution.
15. Centrifuger 20 minutes à 20000 g.
16. Transférer la phase supérieure dans un nouveau tube Eppendorf 1,5 mL.
17. Ajouter 0,7 volume d'isopropanol (propanol-2).
18. Agiter doucement par retournement.
19. Centrifuger 20 minutes à 20000 g.
20. Enlever l'isopropanol et sécher le culot à l'aide du SpeedVac (c.f. étape 10).
21. Ajouter 200 µL de solution TE (Tris-EDTA).
22. Ajouter 0,5 U de Rnase A (Qiagen 1mg / mL) pour 100 µL de solution.
23. Homogénéiser par retournement.
24. Laisser les tubes sous agitation pendant 1 à 2 jours avant de les conserver à 4°C. Attendre la re-suspension complète du culot (au moins 3 jours) avant de quantifier l'ADN.

Remarques :

- Les vitesses et les temps de centrifugation peuvent être adaptés suivant les centrifugeuses disponibles.
- Le PVPP est stocké en salle d'extraction d'ADN, dans une boîte grise marquée « PVP » (réf. P6755 SIGMA).
- Il est extrêmement important de maintenir les échantillons en permanence au froid afin d'éviter l'initiation de tous les processus d'oxydation.
- Lors du transfert du culot séché, il est important de ne pas attendre la totale dessiccation de ce dernier afin de ne pas trop le détériorer avec la pipette Pasteur.
- Les solutions Protein Precipitation Solution (A7953) et Nuclei Lysis Solution (A7943) sont des solutions faisant parties du Wizard Genomic DNA Purification Kit de Promega.

Annexe 2

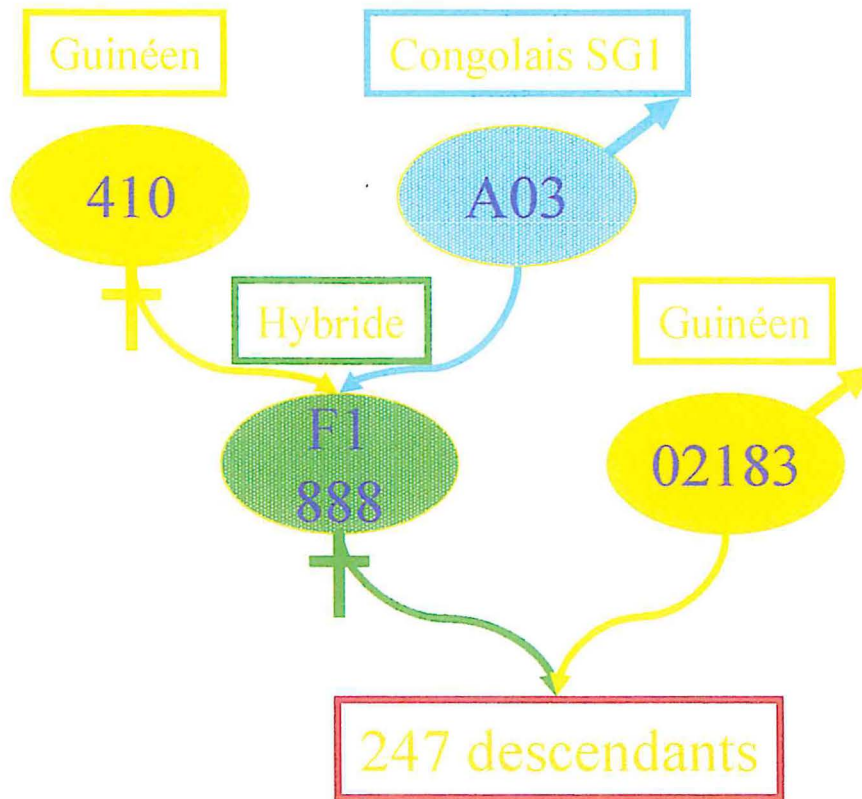


Schéma de croisement à l'origine de la population de cartographie de la carte utilisée. (Berline, 2001)

Protocole tailing avec M13-amorce + M13 marqué à l'IR-700 ou 800, Cirad-Biotrop

Rq : ces concentrations sont spécifiques au laboratoire Biotrop (le tampon 10x contient par exemple déjà du MgCl2)

Marqueur :

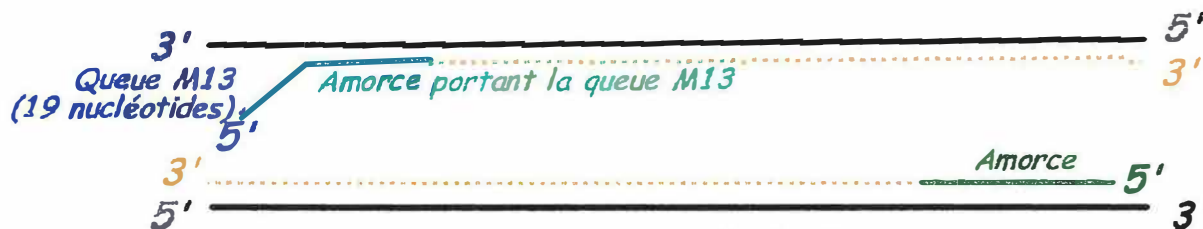
	<i>[Initiale]</i> Unité	<i>Qté finale</i> Unité	<i>Vol. d'ADN par réaction</i>
ADN	0.5 ng/µl	ADN 2.50 ng	ADN 5.00 µl
	<i>[Initiale]</i> Unité	<i>[Finale]</i> Unité	<i>Vol. réactifs ds mix</i>
Tp	10 X	Tp 1.0 X	Tp 270.00 µl
dNTP	2000 µM	dNTP 200 µM	dNTP 270.00 µl
MgCl2	50 mM	MgCl2 0.50 mM	MgCl2 27.00 µl
Am1-M13	10 µM	Am1-M13 0.08 µM	Am1-M13 21.60 µl
Am2	10 µM	Am2 0.10 µM	Am2 27.00 µl
M13 marqué	10 µM	M13 marqué 0.10 µM	M13 marqué 27.00 µl
Taq	2 U/µl	Taq 0.07 U/µl	Taq 94.50 µl
			H2O 612.90 µl
Nombre de réactions :		225	
Volume de réaction :		10 µl	Volume Mix 1350.00 µl
Marge pour le mix :		20 %	Vol. à distribuer 5.00 µl

Composition du mix PCR et quantité requise des différents produits le composant.

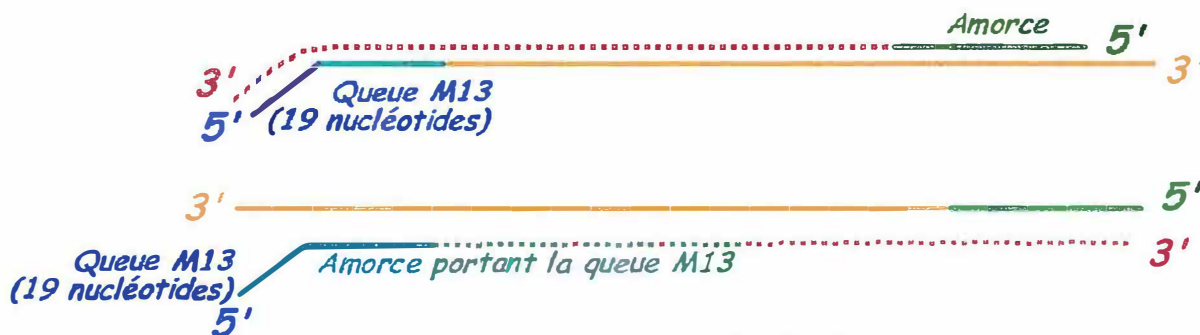
PCR réalisée pour le passage au séquenceur automatique



↓ Cycle 1



↓ Cycle 2



↓ Cycle 3



Annexe 5

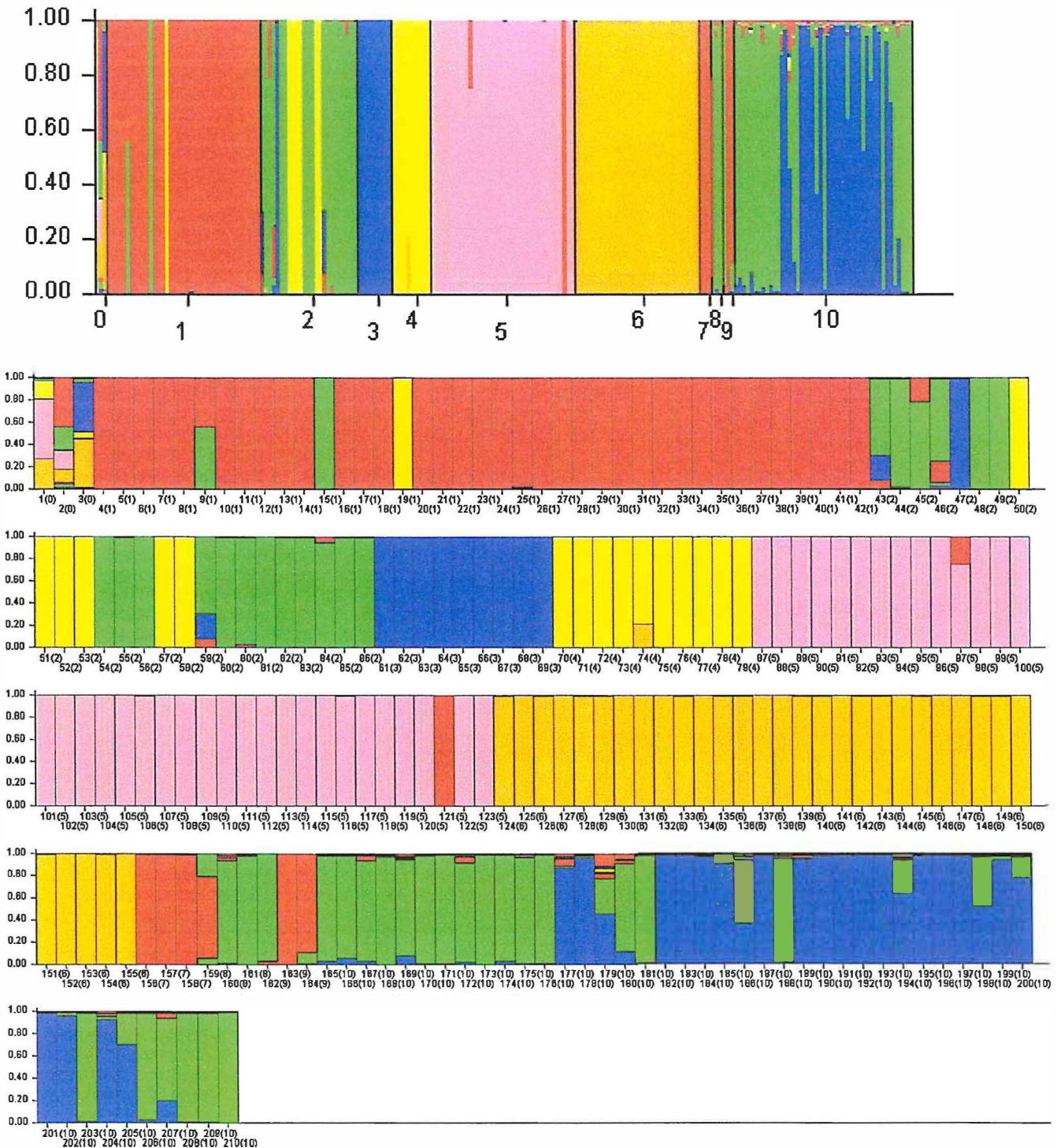
Marker	Major.Allele.Frquency	AlleleNo	Availability	GeneDiversity	PIC
DL011	0.49	10.00	0.84	0.66	0.61
DL013	0.36	13.00	0.98	0.73	0.68
DL025	0.34	7.00	0.83	0.78	0.75
DL026	0.41	9.00	0.97	0.76	0.73
257	0.50	7.00	0.98	0.60	0.52
329	0.31	13.00	0.92	0.82	0.80
351	0.29	17.00	0.84	0.86	0.84
355	0.31	22.00	0.73	0.85	0.83
358	0.44	16.00	0.93	0.74	0.72
360	0.21	14.00	0.87	0.88	0.87
364	0.37	10.00	0.75	0.75	0.71
367	0.39	11.00	0.92	0.78	0.76
368	0.18	20.00	0.97	0.90	0.89
371	0.59	7.00	0.95	0.59	0.55
384	0.43	10.00	0.97	0.73	0.70
394	0.72	10.00	0.91	0.47	0.46
429	0.26	12.00	0.93	0.85	0.83
442	0.14	24.00	0.71	0.91	0.91
445	0.52	10.00	0.89	0.56	0.47
456	0.19	19.00	0.93	0.89	0.88
460	0.47	9.00	0.90	0.71	0.68
461	0.25	15.00	0.60	0.87	0.85
463	0.52	6.00	0.90	0.66	0.61
471	0.26	17.00	0.96	0.84	0.82
495	0.58	7.00	0.91	0.57	0.51
501	0.52	13.00	0.86	0.69	0.67
753	0.33	15.00	0.94	0.78	0.75
755	0.36	16.00	0.97	0.83	0.82
774	0.69	5.00	0.90	0.48	0.44
779	0.22	13.00	0.89	0.85	0.84
782	0.94	4.00	0.97	0.11	0.10
790	0.36	22.00	0.98	0.84	0.83
837	0.28	11.00	0.96	0.84	0.82
838	0.25	11.00	0.93	0.79	0.76
Mean	0.40	12.50	0.90	0.73	0.71

Annexe 6

Tableau : Caractéristiques phénotypiques des populations Guinéennes et Congolaises de *Coffea canephora* . Les groupes B et C sont considérés comme faisant partie du sous groupe 2 congolais. (D'après Montagnon 2000)

		Congolais			Guinéens	
		SG1	SG2		cultivés	sauvages
		cultivés	cultivés	sauvages		
		Togo, Robusta A1, Niaouli, Kouilou Madagascar, Ebobo, Aboisso	Boukoko, Ouganda, C10 Man, Lula, INEAC	Libengue, Doungba, Nana ?	Guinéens cultivés	Ira1, Bafingdala, Ira2, Pine, Kouin, Gbapleu, Fourounbankoro, Bossematie, Maraoué, Gbapleu2, Gonate, Logbonou, Pelezi
feuilles	longueur (mm)	188	203	221	170	176
	largeur (mm)	71	82	93	68	75
	longueur/largeur	2.68	2.49	2.38	2.53	2.36
	Acumen	court à long	long	long	long	long
	Pétiole	court à long	Moyen à long	long	court à long	moyen à long
	domaties	absentes à bien visibles	peu marquées à bien visibles	peu marquées à bien visibles	peu marquées à bien visibles	peu marquées à bien visibles
fruits	pédoncule du fruit	court à moyen	court à moyen	court à long	court à long	court à long
	taille du disque	petit	petit	petit	grand	petit
	relief du disque	plat à saillant	saillant	plat à saillant	plat	plat
	granulométrie (g pour 100 grains, 12 % humidité)	14.2	15.4	9.5	10	9
divers	ramification	moyenne à forte	faible à moyenne	faible à moyenne	moyenne à forte	moyenne à forte
	longueur entrenoeuds	moyen à court	long	long	court	court
	phénologie	très tardif	moyennement tardif	moyennement tardif	précoce	précoce
	caféine (% matière sèche)	2.7	2.3	2.3	2.9	2.6
	rouille orangée (% de plants résistants)	73	84	97	31	28
	scolytes de branchettes (% de plants résistants)	94	73	50	88	77
	résistance à la sécheresse	bonne	faible		bonne	
	qualité organoleptique	Bons arômes et acidité. Amertume et corps moyens	Bons arômes et acidité. Faible amertume et corps		Médiocre : Faibles arômes et acidité. Fortes amertumes et corps	

Annexe 7



Sorties graphique structure. En haut et en bas même sortie mais à échelle différente. La seconde sortie permet de décrire le génome du clone 126. Population 0 = témoins de cartographie et clone 126 (individu n°2) 1 = B, 2 = SG2, 3 = SG1, 4 = C, 5 = Pélézi, 6 = Mouniandougou, 7, 8 et 9 témoins ougandais (Erect, N’Ganda, Farm), 10 = Ougandais sauvages

Philippe Cubry

CIRAD-CP, UMR T52 PIA, TA 80/03, Avenue Agropolis, 34398 Montpellier Cedex 5, France

ANALYSE DE LA DIVERSITE GENETIQUE ET EVALUATION DU DESEQUILIBRE DE LIAISON DANS QUELQUES POPULATIONS NATURELLES ET CULTIVEES DE *Coffea canephora*.

Le café Robusta, produit par *Coffea canephora*, représente environ 35% de la production mondiale de café. Dans le cadre de recherches visant à l'amélioration de la qualité à la tasse du café, la connaissance de la diversité génétique des populations naturelles et cultivées présente un intérêt important. Dans un premier temps, une étude de diversité a été réalisée à l'aide de 34 marqueurs SSR répartis sur l'ensemble du génome. Cette étude comprend des populations d'individus des grands groupes génétiques préalablement identifiés ainsi que des populations sauvages et cultivées d'Afrique de l'Est jamais analysées pour leur diversité. Les résultats obtenus confirment la structuration en groupes décrite précédemment et indiquent l'originalité des populations d'Afrique de l'Est dans l'espèce *Coffea canephora*. Ces populations pourraient constituer une variabilité génétique utilisable par les sélectionneurs.

Une évaluation du déséquilibre de liaison pour des marqueurs entre lesquels la distance génétique est connue a ensuite été menée pour 25 marqueurs microsatellites sur les populations analysées en diversité. Les résultats de cette étude pourront être utilisés pour les programmes d'amélioration des populations en cours et serviront de base à la définition de populations utilisables pour de futures études d'association.

GENETIC DIVERSITY ANALYSES AND LINKAGE DISEQUILIBRIUM EVALUATION IN SOME NATURAL AND CULTIVATED POPULATIONS OF *Coffea canephora*.

Philippe Cubry

CIRAD-CP, UMR T52 PIA, TA 80/03, Avenue Agropolis, 34398 Montpellier Cedex 5, France

Robusta coffee, produced by *Coffea canephora*, represents about 35% of the world coffee production. As part of studies aiming at the improvement of coffee cup quality, the knowledge of the genetic diversity of the wild and cultivated populations is an important task. First, a diversity study was undertaken with 34 SSR markers distributed on the whole genome. This study concerned several populations belonging to the main known coffee diversity groups, as well as wild and cultivated East African populations that were never analyzed for diversity. The results confirmed the groups structure previously described and stressed out the original features of the East African populations that seem to represent a new homogenous group. This new group could allow the breeders to broaden the genetic basis they use for coffee improvement. Second, an estimation of linkage disequilibrium for markers for which the genetic distance is known was made for 25 SSR markers on the same populations as in the diversity study. The results of this work will be useful to breeding programs and will be the basis for future association studies.