

ACADEMIE DE MONTPELLIER

UNIVERSITE MONTPELLIER II  
SCIENCES ET TECHNIQUES DU LANGUEDOC

THESE

présentée à l'Université de Montpellier II Sciences et Techniques du Languedoc  
pour obtenir le diplôme de DOCTORAT

SPECIALITE: Physiologie et Biologie Des Organismes et des Populations

Formation Doctorale: Bases de la production végétale

Ecole Doctorale: Biologie des Systèmes Intégrés-Agronomie, Environnement

**CONTRIBUTION A L'ETUDE GENETIQUE DU CACAOYER PAR  
LES MARQUEURS MOLECULAIRES: DIVERSITE GENETIQUE ET  
RECHERCHE DE QTLs**

par

**Jeanne Andi KOHI ép. N'GORAN**

Soutenue le 03 Octobre 1994 devant le jury composé de:

Mme LANAUD Claire, Responsable de laboratoire  
CIRAD, Montpellier

Directeur de thèse

M. BERVILLE André, Directeur de Recherches  
INRA, Montpellier

Examineur

Mme CASSE-DELBART Francine, Professeur  
Université Montpellier II

Présidente

M. COMBES Daniel, Professeur, Université de Pau

Rapporteur

M. GOUYON Pierre Henri, Professeur, Université de Paris XI

Rapporteur

M. KOUASSI Auguste, Maître-Assistant,  
Université d'Abidjan, Côte d'Ivoire

Examineur

## REMERCIEMENTS

Je voudrais remercier Françoise Casse-Delbart, professeur à l'Université Montpellier II d'avoir accepté de présider le jury de cette thèse.

Je tiens à remercier Pierre Henry Gouyon, Professeur à l'université de Paris XI ainsi que Daniel Combes, professeur à l'Université de Pau qui malgré leurs nombreuses occupations, ont accepté d'être les rapporteurs de ce travail.

André Bervillé, Directeur de recherches à l'INRA de Montpellier a porté un intérêt à ce travail. Ses remarques sur le manuscrit ont été précieuses. Qu'il en soit remercié.

Je remercie Auguste Kouassi, maître-assistant à l'Université d'Abidjan et conseiller du Ministre de l'enseignement supérieur et de la recherche scientifique de Côte d'Ivoire, d'avoir effectué le voyage en abandonnant ses lourdes occupations pour venir juger ce travail. J'ai trouvé en lui réconfort et soutien moral.

Je tiens particulièrement à remercier Claire Lanaud, responsable du laboratoire Agetrop du CIRAD, qui m'a permis d'effectuer cette formation dans son laboratoire. Sans elle, ce document n'aurait jamais vu le jour. C'est elle qui est l'initiatrice de ce travail. Par son enthousiasme et son optimisme elle m'a soutenue dans les moments difficiles.

Je remercie vivement Albertus Eskes qui m'a fait bénéficier de sa rigueur scientifique. Il a consacré beaucoup de temps à la correction du manuscrit et les nombreuses discussions que nous avons eues m'ont aidée à ne pas m'éloigner des réalités de la sélection classique du cacaoyer. Mes remerciements vont aussi à Didier Paulin et à Christian Cilas pour les discussions fructueuses que nous avons eues.

J'aimerais aussi remercier Perla Hamon. Bien que n'étant pas directement concernée par ce travail, elle s'y est beaucoup impliquée et ses nombreuses remarques sur le manuscrit m'ont été d'une aide appréciable. Ma gratitude va aussi à Mathias Lorieux, qui m'a aidée à réaliser les analyses de cartographie et de QTL ainsi qu'Ange Marie Risterucci qui m'a appris certaines techniques de laboratoire.

Mes remerciements vont aussi à Nanga Coulibaly, directeur de l'IDEFOR-DCC qui n'a jamais cessé de me considérer comme membre de son personnel. Il m'a ainsi permis de réaliser les observations en champ au cours des brefs séjours en Côte d'Ivoire. Je remercie aussi Didier Clément, Nathanaël Affian, Gérard Sagou, Zah Paul et toute l'équipe de génétique du cacaoyer à l'IDEFOR-DCC qui m'ont assistée dans la collecte des données au champ.

Je remercie toute l'équipe du laboratoire Agetrop, en particulier Françoise Carreel, Maguy Rodier, Patricia Lebrun et Laurent Grivet qui m'ont aidée à garder le moral haut. Mes remerciements vont aussi à André Bouet pour les nombreux services qu'il m'a rendus et à tous ceux qui de près ou de loin ont été impliqués dans ce travail.

Je voudrais exprimer ma gratitude au ministère de la coopération française pour m'avoir octroyé une bourse de formation et au ministère de l'enseignement supérieur et de la recherche scientifique de Côte d'Ivoire pour avoir accepté mon détachement auprès du laboratoire Agetrop du CIRAD le temps de cette thèse.

Mes remerciements vont enfin à mes enfants, Colette, Roseline, Carmen et Jules, et à mon mari Nestor. Sans leur compréhension, leur patience et leur indulgence, je n'aurais jamais pu accomplir ce travail.

INTRODUCTION.....	1
CHAPITRE I: DONNEES GENERALES SUR LE CACAOYER.....	3
I- IMPORTANCE ECONOMIQUE.....	3
II- MORPHOLOGIE ET BIOLOGIE.....	4
1- Maxonomie.....	4
2- Morphologie.....	4
3- Biologie.....	5
4- Agronomie.....	6
III- LES POPULATIONS DE CACAOYERS.....	7
1- Les cacaoyers Criollo.....	8
2- Les cacaoyers Forastero.....	8
Les Forastero Haut Amazoniens.....	9
Les Forastero Bas Amazoniens.....	9
3- Les cacaoyers Trinitario.....	10
IV- ORGANISATION DE LA DIVERSITE GENETIQUE DU CACAOYER...	10
1-Diversité morphologique.....	10
2-Diversité enzymatique.....	11
3-Diversité révélée par les marqueurs moléculaires.....	12
V- AMELIORATION ET TRAVAUX DE SELECTION.....	13
CHAPITRE II : INTERETS DES MARQUEURS MOLECULAIRES.....	16
I- LES MARQUEURS RFLP.....	16
II- LES MARQUEURS RAPD.....	17
a - La PCR et les RAPD.....	17
b- Cartographie et identification de marqueurs spécifiques de régions chromosomiques.....	19
c- Analyse des populations.....	21
d- Contraintes de l'utilisation des RAPD.....	23
CHAPITRE III: ETUDE DE LA DIVERSITE GENETIQUE DE L'ESPECE.....	24
I- MATERIEL ET METHODES.....	24
1- Matériel végétal.....	24
2- Méthodes.....	26
a- Analyses morphologiques.....	26
b- Analyses moléculaires.....	27
Extraction de l'ADN total.....	27
Dosage de l'ADN.....	27
Réactions d'amplification.....	27
Migration et visualisation.....	27
Tri des amorces.....	28
Essais de répétabilité.....	28

Extraction de produits amplifiés.....	28
Marquage des fragments amplifiés.....	28
Préhybridations et hybridations .....	28
Collecte des données .....	29
Analyse des données .....	29
Répartition des produits d'amplification dans les populations et polymorphisme .....	29
Analyse factorielle des correspondances (AFC)...	30
Classification ascendante hiérarchique .....	30
Paramètres génétiques.....	31
Distances génétiques.....	33
<b>II - RESULTATS</b> .....	<b>34</b>
1-Diversité morphologique.....	34
a- Détermination de la taille minimum des échantillons .....	34
b- Analyses morphologiques.....	35
2-Diversité révélée par les marqueurs RAPD.....	37
a- Mises au point méthodologiques de la technique de RAPD ....	37
Effet de la qualité de l'ADN sur les produits d'amplification .....	37
Effet de la quantité d'ADN.....	38
Effet de la qualité et de la quantité de la Taq polymérase..	38
Effet de la quantité de Magnésium .....	38
Effet de la température d'hybridation.....	39
Répétabilité .....	39
b- Nature des fragments amplifiés.....	39
c- Fréquence des Bandes RAPD dans les populations .....	40
d- Organisation de la diversité génétique révélée par les RAPD...	41
Analyse factorielle des correspondances .....	41
Classification des génotypes.....	41
Classification sur tableau de distances.....	41
Classification à partir des coordonnées des individus sur l'AFC .....	42
Comparaison des différents types de regroupements.....	43
3- Comparaison de la diversité révélée par les caractères morphologiques et celle révélée par les RAPD.....	43
4- Comparaison RFLP/RAPD .....	44
5- Paramètres génétiques de la diversité révélée par les RFLP.....	54
a- Diversité globale .....	54

b- Diversité selon les origines géographiques.....	54
c- Diversité selon les groupes morphogéographiques.....	56
III- DISCUSSION.....	57
A- Diversité révélée par les RAPD.....	57
B- Paramètres génétiques de la diversité.....	60
IV - CONCLUSION.....	63
CHAPITRE IV: CARTOGRAPHIE GENETIQUE DU CACAOYER PAR LES MARQUEURS RAPD ET LOCALISATION DE QTLs.....	65
I - INTRODUCTION.....	65
II - CARTOGRAPHIE GENETIQUE PAR LES MARQUEURS RAPD .....	66
A- MATERIEL ET METHODES.....	66
1- Matériel Végétal.....	66
Clone utilisé comme parent femelle : UPA409.....	67
Clone utilisé comme parent mâle: POR.....	67
2- Réactions d'amplification.....	67
3- Analyses génétiques.....	67
B- RESULTATS.....	69
1- Déterminisme génétique des marqueurs RAPD.....	69
2- Niveau de polymorphisme .....	70
3- Analyses des ségrégations.....	70
2- Répétabilité.....	71
C- DISCUSSION.....	71
III- LOCALISATION DE QTL.....	73
A -MATERIEL ET METHODES.....	73
1- Matériel Végétal.....	73
Clone utilisé comme parent femelle : UPA 402.....	73
Clone utilisé comme parent mâle : UF676.....	73
2- Carte genetique .....	74
3- Caractères mesurées.....	74
Caractères morphologiques observés:.....	74
Caractères agronomiques .....	75
4- Collecte des données .....	75
5- Methodes de recherche de QTL.....	76
L'interval mapping ou cartographie par intervalle:.....	76
Méthode d'analyse de variance.....	76
La méthode des moindres carrés.....	77
B -RESULTATS .....	80
1- Caracteristiques morphologiques des deux parents UPA402 et UF676 .....	80

2- Correlations des caracteres.....	80
3- Localisation des QTL de la population de Bingerville.....	81
4- Localisation des QTL de la population de Divo.....	82
5- Localisation des QTL au sein de la population composite.....	82
C- DISCUSSION.....	83
IV- CONCLUSION.....	86
CONCLUSION GENERALE.....	89
RAPPEL DES PRINCIPAUX RESULTATS.....	89
PERSPECTIVES.....	91

## LISTE DES TABLEAUX

**Tableau 1:** Répartition de la production mondiale de cacao en 1993

**Tableau 2:** Importance économique du cacao dans les pays africains

**Tableau 3:** Liste de clones utilisés dans les différentes analyses de la diversité génétique

**Tableau 4:** Nombre de géotypes dans les différentes analyses

**Tableau 5:** Effectifs de cabosses nécessaires à l'évaluation des caractères des cabosses

**Tableau 6:** Effectifs de fèves nécessaires à l'évaluation des caractères des fèves

**Tableau 7:** Analyses de variance des caractères morphologiques

**Tableau 8:** Comparaison des moyennes des caractères morphologiques par le test de la plus petite différence significative

**Tableau 9:** Fréquences des bandes dans les populations et tests de Chi-deux

**Tableau 10:** Comparaison des fragments RAPD dans les populations

**Tableau 11:** Fréquences alléliques au sein de deux échantillons de cacaoyers, l'un utilisé pour les analyses RFLP et l'autre pour les analyses RAPD

**Tableau 12:** Fréquences alléliques et indices de diversité par locus au sein des populations Criollo et Forastero

**Tableau 13:** Valeurs moyennes des paramètres de génétique des populations des Forastero et des Criollo

**Tableau 14:** Valeurs des paramètres F par locus au sein des Forastero et des Criollo

**Tableau 15:** Valeurs moyennes des paramètres de génétiques des populations des Forastero Haut Amazoniens(HA), des Forastero Bas Amazoniens(BA), des Trinitario(TR) et des Criollo:

**Tableau 16:** Fréquences alléliques et indices de diversité par locus au sein des quatre population, Forastero Haut Amazoniens(HA), des Forastero Bas Amazoniens(BA), des Trinitario(TR) et des Criollo:

**Tableau 17:** Valeurs des Fis par locus au sein des Forastero Haut Amazoniens(HA), des Forastero Bas Amazoniens(BA), des Trinitario(TR) et des Criollo:

**Tableau 18:** Valeurs des  $F_{st}$  entre les populations prises deux à deux

**Tableau 19:** Comparaison des quatre populations prises deux à deux, au niveau des déséquilibres(HW), des paramètres  $F_{st}$ , et de la distance génétique .

**Tableau 20:** Croisements induisant du polymorphisme entre les deux parents

**Tableau 21:** Ségrégation de marqueurs RAPD et tests de  $\chi^2$

**Tableau 22:** Valeurs des parents pour divers caractères morphologiques à Bingerville

**Tableau 23:** Valeurs des parents pour divers caractères morphologiques à Divo

**Tableau 24:** Valeurs des coefficients de corrélations à Bingerville

**Tableau 25:** Valeurs des coefficients de corrélations à Divo

**Tableau 26 :** Caractéristiques des QTL révélés à Bingerville par Interval mapping

**Tableau 27:** Caractéristiques des QTL révélés à Bingerville par analyse de variance

**Tableau 28:** Caractéristiques des QTL révélés à Divo par Interval mapping

**Tableau 29:** Caractéristiques des QTL révélés à Divo par analyse de variance

**Tableau 30:** Caractéristiques des QTL révélés à Bingerville et à Divo par Interval mapping

**Tableau 31:** Caractéristiques des QTL révélés à Bingerville et à Divo par analyse de variance

## LISTE DES FIGURES

**Figure 1:** Courbes de l'évolution de l'offre et de la demande

**Figure 2 :** Plan 1 de l'ACP sur les caractères morphologiques

**Figure 3:** Fréquence des bandes RAPD dans les populations de cacaoyers

**Figure 4:** Plan 1 de l'AFC à partir des bandes RAPD

**Figure 5 a et b:** Plans 1 des AFC à partir des séquences uniques RAPD et des séquences répétées RAPD

**Figure 6:** Phénogramme à partir des indices de Czekanowski

**Figure 7:** Phénogramme à partir des indices de Rogers et Tanimoto

**Figure 8:** Phénogramme à partir des coordonnées de 7 axes factoriels sur les données RAPD

**Figure 9a:** Plan 1 de l'AFC faite à partir des données RFLP/cDNA

**Figure 9b:** Dendrogramme à partir des distances génétiques de Nei

**Figure 10:** Phénogramme des données cDNA

**Figure 11:** Trois groupes de liaison obtenus avec un LODscore de trois

**Figure 12:** Localisation des QTL détectés à Bingerville par interval mapping

**Figure 13:** Localisation des QTL détectés à Divo par interval mapping

**Figure 13:** Localisation des QTL détectés à Bingerville et à Divo par interval mapping

## LISTE DES PHOTOS

**Photo 1:** Cabosse de cacaoyer

**Photo 2:** Arbre de cacaoyer en fleurs

**Photo 3:** Extraits d'ADN

**Photo 4:** Profil simple et profil complexe

**Photo 5:** trois types de profils obtenus par hybridation des bandes RAPD sur de l'ADN total

**Photo 6:** Ségrégation de bandes RAPD dans la descendance UPA409 x POR

## ANNEXES

**Annexe 1 :** protocole d'extraction d'ADN total

**Annexe 2:** Réactions d'amplification

**Annexe 3:** protocole d'extraction d'ADN total pour analyses RFLP

**Annexe 4:** Préhybridation et hybridation

**Annexe 5:** Digestion enzymatique et transfert sur membrane

**Annexe 6:** Récupération de sonde

**Annexe 7:** Caractéristiques des cabosses

**Annexe 8:** Caractéristiques des fèves

**Annexe 9:** AFC avec noms codés et correspondances

**Annexe 10:** Nombre de locus hétérozygotes au sein des génotypes étudiés par RFLP

**Annexe 11:** Histogrammes de distribution des caractères à Bingerville

**Annexe 12:** Histogrammes de distribution des caractères à Divo

## INTRODUCTION

Le cacaoyer est une plante ligneuse des pays tropicaux, originaire d'Amérique du Sud dont la culture est faite en Afrique, en Asie et en Amérique. Le fruit du cacaoyer, la cabosse, contient des graines appelées fèves qui lorsqu'elles sont fermentées et séchées servent à la fabrication du chocolat.

Le cacaoyer est une culture d'exportation, source importante de devises pour certains pays d'Afrique, dont le produit marchand fait l'objet d'un important commerce international.

Pour faire face à la demande en cacao des consommateurs et en matériel végétal amélioré pour les paysans, des programmes de sélection visant à l'amélioration de la productivité, de la résistance aux aléas biotiques et abiotiques et des qualités organoleptiques sont en cours dans différents instituts de recherche. L'espèce *Theobroma cacao* est très polymorphe et la création d'hybrides ou de descendances intergroupes est la principale voie d'amélioration génétique du cacaoyer utilisée dans le monde. Cependant, cette approche se heurte d'une part à la longueur des cycles de sélection et d'autre part à une mauvaise connaissance des clones: une confusion dans la classification et une insuffisance d'informations sur le matériel végétal à améliorer. Aussi, depuis des décennies, un nombre réduit de clones, souvent les mêmes partout dans le monde, sont utilisés dans les schémas d'amélioration. En outre, la menace de disparition de la forêt tropicale et la nécessité de mise en collection d'un grand nombre de génotypes exigent une gestion plus efficace des ressources génétiques du cacaoyer par une évaluation rapide de la diversité génétique des clones et leur classification.

Bien que tous les auteurs s'accordent dorénavant sur l'existence d'une seule espèce de cacaoyers, il n'y a pas de consensus quant aux subdivisions au sein de cette espèce. Cuatrecasas (1964) reconnaît l'existence de deux sous espèces alors que Soria (1970) parle de trois complexes génétiques qui correspondraient à des variétés. Des problèmes d'identification et de classification des génotypes dans les différents groupes se posent souvent (Cas des cacaoyers Nacional d'Equateur, des cacaoyers guyanais) et montrent que jusqu'à présent aucun outil d'identification et de classification n'a donné une entière satisfaction. Les caractères morphologiques, efficaces pour les regroupements sommaires ne permettent pas une identification variétale précise comme l'a reconnu Engels (1986). De même, les marqueurs isozymes, à cause du faible nombre de locus échantillonnés ne permettent pas une discrimination fine des divers groupes génétiques et une identification variétale complète. Plusieurs types de marqueurs RFLP ont été utilisés pour étudier la diversité génétique du cacaoyer et ont fourni

des informations importantes sur l'organisation de la diversité. Mais la technique des RFLP est lourde à mettre en oeuvre. Les différentes approches bien souvent se complètent et montrent qu'aucune d'elles, prise isolément n'est entièrement satisfaisante. Dans un tel contexte de complémentarité d'information, il apparaît utile de développer plusieurs outils afin de mieux appréhender l'organisation de l'espèce. Au sein de cet ensemble d'outils, les techniques RFLP seraient à privilégier mais les coûts des investissements sont élevés et l'acquisition de données est longue et difficile. Au côté des marqueurs RFLP, les marqueurs RAPD (Random Amplified Polymorphic DNA) utilisés chez certaines plantes (Williams *et al.* 1990) ont permis de préciser les classifications. Correspondant à différents types de séquences dans le génome, les RAPD révèlent souvent plus de polymorphisme que les RFLP et sont attrayants du fait de la simplicité et de la rapidité des analyses.

Dans ce travail, nous avons évalué dans quelle mesure les RAPD pouvaient être utiles pour la gestion des ressources génétiques du cacaoyer et son amélioration: pour cela, nous avons réalisé une étude de diversité génétique à l'aide des RAPD puis en avons comparé les résultats avec ceux obtenus par d'autres types de marqueurs. Nous avons complété les études descriptives de la diversité révélée par les RFLP par une estimation de divers paramètres de la structure génétique des populations de cacaoyer. Dans une deuxième partie, nous avons abordé les problèmes de cartographie génétique du cacaoyer d'une part en testant les potentialités des marqueurs RAPD pour la construction d'une carte génétique et d'autre part en localisant des locus impliqués dans l'expression de caractères quantitatifs importants du cacaoyer à l'aide d'une carte génétique intégrant plusieurs types de marqueurs construite parallèlement au laboratoire (Lanaud *et al.* 1993). Cet aspect du travail ouvre la voie à une sélection précoce de génotypes basée sur les corrélations des marqueurs moléculaires avec les caractères quantitatifs intéressants. De plus, il permet d'envisager une sélection assistée par marqueurs par le contrôle précoce des recombinaisons au cours des cycles de sélection et une réduction de la longueur des cycles de sélection.

<b>AFRIQUE</b>	Côte d'Ivoire	Ghana
<b>1277600T</b>	697000	312100
<b>53,76%</b>	29,33%	13,13%
<b>AMERIQUE</b>	Brésil	Equateur
<b>576400</b>	308600	67400
<b>24,26%</b>	12,98%	2,83%
<b>ASIE</b>	Malaisie	Indonésie
<b>522200</b>	225 000 T	240000T
<b>21,98%</b>	9,4%	10,10%
<b>total mondial</b>	<b>2 376200T</b>	

**Tableau 1: Repartition de la production mondiale en 1993**

Pays	Côte d'Ivoire	Ghana	Cameroun	Nigeria
% des exportations	43%	59 %	15 %	2,2%
Familles de planteurs	700 000	500 000	260 000	150 000

**Tableau 2: Importance economique du cacao dans les pays africains**

# CHAPITRE I: DONNEES GENERALES SUR LE CACAOYER

## I- IMPORTANCE ECONOMIQUE

L'utilisation du cacaoyer comme plante alimentaire a commencé en Amérique centrale et australe au temps des Mayas qui le considéraient comme une plante des Dieux, dont ils faisaient la culture bien avant l'arrivée des Européens (Braudeau, 1969; Dand, 1993). Déjà à cette époque, le cacaoyer avait une très grande importance puisque les fèves étaient utilisées comme monnaie d'échanges dans des transactions commerciales.

Les européens introduisirent le cacaoyer à partir du 16<sup>ème</sup> siècle dans les différentes îles de leurs possessions qui dès le 17<sup>ème</sup> siècle commencèrent à exporter le cacao vers les métropoles. Beaucoup plus tard au 19<sup>ème</sup> siècle, le cacaoyer fut introduit en Afrique insulaire (Sao Tome et Principe) puis en Afrique continentale, au Ghana, au Nigeria, au Cameroun et en Côte d'Ivoire; et l'exportation de fèves vers les pays consommateurs commença dès le début du 20<sup>ème</sup> siècle.

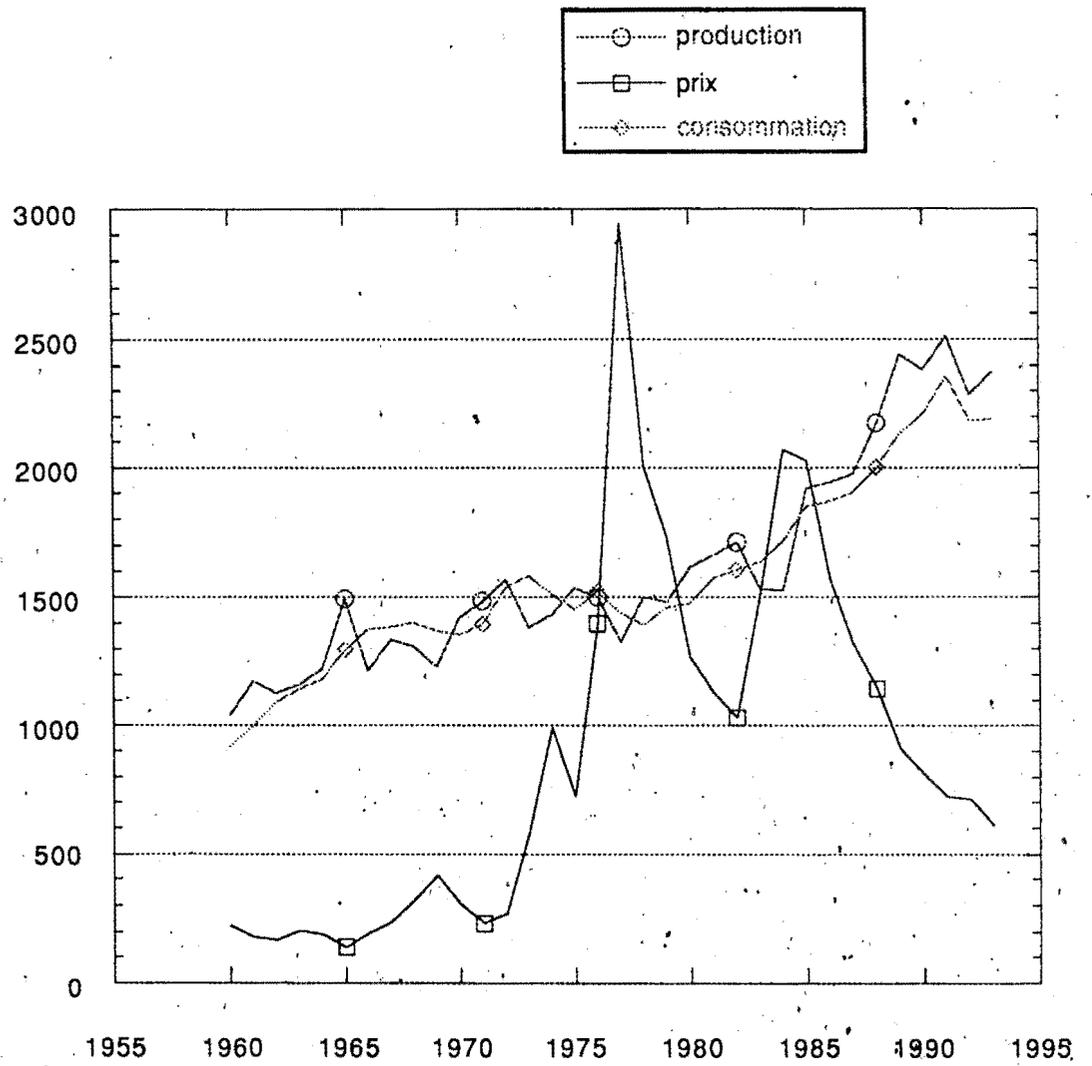
L'introduction du cacaoyer en Asie date du 16<sup>ème</sup> siècle (Indonésie) mais sa culture n'a pris de l'importance que dans la seconde moitié du 20<sup>ème</sup> siècle.

A partir de ces introductions, trois grandes régions deviennent productrices de cacao: l'Afrique, l'Amérique et l'Asie.

La production mondiale de cacao a connu une croissance impressionnante. En 1900, elle n'était que d'environ 125 000 tonnes de cacao marchand provenant presque totalement d'Amérique: la production africaine était de 20 000 tonnes soit 16% de la production mondiale (Dand, 1993). En 1993, la production mondiale passe à 2376200 tonnes soit une croissance moyenne annuelle de 24 206 tonnes dont la répartition est de 53,76% pour l'Afrique, 24,26 % pour l'Amérique et 21,98 % pour l'Asie (Tableau 1).

Le cacao est une source importante de devises pour les pays Africains producteurs. En 1987, le F.M.I. évaluait cette importance par rapport au reste des exportations de certains pays et au nombre de familles impliquées (Jouve *et al*, 1990) comme indiqué dans le tableau 2.

Le cacao est consommé essentiellement dans les pays tempérés, exception faite du Brésil. Deux tiers de la production sont destinés à la fabrication du chocolat et 1/3 est transformé en poudre utilisée dans les industries alimentaires.



**Figure 1:** Evolution de la production de cacao, de la consommation et des prix

Les pays consommateurs sont les Etats Unis, l'Allemagne, les pays Bas, le Royaume Uni, etc.

Pendant longtemps, la consommation équivalait à peu près à la production mais de temps à autre, la demande a dépassé l'offre, conduisant à des flambées de prix comme ce fut le cas entre 1975 et 1980. Cette situation a provoqué la création de nouvelles plantations à travers les zones de production. A partir de 1985, une offre très supérieure à la demande a provoqué une chute des cours du cacao. La croissance de la production dans des pays comme la Côte d'Ivoire et l'entrée dans la production des pays d'Asie, comme l'Indonésie et la Malaisie sont responsables de cette situation. Mais le marché du cacao est cyclique comme le décrit Ruf (1993) et la crise actuelle se résorbera probablement. Les courbes de la figure 1 montrent l'évolution des prix de la tonne de cacao marchand ainsi que celle de l'offre et de la demande au cours des années.

En Afrique, la culture du cacao est faite par de petits planteurs dont la taille des exploitations est rarement supérieure à 10 ha. En Côte d'Ivoire, 67 % des exploitations ont moins de 5 ha et la main d'oeuvre y est de type familial. La mise en oeuvre de nouvelles techniques est plus difficile avec ce type d'exploitant dont le niveau technique est très faible. En Asie ou en Amérique, les grandes exploitations sont fréquentes; 80 % des exploitations en Malaisie et 55% au Brésil ont une taille supérieure à 450 ha.

## II- MORPHOLOGIE ET BIOLOGIE

### 1- TAXONOMIE

Le cacaoyer est une plante pérenne qui se rencontre à l'état naturel dans les étages inférieurs des forêts tropicales humides d'Amérique. Il appartient à la famille des *Sterculiacées* et au genre *Theobroma* dont Cuatrecasas (1964) a décrit 22 espèces. Le cacaoyer cultivé commercialement pour le chocolat appartient à l'espèce *T. cacao* qui a été subdivisée en trois grands groupes: les Forastero, les Criollo et les Trinitario. D'autres espèces du genre *Theobroma* sont exploitées localement pour la fabrication de boissons rafraîchissantes et de chocolat: c'est le cas de *T. bicolor*, *T. grandiflorum*, *T. speciosum* (Braudeau, 1969).

### 2- MORPHOLOGIE.

A l'état sauvage, le cacaoyer est un arbre de plusieurs mètres de hauteur comportant plusieurs étages et qui émet parfois des racines aériennes. En plantation, le semenceau émet un axe orthotrope qui croît de manière continue jusqu'à 18-24 mois, puis émet une couronne de 3 à 5 branches plagiotropes dont la croissance est quant à elle indéfinie. Plus tard, sous la couronne et à la base de la tige principale partent d'autres axes orthotropes encore appelés gourmands qui

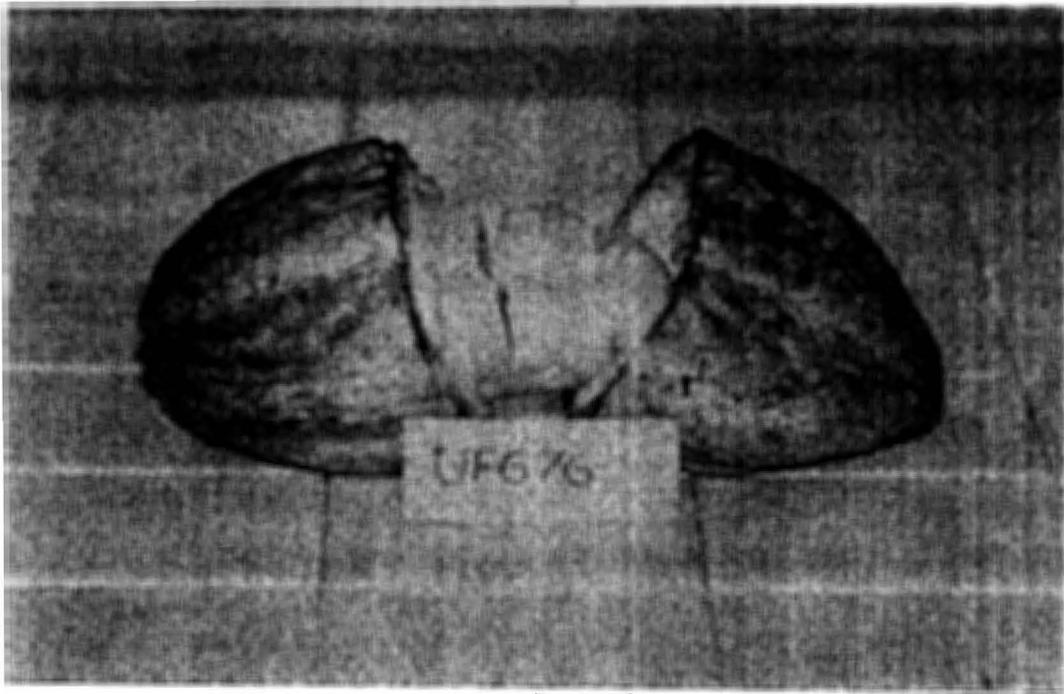


sont éliminés au fur et à mesure de leur apparition. Le cacaoyer adulte en culture n'atteint que 5 à 7 m de hauteur.

### 3- BIOLOGIE

Le cacaoyer est diploïde avec  $2n=20$  chromosomes (Carletto, 1946; Glicenstein et Fritz, 1989). Les chromosomes sont de petite taille (Martinson, 1975) et la taille du génome haploïde est d'environ 0.4 pg (Lanaud *et al.*, 1992; Figueira *et al.*, 1992). Le cacaoyer est une plante essentiellement allogame. Cependant, certaines variétés comme les Amelonado sont autogames, d'autres génotypes appartenant aux Forastero Haut Amazoniens sont préférentiellement allogames. Le cacaoyer possède un système d'autoincompatibilité se traduisant par la chute des fleurs pollinisées avec du pollen incompatible. Ce système d'incompatibilité a été étudié par Knight et Rogers en 1955, et par Cope en 1962. Il apparaît qu'il s'agit d'un système complexe gaméto-sporohytique qui se manifeste avant et après la formation du zygote. Il est déterminé par un groupe d'allèles S qui entretiennent des relations de dominance ou de codominance aboutissant à la fusion ou à la non fusion des gamètes. De plus, bien que la fusion aie eu lieu dans une partie de l'ovaire, la plante mère éliminerait les fruits issus d'autofécondation (Lanaud *et al.*, 1987). La présence de ce caractère d'autoincompatibilité chez certains clones a permis de mettre en place dans certains pays des champs semenciers biconaux pour la production à grande échelle d'hybrides intergroupes en évitant ainsi les contraintes des pollinisations artificielles. Cependant, il existe une gradation de l'incompatibilité et ce schéma de production d'hybrides est remis en question dans de nombreux pays notamment en Côte d'Ivoire à partir des travaux de Lanaud *et al.* (1987). En effet, alors que certains clones sont autoincompatibles quelles que soient les conditions, l'incompatibilité de certains clones peut être levée par apport d'allopollen et d'autopollen en mélange. C'est l'association sporophyte femelle, autopollen, allopollen qui détermine la réaction d'incompatibilité (Lanaud, 1987).

La floraison du cacaoyer est très abondante et intervient sur le tronc et les branches principales de l'arbre à partir de 18 mois après le semis de la graine pour les variétés les plus précoces. Les périodes de floraison sont variables d'une variété à l'autre et dépendent aussi des conditions climatiques. Les Amelonado ouest Africains ont deux pics de floraison dans l'année alors que les Forastero Haut Amazoniens et leurs hybrides fleurissent tout au long de l'année. Les fleurs sont petites et hermaphrodites, de type 5 et le gynécée est composé d'un ovaire supère surmonté d'un style terminé par cinq stigmates. L'ovaire comprend cinq loges contenant chacune une dizaine d'ovules. L'androcée est composé de cinq étamines terminées par 4 anthères protégées par une cuculle. Les étamines alternent avec les pièces stériles que sont les staminodes. La fleur s'ouvre le



**Photo 1: Une cabosse de cacaoyer**



**Photo 2: Un cacaoyer en fleurs**

matin, et le pollen est immédiatement fonctionnel mais sa durée de vie varie selon les saisons entre un et trois jours (Lachenaud, 1991).

Le fruit, appelé chérelle jusqu'à deux mois et cabosse ensuite est composé d'une coque comprenant trois parties: l'endocarpe, le mésocarpe et l'épicarpe. La cavité intérieure du fruit est occupée par les graines ou fèves au nombre d'environ 50 enrobées d'une substance mucilagineuse et portées par un rachis central (Photo 1). Après 150 à 200 jours de développement, les fruits arrivent à maturité. La durée de maturation des fruits dépend des conditions climatiques mais surtout des clones: les clones ayant de grosses cabosses comme certains Trinitario (ICS) ont en moyenne 210 jours de maturation, alors que les Forastero Haut Amazoniens avec leurs petites cabosses ont 165 jours de maturation (Cilas, 1992). Dès la maturité du fruit, les graines sont prêtes à germer. Il n'existe pas de dormance chez le cacaoyer et la graine perd rapidement son pouvoir germinatif surtout si elle est extraite de la cabosse. Les fèves à l'intérieur d'une cabosse non ouverte peuvent conserver leur pouvoir germinatif pendant plusieurs jours (Mossu, 1974). Le nombre de fruits matures produits par le cacaoyer est très faible comparé au nombre de fleurs (Photo 2). Des expériences réalisées par Lachenaud (1991) dans des parcelles différentes en condition de pollinisations naturelles ont révélé un rendement en cabosses de 0,70% à 4% environ. Certains avancent l'hypothèse d'une sous pollinisation puisque seuls 10 % des fleurs recevraient plus de cinq grains de pollen (De Reffye *et al.*, 1978), d'autres estiment qu'il s'agit de problèmes nutritionnels étant donné que même suffisamment pollinisées, les fleurs peuvent chuter (Lachenaud, 1991).

La pollinisation du cacaoyer est entomophile et est assurée par les insectes de tous genres allant des *Forcypomya*, aux thrips en passant par les fourmis (Paulin *et al.* 1983; Lachenaud, 1991). Pour augmenter le nombre de nouaisons et de cabosses produites annuellement dans les champs semenciers certains pays ont recours à des pollinisations manuelles afin de contourner le défaut de pollinisation signalé par certains auteurs.

#### 4- AGRONOMIE

Le cacaoyer se rencontre dans une zone comprise entre 15 ° de latitude au Sud de l'Equateur et 20 ° au Nord de celui ci (Soria, 1970). Il a besoin pour se développer d'un sol profond et riche et de beaucoup d'eau. La plante issue d'une graine émet une racine pivotante de longueur variable selon le type de sol tandis que celle issue d'une bouture a un système racinaire fasciculée qui ne lui permet pas de lutter efficacement contre la sécheresse. Au jeune âge et jusqu'à la fermeture des frondaisons qui survient trois ans après la plantation, le cacaoyer a besoin d'un ombrage temporaire; des bananiers ou des cocotiers sont utilisés selon les pays. La culture sans ombrage est préconisée pour l'obtention de bons

rendements mais cette pratique peut être préjudiciable lorsqu'il y a un manque d'entretien phytosanitaire. Dans certains pays comme le Ghana, la culture du cacaoyer est faite sous ombrage, et ceci réduirait l'incidence des dégâts causés par les insectes.

Des ennemis naturels du cacaoyer provoquent de nombreux dégâts dans les plantations. En jeune plantation, le cortège phytosanitaire est composé de psylles, de thrips, de chenilles défoliatrices et autres insectes qui détruisent le bourgeon terminal et retardent ainsi la croissance de la plante. En plantation adulte et selon les régions, il s'agit de maladies cryptogamiques, virales et d'insectes. Les différentes espèces du *Phytophthora* sont responsables de la pourriture brune des cabosses, qui au Cameroun engendrent 50 % de pertes minimales de production (Blaha et Lotodé, 1976; Babacauh et Partiot, 1976). La moniliose due à *Moniliophthora roreri* provoque aussi des pourritures sur les cabosses en Amérique et la maladie du balai de sorcière due à *Crinepellis pernicioso* attaquent les parties aériennes du cacaoyer en Amérique. Les insectes ravageurs sont les mirides du cacaoyer de type *Sahlbergella*, *Helopeltis* et *Monalonion* qui par leurs piqûres provoquent des nécroses sur les cabosses et les branches et des dessèchements des jeunes pousses. Récemment des différences de sensibilité ont été trouvées entre clones, les cacaoyers des groupes Criollo et Trinitario étant les plus sensibles aux attaques de mirides (N'goran *et al.*, 1992; Sounigo *et al.*, 1993). Le Cocoa Swollen Shoot Virus est la plus redoutable des maladies. Entre 1946 et 1967, l'arrachage de plusieurs millions de cacaoyers au Ghana a été nécessaire pour éviter l'expansion de la maladie (Braudeau, 1969). Les symptômes se traduisent par des gonflements des tiges et l'apparition de mosaïques sur les feuilles. En Asie, ce sont le VSD (Vascular Streak Dieback disease) et le Pod Borer qui provoquent des dégâts dans les plantations.

### III- LES POPULATIONS DE CACAOYERS

Très tôt, les cacaoyers ont été repartis en deux grands groupes selon la couleur, la forme et la taille des cabosses : les **Criollo** aux cabosses rouges ou vertes et les **Forastero** aux cabosses vertes. Considérés au départ selon les auteurs comme deux espèces différentes (Morris 1882, Pittier 1930), comme la même espèce (Cheesman, 1944) ou comme deux sous-espèces (Cuatrecasas, 1964), les cacaoyers font dorénavant partie d'une seule espèce comportant des formes sauvages et cultivées toutes interfertiles.

De nombreuses descriptions de cacaoyers ont été faites et la monographie de Cuatrecasas (1964) a fait le point de ce qui était connu alors. Les travaux conduits à Trinidad et les différentes prospections effectuées ont permis de

préciser la classification des cacaoyers en subdivisant les Forastero en Forastero Amazoniens et Trinitario et les Criollo en Criollo d'Amérique Centrale et d'Amérique du Sud (Cheesman, 1944). Cependant les auteurs reconnaissent qu'il est difficile de savoir les limites d'un groupe en se basant sur les caractères morphologiques. En 1970, Soria décrit les cacaoyers rencontrés en Amérique tropicale en considérant trois grands groupes qui sont les Forastero, les Criollo et les Trinitario. Les Forastero sont communément subdivisés selon leurs lieux d'origine en Forastero Bas Amazoniens et Forastero Haut Amazoniens par certains (Toxopeus, 1972). De nos jours, les groupes de cacaoyers reconnus sont les **Criollo**, les **Forastero Haut** et **Forastero Bas Amazoniens** et les **Trinitario**.

### **1- LES CACAOYERS CRIOLLO**

Ce sont les cacaoyers cultivés depuis la période précolombienne en Amérique centrale (Mexique et Nicaragua) ou au Venezuela par les Indiens (Reyes, 1992) à partir de semences récupérées dans la forêt. Leur aire d'origine est l'Amérique centrale et le nord de l'Amérique du Sud. Au sein des Criollo, Soria (1970) recense plusieurs variétés qui ne sont pas toujours des variétés de Criollo pures comme le fait remarquer Lockwood et End (1992). Ce sont: Lagarto, Criollo de Nicaragua ou du Venezuela, Criollo de Colombie, Criollo du Mexique, Porcelana. Toutes ces variétés produisent un cacao apprécié, jugé fin. Les caractères morphologiques et agronomiques marquants des Criollo sont les suivants:

- faible vigueur des arbres
- sensibilité à la majorité des maladies et insectes
- cabosses verruqueuses
- prépondérance de coloration rouge des cabosses mais présence aussi de cabosses vertes dans certaines populations.

Du fait de leur faible vigueur, les Criollo sont très peu cultivés dans le monde et sont de plus en plus remplacés par des Forastero ou des Trinitario. Des efforts de prospections dans les anciennes plantations de cacaoyers Criollo sont faits par différents pays d'Amérique Centrale et du Sud (Reyes, 1992, Lockwood et End, 1992) afin d'éviter la disparition de ces cacaoyers classés comme la première priorité en matière de conservation par l'IBPGR (Kennedy *et al.*, 1987) à cause de la disparition rapide des forêts. La collection internationale du CATIE au Costa Rica regroupe une centaine de génotypes Criollo.

### **2- LES CACAOYERS FORASTERO**

Ils constituent la majorité des cacaoyers cultivés dans le monde. Ils se rencontrent naturellement dans le bassin amazonien, le long de l'Orénoque et en Guyane. Ils sont souvent classés en Forastero Haut Amazoniens et Forastero Bas Amazoniens qui intègrent aussi par abus de langage les Forastero de l'Orénoque

et de Guyane. La collection internationale ICG (International Cocoa Genebank) à Trinidad s'occupe de leur conservation et de leur caractérisation.

### **LES FORASTERO HAUT AMAZONIENS**

Ils se rencontrent dans les bassins de la Haute Amazonie et ont été décrits pour la première fois par Pound en 1938 et en 1945 lors de prospections au Pérou et en Equateur visant à trouver des cacaoyers résistants à la maladie du " balai de sorcière". Du fait de nombreuses qualités rencontrées au sein de ces populations, plusieurs autres prospections ont été réalisées dans cette région (Allen et Lass, 1983; Chalmers, 1968-1973). Les Forastero Haut Amazoniens sont à la base de nombreux hybrides à travers le monde. Ils présentent une grande diversité et leurs caractéristiques principales sont:

- des cabosses allongées vertes dont la surface n'est pas toujours lisse
- des fèves violettes , plates et de taille variable: certains géotypes comme SCA6 ont de petites fèves d'autres comme MO9 sont caractérisés par de grosses fèves.
- des arbres précoces, vigoureux et productifs.
- des tolérance à certaines maladies (N'goran *et al.*, 1992).

### **LES FORASTERO BAS AMAZONIENS**

Ils se rencontrent en Basse Amazonie, le long de l'Orénoque et en Guyane. Ils constituent la majorité des cacaoyers cultivés au Brésil et en Afrique de l'Ouest. Ils sont composés de plusieurs sous populations. Les variétés décrites par Braudeau (1969) et Soria (1970) sont: "para", "maranhao", "comun", "catongo", "almeida", "matina", et l'"Amelonado" d'Afrique de l'Ouest. Certaines sous populations dérivent de populations préexistantes: les "catongo" et "almeida" sont apparus à la faveur d'une mutation au sein des sous populations "comun" et "para".

Les Forastero Bas Amazoniens ont généralement des cabosses arrondies de petite taille, presque lisse et de couleur verte. En dehors des Catongo et des Almeida qui ont des fèves blanches, les fèves sont aplaties et de couleur violette. Les arbres sont généralement moins vigoureux et plus tardifs que ceux des Forastero Haut Amazoniens mais comportent des formes très productives (clone IFC5) et des formes peu sensibles aux insectes: clone IFC2 de Côte d'Ivoire (N'goran *et al.*, 1992). Considérés au départ comme très homogènes, les Forastero Bas Amazoniens apparaissent dorénavant comme variables au vu de la diversité rencontrée lors de prospections au Vénézuéla (Lanaud *et al.*, 1986) et en Guyane Française (Sallée 1986, Lachenaud 1992) ainsi qu'au Brésil (Pires *et al.* 1992).

### 3- LES CACAOYERS TRINITARIO

Ce sont des cacaoyers qui proviendraient d'hybridations naturelles survenues à Trinidad entre les Criollo et les Forastero ou encore auraient été introduits à Trinidad à partir de l'estuaire de l'Orénoque (Toxopeus, 1972) . Ils présentent à la fois les caractères des Criollo et des Forastero et de ce fait sont très variables. On y rencontre toutes les combinaisons de forme et de couleur possibles. Ces hybrides du fait de leur vigueur et de leurs qualités technologiques ont de nos jours remplacé les Criollo dans presque toutes les plantations d'Amérique centrale et sont prépondérants en Colombie, en Equateur, à Trinidad, en Papouasie Nouvelle Guinée et au Cameroun .

## IV- ORGANISATION DE LA DIVERSITE GENETIQUE DU CACAOYER

### 1-DIVERSITÉ MORPHOLOGIQUE

De nombreux caractères portant sur les cabosses, les fèves et les fleurs servent à identifier des cacaoyers. Enriquez et Soria (1968) ont étudié la longueur, la largeur et le poids des fèves pelées. Ils ont montré que ces caractères pouvaient être utilisés comme descripteurs des clones et ont défini une valeur discriminante estimée par le rapport de la variance interclonale sur la variance intraclonale. Plus cette valeur est élevée et plus le caractère a une valeur descriptive importante. Engels (1983) a utilisé des descripteurs qualitatifs et quantitatifs des fleurs, des cabosses, des fèves et des feuilles en définissant un pouvoir discriminatif pour chaque caractère correspondant au rapport du nombre de différences significatives observées sur le nombre de comparaisons possibles. La valeur de ce paramètre est élevée lorsque la variation interclonale est grande. Selon Engels (1985), les caractères les plus descriptifs des fèves sont par ordre d'importance, le nombre de fèves par cabosse, la longueur de la fève, le poids et l'épaisseur de la fève alors qu'au niveau des cabosses, les caractères les plus descriptifs sont l'épaisseur, le diamètre et la longueur de la cabosse. Engels (1986) a étudié la diversité morphologique de 294 clones issus de toutes les origines. Il a défini 39 descripteurs qui lui ont permis de distinguer les deux groupes originaux de cacaoyers (Criollo et Forastero). Cependant les difficultés de codage et d'observation l'ont conduit à réduire le nombre de descripteurs à 10 pour l'identification et la classification des clones. Ces descripteurs sont actuellement utilisés pour l'évaluation morphologique des clones de la collection internationale de Trinidad.

Le déterminisme génétique de la majorité des descripteurs morphologiques reste inconnu. Le marqueur "axil spot" (Harland et Frechville, 1927) est un de ceux dont le déterminisme a complètement été élucidé. Il est responsable de l'apparition de tâches à la base des pétioles des premières feuilles et a beaucoup été utilisé pour le repérage de certains génotypes. Il est présent à l'état homozygote chez le clone K5 et est gouverné par deux gènes dominants. La taille des cabosses représentée par le poids, la longueur et le diamètre a été étudiée par Soria *et al.* (1974). Une héritabilité de 55 % pour la longueur, de 57 % pour le poids et de 63 % pour le diamètre a été trouvée. Glendinning (1963) a montré sur trois descendance F2 que la taille des fèves était constante sur un arbre et positivement corrélée à la taille des cabosses. Des gènes majeurs interviendraient mais leur nombre et leur mode d'action restent inconnus. Dans des croisements faisant intervenir des variétés à fèves blanches et des variétés à fèves violettes, la descendance contient un mélange de fèves blanche et de fèves violettes, avec cependant un faible nombre de fèves blanches. Le caractère "fève blanche" serait donc récessif (Wellensiek, 1931). Le nombre d'ovules par ovaire est considéré comme l'un des plus descriptifs: variable entre clones mais stable au sein d'un clone (Engels 1983; Pound, 1933; Jacob et Atanda, 1973; Soria et Enriquez, 1981). Des travaux récents (Lachenaud, 1991) montrent cependant que le nombre d'ovules par ovaire est variable selon les conditions environnementales. Soria et Esquivel (1970) estiment que le nombre d'ovules par ovaire est un caractère dominant. Les autres caractéristiques des fleurs que sont la longueur et la largeur de l'ovaire, la longueur du style, la longueur et la largeur des sépales et la longueur des staminodes ont été étudiées par Raboin *et al.* (1993). Des effets d'aptitude générale à la combinaison ont été mis en évidence entre les croisements. Ces caractères ont une forte héritabilité, exception faite de la largeur de l'ovaire. Les qualités organoleptiques du cacao marchand ont été étudiées par plusieurs auteurs. Clément *et al.* (1991), ont montré qu'il n'y avait pas de différence entre les génotypes qu'ils ont analysés pour l'amertume et l'astringence du cacao, alors que la flaveur est liée au génotype. De même, Clapperton *et al.* (1991) ont montré l'existence d'un effet de l'origine du matériel sur la flaveur même si le traitement post-récolte de la fève est déterminant.

## 2-DIVERSITÉ ENZYMATIQUE

Dix gènes provenant de six systèmes enzymatiques ont été identifiés et utilisés pour étudier la diversité génétique de l'espèce représentée par 350 génotypes (Lanaud, 1987). Les résultats obtenus montrent que la majorité des allèles sont représentés dans la population de Haute Amazonie qui est considérée comme le centre de diversité primaire du cacaoyer. Deux allèles (MDHA/1 et PAC 1) spécifiques aux Criollo et Trinitario ont été observés. D'autres allèles sont rares

dans certaines populations. Les études enzymatiques ont permis de révéler que des centres de diversifications secondaires ont pu exister à côté du centre de diversification primaire que constitue la région de Haute Amazonie.

Le développement de cet outil d'analyse a contribué à révéler des problèmes au niveau de la production des semences hybrides. Il est utilisé pour vérifier la légitimité des semences. Intéressants sur bien des aspects, les marqueurs enzymatiques sont cependant en nombre très limité chez le cacaoyer et n'ont pas permis de différencier sans équivoque les différentes populations qui se recoupent partiellement.

### 3- DIVERSITÉ RÉVÉLÉE PAR LES MARQUEURS MOLÉCULAIRES

Plus importants numériquement que les isozymes, les marqueurs moléculaires ont été utilisés pour étudier la diversité du cacaoyer au niveau de l'ADN nucléaire et cytoplasmique. Plusieurs types de marqueurs ont servi à structurer la variabilité: des marqueurs RFLP pour lesquels des sondes cDNA, des sondes rDNA, des sondes mitochondriales et des sondes chloroplastiques ainsi que des marqueurs RAPD ont été utilisés. Laurent *et al.* (1994) ont utilisées 31 sondes de cDNA issues de fèves germées pour étudier la diversité de 203 clones d'origines diverses et il a été montré une différenciation marquée entre les deux grands groupes: Forastero et Criollo.

Contrairement aux cDNA, les sondes nucléaires de rDNA codant pour les ARN ribosomiques sont des séquences répétées en tandem comportant une partie transcrite très conservée et un grand espaceur intergénique très variable (Lapitan, 1992; Delseny *et al.* 1988). Au sein de 192 génotypes étudiés par Laurent *et al.* (1993) trois profils majoritaires ont été identifiés (A, B, C). Le profil A est présent chez la majorité des clones Forastero et une partie des Trinitario, le profil C est présent chez la majorité des clones Criollo et le profil B, qui correspond à un profil hybride entre A et C a été trouvé chez une partie des Trinitario. Figueira *et al.* (1993) ont étudié 30 clones par les sondes rDNA mais n'ont pu caractériser spécifiquement les groupes.

Les génomes mitochondrial et chloroplastique ont généralement une transmission uniparentale. Ils retracent l'histoire évolutive d'un seul parent. Utilisés comme sondes par Laurent *et al.* (1993) sur 177 génotypes, ils ont permis de mettre en évidence l'origine unique des Forastero issus de basse et de haute Amazonie et une grande diversité au sein des Criollo.

Des marqueurs RAPD ont été utilisés par Lerceteau (1992) sur 56 génotypes et par Figueira *et al.* (1993) sur 30 génotypes. Bien que le nombre d'amorces universelles employées et le nombre de fragments amplifiés dans ces études soient élevés, les auteurs n'ont pas réussi à distinguer les groupes morpho-géographiques. Figueira *et al.* (1993) ont pu différencier les cacaoyers sauvages des

minisatellites hypervariables qui sont des séquences répétées dispersées dans tout le génome. A un locus peuvent être associés de nombreux allèles. La technique RFLP est cependant coûteuse et lourde à mettre en oeuvre et l'utilisation fréquente de produits radioactifs la rend quelque peu dangereuse et inaccessible à certains laboratoires.

## II- LES MARQUEURS RAPD

### A - LA PCR ET LES RAPD

La PCR (polymerase chain reaction) est une technique qui permet l'amplification génique *in vitro* de séquences d'ADN. A l'issue de  $n$  cycles d'amplification, la séquence d'ADN cible est multipliée de manière exponentielle;  $2^n$  molécules sont alors produites. La technique a été mise au point par Saiki *et al.* (1985) et utilisée pour la première fois en biologie humaine. La Klenow polymérase isolée à partir de *Escherichia coli* utilisée autrefois pour catalyser la réaction a été remplacée par la Taq polymérase, plus stable, isolée à partir de la bactérie *Thermus aquaticus* qui conserve son activité à des températures élevées (Innis *et al.* 1988 ).

L'amplification génique *in vitro* s'appuie sur la propriété qu'a l'ADN de se dénaturer et de se rénaturer sous certaines conditions de température. En effet, l'ADN bicaténaire lorsqu'il est chauffé à plus de 90°C se sépare en deux brins qui peuvent se réassocier par abaissement progressif de la température. En présence de nucléotides et de polymérase de nouveaux fragments peuvent être ainsi synthétisés à partir des amorces.

Trois étapes se succèdent au cours d'une réaction typique d'amplification. Elles sont facilitées par l'utilisation d'automates spécialement conçus pour ce type de réaction:

1- La **dénaturation** permet aux brins d'ADN matrice de se dissocier par rupture des liaisons hydrogènes. Lorsqu'il s'agit d'ADN génomique, il faut une étape de dénaturation longue et une température élevée (au delà de 90 °C) .

2- L'**hybridation** ou appariement: par abaissement de la température, on permet aux primers ou amorces monocaténares présents dans le milieu de réaction de s'apparier aux régions complémentaires de l'ADN matrice.

3- L'**élongation** et la polymérisation: une fois la synthèse amorcée, il faut la poursuivre en augmentant de nouveau la température jusqu'à 70°C ou 75 °C afin d'accroître la stabilité de l'hybride néoformé et de se situer en zone d'action optimale de la Taq polymérase.

Certains facteurs interviennent de manière déterminante sur les produits de la réaction et leur qualité. Ce sont, l'ADN matrice, les amorces, la température et les ions magnésium et la Taq polymérase.

De très faibles quantités d'ADN sont nécessaires mais varient selon les espèces et la taille du génome ; cependant il en faut suffisamment pour commencer la réaction et obtenir une stabilité des produits. Dans une réaction de PCR, l'amorce est le point de départ de la synthèse et la technique d'amplification telle que décrite par Saiki et al(1988) demande que les primers utilisés correspondent aux séquences bordantes de la région à amplifier. Dès lors des informations sur la séquence cible sont nécessaires pour la synthèse des primers qui doivent remplir certaines conditions pour que l'amplification se fasse avec le maximum d'efficacité:

- être composés de 20 à 30 nucléotides
- avoir une composition en bases équilibrée et analogue à celle du fragment à amplifier
- avoir au moins 50% de G+C (car les liaisons GC sont plus stables que les AT)
- être distants de 400 paires de base au moins

Il faut aussi tester la complémentarité des primers afin d'éliminer ceux susceptibles de donner des produits primer dimers. La composition en bases, la concentration et la température de fusion des primers déterminent la température d'hybridation au cours de la réaction d'amplification.

$T_m = 2(A+T)+4(G+C)$  pour les primers ayant moins de 18 bases.

et  $T_m = 81,5-16,6(\log_{10}(Na^+))+ 0,41(\%GC)+69,3 - 600/N$  avec N le nombre de bases de l'amorce pour ceux de plus de 20 bases (Sambrook et al. 1989).

$T_m$ = température de fusion.

Généralement, il faut soustraire 5 °C de la  $T_m$  pour obtenir une température d'hybridation satisfaisante.

Les ions  $Mg^{++}$  jouent un rôle important dans la stabilité du complexe primer/matrice (Shin et al., 1990) et leur déficit entraîne l'obtention de faibles produits tandis que leur excès provoque l'accumulation de produits non spécifiques. L'enzyme catalysant la réaction d'amplification a une activité entre 70 et 80°C. Elle a un taux d'erreur d'incorporation de nucléotides de  $2 \times 10^{-4}$  par cycle (Saiki et al., 1988; Le Pennec, 1991).

Les marqueurs RAPD (Random amplified polymorphic DNA) sont des marqueurs moléculaires issus d'une amplification génique *in vitro* utilisant des amorces de 9 à 10 bases. Les fragments d'ADN à amplifier sont de séquences inconnues mais du fait de la petite taille des amorces, il est souvent possible de

trouver plusieurs séquences complémentaires des amorces dans le génome et qui pourront ainsi être à l'origine des amplifications. A l'issue de l'amplification, la présence ou l'absence de fragments amplifiés permet de mettre en évidence un polymorphisme entre des individus lié soit à des mutations sur les sites d'hybridation des amorces, soit à une insertion ou une délétion au niveau des sites. La détection des produits se fait après migration dans un gel d'agarose, de polyacrylamide ou dans un gel dénaturant. Ces marqueurs ont été utilisés pour la première fois par Williams *et al.* (1990) et par Welsh et McClelland (1990). Une seule amorce suffit pour obtenir une amplification si deux copies de la séquence sont suffisamment proches et en position inversée dans le génome mais pour varier les produits, les amorces peuvent être utilisées par paire (Klein-Lankhorst *et al.*, 1991). Des changements de nucléotides au sein d'une amorce aboutissent à l'obtention de profils d'amplification différents (Williams *et al.* 1991). Différentes applications des RAPD détaillées ci-dessous sont possibles.

## **B- CARTOGRAPHIE ET IDENTIFICATION DE MARQUEURS SPÉCIFIQUES DE RÉGIONS CHROMOSOMIQUES**

Les marqueurs RAPD sont généralement dominants (Hunt et Page, 1992; Heun et Helentjaris, 1993; Yang et Quiros, 1993) et qualitativement seuls deux phénotypes peuvent être observés quelles que soient les circonstances : les homozygotes récessifs ne présentant pas de produit d'amplification et les individus hétérozygotes ou homozygotes dominants produisant un fragment d'amplification. Les marqueurs RAPD codominants sont rares (Tinker *et al.*, 1993) et il est difficile de quantifier les produits par leur intensité pour déterminer les hétérozygotes (Hunt et Page 1992). Dans ces circonstances, seules certaines populations de cartographie peuvent être utilisées. Alors qu'avec les marqueurs codominants de type RFLP ou isozymes les populations F2 sont les meilleures pour l'établissement des cartes, elles s'avèrent non adaptés lorsque les marqueurs RAPD sont utilisés et la précision des estimations de fréquences de recombinaisons s'en ressent. Les populations idéales pour les RAPD sont des lignées recombinantes, des populations de backcross (Gebhardt *et al.* 1989) mais aussi des populations de gamétophytes. Lorsque ces populations idéales ne sont pas disponibles, des descendances de parents hétérozygotes sont utilisées en considérant que l'on se situe dans un cas de backcross pour certains locus, appelé encore pseudobackcross (Carlson *et al.*, 1991, Echt *et al.* 1992).

Quelques cartes génétiques ont été établies uniquement au moyen des RAPD. Chez les conifères, les mégagamétophytes, tissus nutritifs des graines peuvent être utilisés comme populations de cartographie. Ils proviennent de la

division de la mégaspore et sont haploïdes. L'accès direct aux gamètes via les mégagamétophytes est ainsi possible et le caractère dominant des marqueurs RAPD ne constitue pas un réel problème car tous les génotypes sont identifiés sans équivoque. Ainsi Tulsieram *et al.*, (1992) et Nelson *et al.*, (1993) ont construit respectivement une carte de *Picea glauca* et de *Pinus elliotii* avec 61 et 73 marqueurs RAPD.

Chez *Arabidopsis thaliana* où une population de lignées recombinantes de 150 individus était disponible, Reiter *et al.* (1992) ont construit en peu de temps, une carte comprenant 252 marqueurs RAPD. Chez *Helianthus*, la difficulté de production de descendances F2 et backcross a conduit Rieseberg *et al.* (1993) à recourir à une descendance issue d'un croisement interspécifique pour établir une carte génétique de 161 RAPD répartis sur 18 groupes de liaison. Bien qu'il soit souvent plus facile d'établir des cartes RAPD que des cartes RFLP, la tendance actuelle est à l'utilisation de RAPD en mélange avec d'autres types de marqueurs car l'universalité des cartes RAPD est de plus en plus remise en cause (Kesseli *et al.* 1991). En effet, Fauré *et al.* (1993) ont utilisé deux populations pour cartographier le génome du bananier: certains marqueurs RAPD ont été cartographiés dans des groupes de liaison différents d'une population à l'autre. Les marqueurs RAPD sont de plus en plus utilisés pour la saturation de régions spécifiques du génome et la détection de liaisons entre RAPD et caractères d'intérêt agronomique. La présence de lignées isogéniques chez certaines plantes facilite le marquage régional. Chez la tomate, des lignées isogéniques ont permis d'identifier des marqueurs RAPD liés au gène responsable de la résistance au nématode sur le chromosome 6 (Klein-Lankhorst *et al.* 1991) ainsi que des marqueurs liés au gène de la résistance au *Pseudomonas syringae* sur le chromosome 5 (Martin *et al.*, 1991). Grâce à la mise en évidence de deux marqueurs RAPD liés au gène de la résistance à la rouille, il est désormais possible d'envisager l'introggression de la résistance à la rouille dans les génotypes sensibles de haricot (Haley *et al.*, 1993). Chez l'avoine, Penner *et al.* (1993) ont mis en évidence une liaison entre un marqueur RAPD et le gène de résistance à la rouille. En l'absence de lignées isogéniques, la BSA (Bulk segregant analysis) peut être utilisée pour saturer une région particulière du génome dès lors qu'on dispose d'une carte génétique et d'une population en ségrégation. La région à traiter doit être délimitée par deux marqueurs et sa taille inférieure ou égale à 15 cM. Les individus homozygotes pour la région cible sont regroupés en deux lots correspondant aux deux types d'homozygotes. Tous les autres marqueurs hors de la région se trouvent en équilibre de liaison et lorsque les deux lots d'individus sont amplifiés, on n'observera du polymorphisme que dans la région cible. Il est

alors possible de la marquer plus finement (Giovannoni *et al.*, 1991; Kesseli *et al.*, 1991, Michelmore *et al.*, 1991, King *et al.*, 1993).

### C- ANALYSE DES POPULATIONS

Un des domaines d'application des RAPD est l'analyse des populations de diverses espèces à travers des identifications variétales ou fingerprint, des déterminations de parenté ou des études de diversité génétique.

Les marqueurs RAPD sont bialléliques contrairement aux marqueurs minisatellites ou microsateellites. Ceci peut constituer un problème à première vue, cependant, le nombre de produits obtenus après une amplification est élevé et compense largement le faible nombre d'allèles possibles. L'amplification non spécifique avec des amorces de courte séquence aboutit à la formation de nombreux produits utilisables pour générer des empreintes génétiques. Le nombre de produits obtenus est variable d'une amorce à l'autre et dépend des espèces analysées. Alors que Echt *et al.* (1992) ont obtenus entre 3 et 12 produits par primer chez la luzerne, Huff *et al.* (1993) ont générés jusqu'à 29 fragments d'ADN avec certaines amorces chez *Buchloë dactyloides*. Si l'on considère qu'un fragment amplifié correspond à un marqueur, il est possible d'utiliser un nombre limité de primers à des fins d'identification et dans ce cas le caractère multilocus des RAPD apparaît comme un avantage. Pour être utile à des identifications, un marqueur doit pouvoir révéler du polymorphisme intragroupe et intergroupe et être de surcroît stable vis à vis de l'environnement. Les marqueurs RAPD révèlent souvent de la variabilité entre individus très proches et pour certains auteurs seront certainement utiles dans les certifications des accessions (Adams *et al.*, 1993). Chez l'orge, il a été possible de détecter du polymorphisme entre individus présentant un coefficient de parenté de 0,92 (Tinker *et al.*, 1993). Ainsi des identifications ont été réalisées chez de nombreuses espèces avec un nombre limité de primers ou au contraire avec plusieurs primers. Une seule amorce a généré 30 fragments d'ADN sur 20 accessions de Tagasaste (*Chamaecytisus proliferus*) et a permis l'identification de plusieurs génotypes (Francisco-Ortega *et al.*, 1993). De même 32 clones de peupliers ont été identifiés avec quatre amorces qui ont produit 120 bandes (Castiglione *et al.*, 1993). Par contre, il a fallu 9 amorces pour identifier sans équivoque 13 clones de cacaoyers (Wilde *et al.*, 1992).

Du fait de la rapidité des analyses, les RAPD sont de plus en plus utilisés pour évaluer la diversité génétique au sein des populations et mettre en évidence des liens de parenté (Kazan *et al.*, 1993). Les travaux de Huff *et al.* (1993) ont mis en évidence la puissance discriminatoire de certains primers et la nécessité d'un choix sélectif de ceux-ci pour l'analyse des populations. Dans le cas de structuration géographique forte, un nombre limité de primers suffisent à

différencier les origines mais au sein d'une origine, il peut s'avérer nécessaire d'utiliser plusieurs primers pour arriver à distinguer les clones.

Lorsque de nouveaux marqueurs apparaissent, il est courant de comparer leurs résultats à ceux obtenus par des marqueurs plus classiques pour savoir si l'utilisation de nouveaux outils améliorent la compréhension d'un système biologique donné. Chez le riz, Glaszmann (1987) a étudié la variabilité enzymatique des riz sur un grand nombre de génotypes d'origines diverses: il a mis en évidence l'existence de deux groupes majeurs et de quelques groupes mineurs. Par la suite, les études par les marqueurs RFLP de Second et Ghesquière (1994) entre autres ont confirmé l'existence de deux groupes majeurs et laissent supposer que les groupes mineurs trouvés par Glaszmann sont des hybrides entre les deux groupes majeurs. Fukuoka *et al.* (1992) ont utilisé 16 accessions de riz pour évaluer la faisabilité des RAPD pour l'étude de la diversité et la cartographie du riz. Chaque génotype était caractérisé par un marqueur RAPD et les regroupements obtenus étaient en accord avec ceux découlant des analyses isozymiques et RFLP. Ils en concluent que l'avantage des analyses RAPD chez le riz réside dans leur simplicité ainsi que dans le plus grand taux de polymorphisme détecté. Chez la tomate, Williams et St-Clair (1993) ont recherché des liens entre plusieurs groupes de variétés au moyen des RFLP et des RAPD. Quel que soit l'outil utilisé, un faible taux de polymorphisme est révélé au niveau intraspécifique. Par ailleurs, ils ont insisté sur les précautions à prendre dans les analyses RAPD afin d'aboutir à un maximum de robustesse de la technique (stabilité des bandes et vérification de leur Mendélisme). Chez le peuplier, une comparaison des résultats entre RAPD, RFLP et isozymes a été réalisée (Liu et Furnier, 1993). Des loci spécifiques des deux espèces ont été mis en évidence avec les RAPD, RFLP et isozymes ainsi qu'une grande diversité chez *Populus tremuloïdes* par rapport au *Populus grandidentata*. Le niveau de variabilité détecté par les RAPD est toujours plus élevé, dû probablement à une plus grande capacité de ces marqueurs à révéler du polymorphisme mais les RAPD semblent moins propices pour des estimations d'un certain nombre de paramètres génétiques dans des populations allogames. Une concordance générale existe entre les résultats obtenus par les RFLP, les isozymes et les RAPD. Souvent, la mise en évidence de grands groupes reste constante mais les relations entre individus à l'intérieur des groupes sont mieux définis par les marqueurs d'ADN. Cependant des cas de désaccord existent: chez le sorgho, Vierling *et al.* (1994) ont comparé 18 lignées au moyen des RFLP et des RAPD: le phénogramme obtenu par les RFLP correspondait le mieux aux phénogramme selon les pedigrees. Chez l'avoine sauvage (*Avena sterilis*), Heun *et al.* (1994) ont montré que les isozymes ainsi que les RAPD distinguaient sans équivoque les 24

accessions étudiées et bien que les regroupements fussent comparables, les RAPD permettaient une séparation plus radicale des groupes.

#### **D- CONTRAINTES DE L'UTILISATION DES RAPD**

Les RAPD en tant que marqueurs moléculaires font l'objet de critiques quant à leur répétabilité. Leur utilisation nécessite l'application et la mise en place d'un certain nombre de précautions.

La réaction de PCR aboutissant aux RAPD est non spécifique et très sensible, un ADN contaminant peut être facilement amplifié, aussi faut-il veiller à travailler dans des conditions d'asepsie très strictes. Lorsque ces considérations techniques sont maîtrisées, il faut veiller à trouver des conditions répétables car en dehors des contaminations extérieures, la réaction d'amplification contient en elle-même de nombreux facteurs de variations qu'il ne faut pas négliger. Il semble en effet que la zone optimum de conditions favorables soit étroite et qu'il ne faille pas s'en éloigner de crainte d'obtenir des amplifications non spécifiques.

Du fait de la non spécificité des RAPD, une bande d'une taille donnée peut provenir de l'amplification d'ADN au niveau de plusieurs locus et ainsi sera inadaptée pour certaines études.

Liste des clones Forastero haut-amazoniens étudiés

Origine	Clone	Collection	
?	Amazon15-15	IRCC	RAPD morphologie
Colombie	EBC5	IRCC	RFLP
	EBC6	IRCC	RFLP
Equateur	EBC10	IRCC	RAPD RFLP
	SPA5	CATIE	RAPD RFLP
	SPA11	CATIE	RFLP
	SPA17	CATIE	RAPD RFLP
	IMC5	IRCC	RFLP
	IMC31	IRCC	RAPD RFLP morphologie
	IMC67	IRCC	RAPD RFLP morphologie
	IMC78	IRCC	RFLP
	LCTEEN37	IRCC	RAPD RFLP
	LCTEEN84	CRU	RAPD RFLP
	LCTEEN127	CRU	RAPD RFLP
	LCTEEN202	CRU	RFLP
	LCTEEN295	CRU	RFLP
	LCTEEN325	CRU	RAPD RFLP
	LCTEEN326	CRU	RFLP
Pérou	LCTEEN355	IRCC	RFLP
	LCTEEN371	IRCC	RAPD
	SCA6	IRCC	RAPD RFLP morphologie
	SCA9	IRCC	RAPD RFLP
	SCA12	IRCC	RAPD RFLP
	PA13	IRCC	RAPD RFLP morphologie
	PA121	IRCC	morphologie
	PA150	IRCC	RAPD
	P1	IRCC	RFLP
	P2	IRCC	RAPD RFLP
	P7	IRCC	morphologie
	P16	IRCC	RFLP
	P19A	RCC	morphologie
	P13B	IRCC	morphologie
	P32A	IRCC	RAPD RFLP morphologie
Ghana	MO9	IRCC	RFLP morphologie
	MO81	IRCC	RAPD RFLP morphologie
	MO98	IRCC	RFLP morphologie
	NA32	IRCC	RAPD RFLP
	IFC312	IRCC	RFLP
	T60/887	IRCC	RAPD RFLP morphologie
	T63/967	IRCC	RAPD RFLP morphologie
Côte d'Ivoire	T79/501	IRCC	RFLP
	T85/799	IRCC	RAPD RFLP
	UPA401	IRCC	RFLP morphologie
	UPA402	IRCC	morphologie
	UPA409	IRCC	morphologie
	UPA413	IRCC	RAPD RFLP morphologie
	UPA603	IRCC	RAPD RFLP
UPA608	IRCC	RFLP	
	UPA620	IRCC	morphologie

Tableau 3: Liste des clones dans les différentes études

Liste des clones Trinitario étudiés

Origine	Clone	Collection	Etudes
Trinidad	ACT2-11	IRCC	RAPDRFLP, morphologie
	ICS6 (EET275)	IRCC	RAPDRFLP, morphologie
	ICS16	IRCC	RAPD RFLP
	ICS46	IRCC	RFLP
	ICS53	IRCC	RAPDRFLP, morphologie
	ICS75	IRCC	RFLP
	ICS89	IRCC	RAPDRFLP morphologie
	ICS95 (ERJOH21)	CEPEC	RAPDRFLP, morphologie
	ICS98 (ERJOH18)	CEPEC	RAPD RFLP
Grenade	GS29	IRCC	RFLP
	GS36	IRCC	RFLP
Venezuela	CHUAO 24		RAPD RFLP
	CHUAO 120	CATIE	RFLP
	CNS 22	CATIE	RFLP
Côte d' Ivoire	CNS 23	CATIE	RAPD
	IFC6	IRCC	RAPDRFLP, morphologie
	IFC7	IRCC	RFLP
	IFC11	IRCC	RFLP
	IFC19	IRCC	RAPDRFLP, morphologie
	IFC413	IRCC	RAPDRFLP, morphologie
	IFC420	IRCC	RFLP
	IFC422	IRCC	RFLP
Ghana	IFC304	IRCC	RFLP
	ACU85	IRCC	RAPDRFLP, morphologie
	K5 (IFC305)	IRCC	RAPD RFLP
Indonésie	DR1	IRCC	RAPDRFLP, morphologie
	G23	CATIE	RAPD RFLP
Samoa	LAFI 7		RAPD RFLP
Java	WA40	IRCC	RAPDRFLP, morphologie
Equateur	EQX27	IRCC	RFLP
	EQX94	IRCC	RAPD RFLP
	EQX107	IRCC	RAPDRFLP, morphologie
	MOQ122	CRU	RFLP
	MOQ216	IRCC	RFLP
	MOQ413	IRCC	RAPDRFLP, morphologie
	MOQ647	IRCC	RFLP morphologie
Honduras	MT1	CATIE	RAPD RFLP
	TJ1	CATIE	RAPD RFLP
Nigeria	N38	IRCC	RAPD RFLP
Mexique	RIM8	CATIE	RFLP
	RIM15	CATIE	RFLP, RFLP, morphologie
	RIM19	CATIE	RFLP
	RIM76,	CATIE	RFLP
	RIM105	CATIE	RFLP
	RIM113	CATIE	RFLP
Colombie	SC5	CATIE	RAPD RFLP
	SC6	CATIE	RFLP
	SPEC54-2	IRCC	RFLP
	SPEC138-8	IRCC	RFLP
Cameroun	SPEC160-9	IRCC	RFLP, RFLP, morphologie
	SNK12	IRCC	RFLP morphologie
	SNK109	IRCC	RFLP
Costa Rica	CC10	IRCC	RFLP
	CC39	IRCC	RAPD RFLP
	UF10	CATIE	RFLP
	UF221	IRCC	RAPD RFLP
	UF296 (ERJOH 17)	CEPEC	RFLP
	UF667 (IFC409)	IRCC	RFLP morphologie
	UF676 (IFC410)	IRCC	RFLP
Panama	UF168	CATIE	RFLP

# CHAPITRE III: ETUDE DE LA DIVERSITE GENETIQUE DE L'ESPECE

## I- MATERIEL ET METHODES

### 1- MATÉRIEL VÉGÉTAL

Le matériel végétal utilisé dans ces études (Tableau 3) provient de plusieurs origines et a été fourni par diverses institutions de recherches sur le cacao dans le monde (IDEFOR-DCC en Côte d'Ivoire, le CEPEC au Brésil, le CATIE au Costa Rica, le CRU à Trinidad, le CIRAD à Montpellier et le FONAIAP au Venezuela) . Dans chacune des études réalisées, un nombre différent de clones a été échantillonné , cependant toutes les origines sont représentées. Certains individus sont communs à toutes les analyses. Le nombre de génotypes utilisés dans chaque analyse est donné dans le tableau 4.

	Analyses morphologiques	Analyses RAPD	Analyses RFLP
Forastero Haut Amazoniens	22	25	40
Forastero Bas Amazoniens	5	26	41
Criollo	5	28	41
Trinitario	20	27	59

**Tableau 4:** Nombre de génotypes de cacaoyers utilisés dans les différentes analyses de diversité génétique.

Les observations morphologiques ont été réalisées en Côte d'Ivoire sur 52 clones : La faible représentation des Criollo dans l'analyse morphologique s'explique par la présence de très peu d'individus de ce groupe au sein de la collection de Côte d'Ivoire. Par contre bien qu'un nombre important de clones Forastero Bas Amazoniens soit présent en collection, il n'a pas toujours été possible d'obtenir un nombre suffisant de cabosses sur une période donnée pour réaliser les observations.

L'étude de diversité par les RAPD a porté sur 106 génotypes appartenant à tous les groupes connus.

Les paramètres de la structure des populations ont été calculés sur 181 clones, dont le typage RFLP a été réalisé par Laurent *et al.* (1994) qui ont par

ailleurs décrit la diversité des cacaoyers par les RFLP à l'aide de méthodes d'analyses descriptives.

Les clones Forastero Haut Amazoniens étudiés comportent :

- des clones provenant des prospections de Pound en 1938 et 1945, au Pérou et en Equateur le long des fleuves Nanay, Parinari, Moroni etc...(clones PA, SCA, P, IMC, MO, NA) et leurs descendances sélectionnées. Les descendances de première génération ont été obtenues au Ghana : ce sont les clones T60, T63, T79, T85. Les descendances de deuxième génération proviennent de Côte d'Ivoire et sont dénommées UPA. Les clones prospectés par Pound ont été très largement diffusés et sont utilisés à travers le monde entier dans les programmes d'amélioration et de sélection. Certains sont connus pour être de bons géniteurs (IMC67) d'autres sont tolérants à des maladies (SCA6).

-des clones issus des prospections réalisées en Colombie et en Equateur par la mission Anglo-Colombienne en 1982 et 1984. Ce sont les clones LCTEEN et EBC.

Les clones Forastero Bas Amazoniens analysés proviennent surtout du Brésil, du Venezuela mais aussi d'Afrique:

-les clones sauvages ERJOH sont issus de prospections effectuées au Brésil surtout dans sa partie Ouest. Certains proviennent de zones frontalières (Pérou, Equateur, Colombie), d'autres de zones plus internes.

-les clones SIC et SIAL du Brésil proviennent de la frontière avec le Venezuela.

-les clones PARA, COMUN, sont des sélections de Bahia au Brésil. Les clones MATINA qui leur sont proches sont aussi issus du bassin Amazonien, près de Belem .

-les clones EET, ECNR et NACIONAL font partie de la variété Nacional qui donne un cacao de qualité "arriba". Ils proviennent d'Equateur mais leur classification est controversée.

-les Forastero Bas Amazoniens du Venezuela (VEN) sont des populations sauvages du Centre et du Sud du Venezuela prospectées par Lanaud *et al.* (1986) le long des fleuves Cuchivero, Orinoco et Padamo;

- les Forastero Bas Amazoniens sauvages de Guyane (GU), prospectés en Guyane française(Sallée)

-les Forastero Bas Amazoniens d'Afrique (clones IFC) sont constitués de clones sélectionnés dans les cacaoyères créées au début des introductions de cacaoyers en Afrique de l'Ouest;

Les clones Trinitario analysés sont très divers:

-les clones ICS (Imperial College Selection) ont été sélectionnés à Trinidad et les clones GS à Grenade;

- les clones UF (United Fruit) et CC sont originaires du Costa Rica;
- les Trinitario d'Equateur sont composés de clones sauvages provenant des prospections de Pound en Equateur (MOQ) et d'autres sélections (EQX);
- les clones RIM qui ont été analysés proviennent du Mexique;
- les Trinitario d'Afrique sont composés de clones sélectionnés en Côte d'Ivoire (IFC), au Cameroun (SNK) et au Ghana (ACU et K5).

Les clones Criollo échantillonnés proviennent majoritairement du Vénézuéla, et du Nicaragua. Les Criollo ne se rencontrent pratiquement plus en culture pure du fait de leur faible potentiel agronomique (Wood et Lass, 1985) et sont souvent des hybrides entre Criollo et Trinitario.

Du fait de l'éloignement des différentes collections du laboratoire Agotrop/CIRAD où les analyses moléculaires ont eu lieu, le matériel végétal a été reçu sous forme de feuilles séchées à la température ambiante enfermées dans des gaines plastiques et conservées dans une chambre froide à +4 °C pour éviter les reprises d'humidité.

## 2- MÉTHODES

### A- ANALYSES MORPHOLOGIQUES

Dans un essai préliminaire, la taille des échantillons nécessaires aux mesures de fèves et de cabosses a été déterminée. Pour cela une étude a été conduite sur la parcelle A18/2 de Divo (Côte d'Ivoire) sur 10 clones à raison de 5 arbres par clone. Chaque arbre a été pollinisé manuellement afin d'obtenir 10 cabosses par arbre.

Si on veut détecter une différence  $d$  correspondant à 5 % de la moyenne, alors la formule permettant de déterminer la taille d'échantillon est (Engels *et al.*, 1980):

$$r = 4 \times S^2 / d^2 \quad S^2 \text{ est la variance, } r \text{ est l'effectif de cabosses ou de fèves}$$

Sur les 52 clones analysés, plusieurs caractères de cabosses et de fèves ont été mesurés: le poids de cabosses, le nombre de fèves par cabosse, le diamètre de cabosse, la longueur de cabosse, l'épaisseur du cortex, le poids de cabosse vide, le poids de fève, la longueur de fève et l'épaisseur de fève. Pour chacun de ces caractères, la moyenne et l'écart type ont été déterminés pour chaque clone. Par la suite, les individus d'un même groupe ont été rassemblés et une analyse de variance à un facteur a été réalisée pour chaque caractère, suivie d'une comparaison de moyennes par le test de Fisher de la plus petite différence significative. Ces analyses ont été réalisées à l'aide du logiciel Statview. Les moyennes des clones ont ensuite été utilisées pour une analyse en composantes principales (ACP), analyse multivariée qui donne une image plus synthétique des données pour l'ensemble des caractères.



Photo 3: Extraits d'ADN de cacaoyer

## **B- ANALYSES MOLÉCULAIRES**

### **Extraction de l'ADN total**

La méthode d'extraction utilisée ( Laurent *et al.* 1993) est donnée dans l'annexe 1. On obtient des solutions d'ADN de bonne qualité après une étape de purification au chlorure de césium (Photo 3).

### **Dosage de l'ADN**

Trois types de dosages ont été réalisés. La méthode au spectrophotomètre estime la concentration d'ADN double brin par l'absorbance à 260 nm en considérant qu'une densité optique de 1 sur un trajet optique de 1 cm correspond à 50 µg/µl d'ADN. La pureté de la solution d'ADN peut être évaluée par le rapport des densités optiques à 280 et à 260 nm. Un rapport voisin de 1,8 est souhaitable et correspond à une bonne pureté de l'ADN bicaténaire par contre si ce rapport est inférieur à 1,8 c'est que l'ADN est contaminé par des protéines ou du phénol. Bien que purifié par une ultracentrifugation par gradient de chlorure de césium, l'ADN du cacao reste souvent contaminé et les concentrations d'ADN mesurées par le spectrophotomètre sont souvent surestimées.

Les concentrations d'ADN ont aussi été estimées par un fluorimètre. Bien qu'ils donnent des estimations meilleures que le spectrophotomètre, les dosages restent toujours insatisfaisants et par ailleurs, plus difficiles et plus contraignants à réaliser que ceux par spectrophotomètre.

Nous avons obtenu de meilleurs résultats en estimant les concentrations en ADN sur un gel d'agarose en présence d'une gamme d'ADN de phage lambda de concentrations connues, après avoir dosé préalablement l'ADN au spectrophotomètre ou au fluorimètre. Ce genre de dosage permet aussi de se rendre compte si l'ADN est dégradé. Après le dosage de l'ADN, des dilutions ont été faites afin d'obtenir des concentrations de 1 ng/µl d'ADN pour les solutions à amplifier. Plusieurs fractions aliquotes ont été faites et conservées au congélateur à -20 °C pour éviter la dégradation des solutions.

### **Réactions d'amplification**

Les réactions d'amplifications ont été mises au point étape par étape en testant plusieurs facteurs (concentrations d'ADN, concentration d'enzyme etc...). Elles ont été réalisées dans un volume de 25 µl et les concentrations des différents éléments ont été ajustées au fur et à mesure que les tests étaient réalisés. Les thermocycleurs utilisés sont du type PHC2 de la société TECHNE. Le protocole d'amplification est donné en annexe 2.

### **Migration et visualisation**

La migration des produits amplifiés a été faite sur gels d'agarose à 1,4% dans le tampon TBE meilleur que le TAE pour la résolution des bandes. Les

produits ont été migrés pendant deux heures à 100 V puis colorés au Bromure d'éthidium pendant 30 minutes.

#### **Tri des amorces**

Les amorces utilisées pour les réactions d'amplification provenaient des kits Operon commercialisés par la firme Operon Technologies Inc. Le tri a été effectué sur un échantillon de 19 génotypes provenant de toutes les origines. Un tri effectué lors de la cartographie génétique avait permis de sélectionner uniquement les primers qui induisaient la formation de produits d'amplification. Pour l'étude de la diversité génétique 20 amorces ont ainsi été sélectionnées.

#### **Essais de répétabilité**

Les essais de répétabilité ont été couplés avec le tri des primers. Une première amplification a été réalisée sur 19 individus avec trois doses d'ADN provenant de la même extraction. Une autre amplification a ensuite été réalisée dans les mêmes conditions expérimentales que la première et le nombre de bandes et leur netteté ont été notés et comparés aux bandes produites lors de la première amplification. Si des différences apparaissaient alors une troisième amplification était réalisée et les bandes polymorphes les plus stables étaient sélectionnées. Si aucune bande n'était restée stable d'une amplification à l'autre alors le primer n'était pas sélectionné.

#### **Extraction de produits amplifiés**

Afin de pouvoir isoler des fragments polymorphes amplifiés à utiliser comme sondes pour réaliser des hybridations sur de l'ADN total, une électrophorèse sur gels d'agarose à bas point fusion de 1,4 % a été réalisée à 100 V et 50 mA pendant deux à trois heures selon la proximité des fragments. Les gels ont ensuite été colorés et observés aux U.V.. A l'aide de fragments de paille, les fragments amplifiés ont été récupérés à la pipette puis stockés dans des tubes à vis au réfrigérateur. Au moment de l'utilisation des bandes RAPD comme sondes, les tubes ont été chauffés au bain marie à sec à 65 °C pour faire fondre l'agarose. Toutes les bandes amplifiées et polymorphes n'ont cependant pas pu être isolées car certaines étaient très proches les unes des autres et mêmes les longues migrations n'ont pas réussi à les séparer.

#### **Marquage des fragments amplifiés**

Le marquage des fragments amplifiés a été fait au  $^{32}\text{P}$  à l'aide du kit "Megaprime labelling systems" de Amersham par la technique du "random priming" ou amorçage aléatoire. La qualité du marquage a été vérifiée par chromatographie sur couche mince. Le protocole est donné en annexe 3.

#### **Préhybridations et hybridations**

Les préhybridations et hybridations ont été faites dans des tubes en verre pendant respectivement 5 et 15 heures dans des fours à hybridation. Les

fragments amplifiés marqués ont été hybridés sur des membranes hybridant l'ADN total de cinq clones POR, UPA409, SNK12, UF676 et UPA402 digéré par les enzymes de restriction Hind III, EcoR V, Xba I, Eco RI et Bgl II (protocole en annexe 4).

#### Collecte des données

Les résultats d'amplification ont été lus sur les photographies des gels colorés au BET après électrophorèse. Plusieurs lectures ont été faites à des moments différents, puis les différentes notations ont été comparées. La présence de marqueurs de taille sur les gels a permis de situer les fragments les uns par rapport aux autres et de déterminer leur taille. Lorsqu'un produit était présent, la note 1 était attribuée au clone pour ce produit et s'il était absent l'individu était noté 0. Lorsqu'on ne réussissait pas à déterminer la présence ou l'absence d'un produit chez un individu après plusieurs lectures, alors on a considéré chez cet individu la donnée comme manquante pour ce produit. Les tableaux de données ont ensuite été traités à l'aide du logiciel ADDAD.

#### Analyse des données

##### *Répartition des produits d'amplification dans les populations et polymorphisme*

Toutes les bandes amplifiées n'ont pas été répertoriées. Seules celles qui étaient polymorphes et stables au sein de l'échantillon étudié ont été analysées. Le pourcentage des bandes présentes dans les populations a été calculé et des tests de Chi-deux ont été utilisés pour d'une part savoir si la répartition des bandes était homogène dans l'ensemble des populations et d'autre part comparer les populations deux à deux et dégager les similitudes au niveau de la fréquence des bandes.

Soit  $p$ , le nombre de groupes comparés (ici 4),  $q$  le nombre de classes phénotypiques dans chaque groupe (ici 2) et  $N$  le nombre total d'individus. On peut construire un tableau à quatre lignes et deux colonnes dont les cellules correspondent aux différentes classes observées.

$$Chi^2 = \sum \sum (n_{ij} - N \times P_{ijOBS})^2 \quad (\text{Dagnélie, 1975})$$

avec  $n_{ij}$  le nombre d'individus dans la classe  $ij$ ,  $P_{ij}$  la fréquence de la classe  $ij$  et  $(p-1)(q-1)$  le nombre de degrés de liberté. Pour la comparaison des populations deux à deux, six tests de Chi-deux sur le modèle suivant ont été réalisés chacun à un degré de liberté pour chacune des bandes RAPD:

le calcul devient plus simple car  $p=2$  et  $q=2$

$$\text{Chi}^2 = \frac{N(n_{11} \times n_{22} - n_{12} \times n_{21})^2}{n_{1.} \times n_{2.} \times n_{.1} \times n_{.2}}$$

$n_{1.}$   $n_{2.}$   $n_{.1}$   $n_{.2}$  sont les sommes marginales

### *Analyse factorielle des correspondances (AFC)*

C'est une méthode de statistique descriptive qui permet de synthétiser des tableaux de données à double entrée par l'extraction d'un nombre réduit de facteurs correspondant à des combinaisons linéaires des variables initiales. Tout tableau contenant des nombres positifs peut être traité par l'AFC. La démarche aboutit à une représentation graphique des points dans un espace de dimension réduite permettant de mettre ainsi en évidence des structures au sein des données. Les ressemblances entre individus sont mesurées par la distance du  $\text{Chi}^2$  (Jambu, 1989).

Le logiciel ADDAD utilisé pour le traitement informatique des données ne sait pas gérer les données manquantes, nous avons mis les individus présentant des données manquantes en supplémentaire pour ne pas biaiser le calcul des coordonnées des axes.

### *Classification ascendante hiérarchique*

C'est une méthode de classification automatique élaborée qui permet de créer des partitions. Il faut pour cela définir un indicateur de liaison entre individus et un critère d'agrégation des classes. Grâce à l'indicateur de liaison qui peut être une mesure de distance ou d'indice de similarité ou de dissimilarité, les individus les plus proches sont regroupés. Ensuite intervient le critère d'agrégation qui permet de regrouper les classes (Jambu, 1989; Fenelon, 1981).

Pour des tableaux de présence et d'absence, certaines distances ou indices de distances sont plus utilisés que d'autres. Certains d'entre eux tiennent compte de l'absence simultanées des bandes chez deux individus, d'autres n'en tiennent pas compte. Par soucis de comparaison, deux types de distances ont été utilisées. Nous pensons que dans le cas des RAPD qui sont des marqueurs dominants, l'absence d'un fragment d'amplification est aussi importante que sa présence. C'est pour cela que nous avons utilisé deux types d'indice l'un tenant compte que de la présence des fragments (Czekanowski) et l'autre tenant compte aussi bien des présences que des absences simultanées pour définir la proximité des génotypes (Rogers et Tanimoto):

Soit  $n(i,j)$  le nombre de bandes présentes simultanément chez les individus  $i$  et  $j$

soit  $n_i$  le nombre de bandes présentes chez  $i$  et absentes chez  $j$ .  
 soit  $n_j$  le nombre de bandes présentes chez  $j$  et absentes chez  $i$   
 soit  $0(i,j)$  le nombre de bandes absentes simultanément chez  $i$  et  $j$

*Indice de similarité de Czekanowsky(Addad, 1985)*

$$d(i,j) = \frac{2n(i,j)}{2n(i,j) + n_i + n_j}$$

*Indice de similarité de Rogers et Tanimoto(Addad,1985)*

$$d(i,j) = \frac{n(i,j) + 0(i,j)}{n(i,j) + 0(i,j) + 2(n_i + n_j)}$$

Plusieurs critères d'agrégation sont disponibles dont le saut minimum qui permet de regrouper les classes sur la base de la plus petite distance entre les éléments de la classe déjà formée et les autres éléments, et l'agrégation par la moyenne qui calcule la distance moyenne entre les éléments pour ensuite choisir la distance moyenne la plus petite. Notre choix s'est porté sur la moyenne encore appelée UPGMA (Unweighted Pair Group Method using arithmetic Averages). Sur tableaux de distances, le programme ADDAD CAHDIS n'utilise que les individus ayant des données complètes pour la réalisation des classifications.

La classification ascendante hiérarchique a aussi été effectuée à partir des coordonnées des facteurs issus d'une analyse factorielle des correspondances. Ce type de classification permet de classer aussi bien les individus avec des données complètes que ceux dont les données sont incomplètes car bien que ne participant pas à la construction des axes parce que introduits dans l'analyse comme individus supplémentaires, ils ont quand même des coordonnées sur les axes. Le programme CAHVOR utilisé est un programme de classification qui utilise la méthode des voisins réciproques et tient compte de l'information apportée par les voisins.

### *Paramètres génétiques*

Du fait que certaines sondes cDNA ont été cartographiées et que d'autres donnaient un profil simple à une ou deux bandes permettant de faire l'hypothèse d'un déterminisme génétique simple à un locus, il a été possible de raisonner en fréquences alléliques et d'étudier la structure génique des populations en présence. Ces analyses ne sont pas toujours possibles avec les marqueurs RAPD qui sont des marqueurs dominants. Sur les 31 combinaisons enzyme/sonde utilisées par Laurent et al. (1994) 27 ont pu être considérées chacune comme un locus et ont été utilisées pour l'analyse des paramètres génétiques.

181 clones appartenant à quatre populations morphogéographiques ont été analysés globalement dans un premier temps. Deux types de subdivision ont ensuite été réalisées. Les clones ont été répartis en Forastero et Criollo qui, selon les hypothèses de Cuatrecasas (1964) sont originaires de deux zones géographiques distinctes: l'Amérique du Sud et l'Amérique centrale. Nous avons exclu de cette analyse spatiale, les cacaoyers Trinitario, hybrides entre Forastero et Criollo et quelques génotypes dont la classification est controversée: les génotypes Nacional et les cacaoyers Guyanais. Une autre subdivision est basée sur la classification classique en Forastero Haut Amazoniens, Forastero Bas Amazoniens, en Criollo et en Trinitario. Les paramètres des populations ont été estimés à l'aide des logiciels GENEPop (Raymond et Rousset: Adresse.: Laboratoire de génétique et environnement. URA CNRS 327. Institut des Sciences de l'évolution. CC065. USTL. Place E. Bataillon, 34095. Montpellier cedex 05, France), BIOSYS (Swofford et Selander, 1981) et DIPLOID.FOR (Weir, 1990).

Outre les fréquences alléliques, le nombre de locus polymorphes, le taux d'hétérozygotie et le test d'équilibre de Hardy Weinberg, pour les diverses populations considérées, la différenciation génétique a été testée en considérant que l'on se situe dans un cas de modèle aléatoire à partir duquel on pourrait tirer des conclusions sur le cacaoyer en général. Le test de Hardy Weinberg permet d'émettre des hypothèses sur la façon dont les gamètes sont associés au sein d'un individu en partant de l'hypothèse nulle de rencontres aléatoires. Si cette hypothèse n'est pas vérifiée alors les associations sont préférentielles et certains génotypes sont plus fréquents que d'autres, en particulier les hétérozygotes pourraient être en surnombre ou au contraire en déficit. Le taux d'hétérozygotes en condition de panmixie est encore appelé diversité de Nei ( $H$ ) et permet d'évaluer la variabilité des populations. Pour des échantillons de taille réduite,  $H$  peut être biaisé et il faut le corriger en tenant compte du nombre d'individus. Pour un locus A avec  $n$  allèles.

$$H_A = 1 - \sum_{i=1}^n p_i^2$$

$p_i$  est la fréquence de l'allèle  $i$  au locus A (Nei, 1972)

Si  $N$  est l'effectif de la population, l'estimateur non biaisé qui tient compte de la taille de l'échantillon est (Nei, 1978):

$$H_A = \frac{2N}{2N-1} \left( 1 - \sum_{i=1}^n p_i^2 \right)$$

En faisant la moyenne sur tous les locus on obtient la diversité moyenne.

Le taux d'hétérozygotes observés  $H_o$  permet d'évaluer le pourcentage d'hétérozygotes réels dans chaque population.

Le degré de différenciation entre populations est estimé par les paramètres statistiques  $F$  (indice de fixation) de Wright (1978) et il existe une relation entre l'indice de fixation de Wright et la diversité de Nei:

$$F = 1 - \frac{H_o}{H}$$

$H_o$  correspond au taux d'hétérozygotes observé et  $H$  est le taux d'hétérozygotes en condition de panmixie. A partir des valeurs des indices de fixation au sein des populations, les taux d'alofécondation ont été estimés par la formule de Nei et Syakudo (1958):

$t_{eq} = (1-F)(1+F)$  ou  $t_{eq}$  est le taux d'alofécondation à l'équilibre et  $F$  l'indice de fixation

$F$  au sein d'un échantillon mesure le déficit total en hétérozygotes et est dénommé  $Fit$ . Il permet d'estimer la corrélation entre les gamètes au sein d'un individu (Weir, 1990):

$F$  au sein d'une population est appelé  $Fis$  et mesure le déficit en hétérozygotes au sein de la population: c'est la corrélation entre gamètes à l'intérieur d'une population.

Lorsque plusieurs populations sont considérées,  $Fst$  estime les corrélations entre des gamètes des individus issus de plusieurs populations. C'est  $Fst$  qui permet de déterminer si la population globale est structurée en sous populations.

Il découle de ces définitions que lorsqu'une espèce n'est pas subdivisée en plusieurs populations,  $Fst$  ne peut pas être estimé et  $Fis$  et  $Fit$  sont confondus. Plusieurs niveaux de hiérarchie sont nécessaires à l'estimation de ces paramètres en l'occurrence , plusieurs gamètes, plusieurs individus, et plusieurs populations.

#### *Distances génétiques*

La proximité ou l'éloignement des populations peut se mesurer à partir des fréquences alléliques par la distance génétique. Plusieurs distances sont proposées

mais deux seulement sont utilisées ici . Il s'agit de la distance de Nei (1972) et de celle de Rogers modifiées (Wright 1978).

Si deux populations P1 et P2 ont des fréquences alléliques X1 et X2 pour l'allèle 1 d'un locus L, les probabilité de tirer deux fois l'allèle 1 du locus L dans l'une ou l'autre population P1 et P2 est

$$J_{P1P1} = X1^2$$

$$J_{P2P2} = X2^2$$

$J_{P1P2} = X1X2$  est la probabilité de tirer l'allèle 1 dans la population P1 et dans P2. Le coefficient de similarité de Nei est pour le locus L:

$$I = \frac{J_{P1P2}}{J_{P1P1} \times J_{P2P2}}$$

Si plusieurs locus sont considérés, le coefficient de similarité moyen est la moyenne arithmétique des différents coefficients de similarité pour les locus considérés.

La distance génétique de Nei est

$$D = -\text{Log } I$$

La distance de Nei est très utilisée car elle peut être reliée au taux de substitution en acide aminé et donc possède une signification d'un point de vue évolutif (De Vienne et Damerval, 1985).

Une autre distance a été estimée: celle de Rogers modifiée: MRD(modified Rogers distance)

$$MRD = \frac{\left[ \sum_1^n (X1_{P1} - X1_{P2})^2 \right]^{1/2}}{(2N)^{1/2}} \quad N = \text{nombre total de locus, } X1_{P1} \text{ et } X1_{P2}$$

sont les fréquence de l'allèle 1 d'un locus L dans les population P1 et P2.

## II - RESULTATS

### 1-DIVERSITÉ MORPHOLOGIQUE

#### A- DETERMINATION DE LA TAILLE MINIMUM DES ÉCHANTILLONS

Les tableaux ci dessous donnent les effectifs minimum des échantillons de cabosses et de fèves. Il y a d'une part, les caractères dont l'évaluation nécessite un échantillon de taille importante: ce sont le poids de cabosses, le nombre de fèves par cabosse et le poids de fèves et d'autre part les autres qui sont en fait des caractères de forme. Pour le poids de cabosses, la taille de l'échantillon de cabosses par arbre servant à estimer les variations entre les arbres d'un même clone est importante. La variation entre cabosses d'un même arbre est donc élevée. Ces

variations peuvent être dues à l'environnement mais aussi à la position des cabosses sur l'arbre. En effet les cabosses sont portées par le tronc principal et les branches et les cabosses situées dans les branches pourraient être défavorisées par leur position proche des feuilles au sein desquelles se réalise la synthèse des éléments de réserves qui sont distribués du bas vers le haut de l'arbre.

Au niveau des caractères de fèves, un échantillon plus important est nécessaire pour évaluer le poids de fèves pour une cabosse que pour l'évaluer au sein d'un clone. Cette apparente aberration rend en fait compte de la variabilité intracabosse élevée et qui est plus importante que celle qui existe au sein d'un clone.

Caractères	effectif de cabosses	
	par arbre	par clone
Poids de cabosse	20	49
Longueur de cabosse	4	9
Diamètre	4	8
Nombre de fèves	7	39

**Tableau 5:** Effectifs de cabosses nécessaires à l'évaluation des caractères de cabosses

Caractères	Nombre de fèves à observer		
	par cabosse	Par arbre	Par clone
Longueur de fève	4	3	5
Poids de fève	16	14	11
Epaisseur de fève	6	4	15

**Tableau 6:** Effectifs de fèves nécessaires à l'évaluation des caractères de fèves

Ces estimations d'effectifs permettent de se rendre compte des difficultés qui peuvent être rencontrées pour l'évaluation de caractères morphologiques très variables au niveau clonal, car la production de cabosses peut être insuffisante pour un arbre au cours d'une période donnée.

### **B- ANALYSES MORPHOLOGIQUES**

Les moyennes des clones pour les caractères des cabosses et des fèves sont présentées dans les annexes. Les caractères analysés ne sont pas indépendants: de fortes corrélations sont observées entre le poids de cabosse pleine de fèves et le

## ANALYSES DE VARIANCE

### 1)- Poids de cabosse

	DDL	CM	F
Inter groupe	3	4260599	118,89
Intra groupe	3265	35834,42	p=0,0001
Total	3268		

### 2)-Nombre de fèves par cabosse

	DDL	CM	F
Inter groupe	3	2130,81	16,13
Intra groupe	3238	132,07	p=0,0001
Total	3241		

### 3)-Longueur de cabosse

	DDL	CM	F
Inter groupe	3	62936,36	95,67
Intra groupe	1231	663,16	p=0,0001
Total	1234		

### 4)-Diamètre de cabosse

	DDL	CM	F
Inter groupe	3	1518,30	14,98
Intra groupe	1231	101,35	p=0,0001
Total	1234		

### 5)-Poids vide de cabosse

	DDL	CM	F
Inter groupe	3	1460051,249	55,34
Intra groupe	1197	26384,09	p=0,0001
Total	1200		

### 6)-Epaisseur de cortex

	DDL	CM	F
Inter groupe	3	10713,27	35,54
Intra groupe	1231	301,41	p=0,0001
Total	1234		

### 7)-Forme de cabosse(Longueur/Diamètre)

	DDL	CM	F
Inter groupe	3	43439,56	67,63
Intra groupe	1231	642,30	p=0,0001
Total	1234		

### 8)-Poids de fève

	DDL	CM	F
Inter groupe	3	11830,20	673,14
Intra groupe	3623	17,575	p=0,0001
Total	3626		

### 9)-Longueur de fève

	DDL	CM	F
Inter groupe	3	72886	101,88
Intra groupe	3623	715,366	p=0,0001
Total	3626		

### 10)-Epaisseur de fève

	DDL	CM	F
Inter groupe	3	56753,88	258,93
Intra groupe	3623	219,19	p=0,0001
Total	3626		

**Tableau 7:** Analyses de variance à un facteur des caractères morphologiques

## COMPARAISONS DE MOYENNES

	Haut Amazoniens	Bas Amazoniens	Trinitario	Criollo
Poids de cabosse(g)	465,11(1424)a s=176,57	376,16 (330)b s=98,23	568,12(1275)c s=223,28	514,26(219)d s=155,71
Nombre de fèves par cabosse	33,41(1424)e s=12,8	37,04(324)g s=9,56	32,14(1275)f s=10,48	32,47(219) ef s=10,72
Longueur (cm) de cabosse	16,05 (546)h s=2,44	14,03 (119)i s=1,40	17,9 (455)j s=2,74	17,72 (115)j s=3,33
Diamètre (cm) de cabosse	8,36(547)k s=1,09	8,03 (119)l s=0,66	8,60 (454) m s=1,03	8,6 (115) km s=0,72
Poids vide (g) de cabosse	383,88 (513)n s=152,48	295,33 (119)o s=72,05	484,3 (454)p s=196,3	415,38 (115)n s=118,27
Epaisseur cortex (mm)	4,93(547) qr s=1,42	3,16(119)s s=1,06	4,85(454) qt s=2,20	4,86 (115) rt s=1,503
Forme de cabosse (longueur /diamètre)	1,92(546)u s=0,21	1,75(119)v s=0,19	2,08(455) w s=0,28	2,05(115) w s=0,31
Poids de fève(g)	1,59(1617)x s=0,36	1,53(402)y s=0,25	2,12(1308)z s=0,46	2,46(300)a' s=0,66
Longueur de fève(cm)	2,30(1617) b' s=0,26	2,30(402)b' s=0,2	2,43(1308)c' s=0,28	2,52(300)c' s=0,29
Epaisseur de fève(mm)	7,94(1617) e' s=1,49	7,75(402) e' s=1,12	9,19(1308)f' s=1,55	9,60 (300)g' s=1,5

Tableau 8: Comparaisons des moyennes des différents groupes par le test de Fisher de la plus petite différence significative (PPDS). Le nombre d'observations par caractère est entre parenthèse et les moyennes non significativement différentes portent la même lettre. s=écart type.

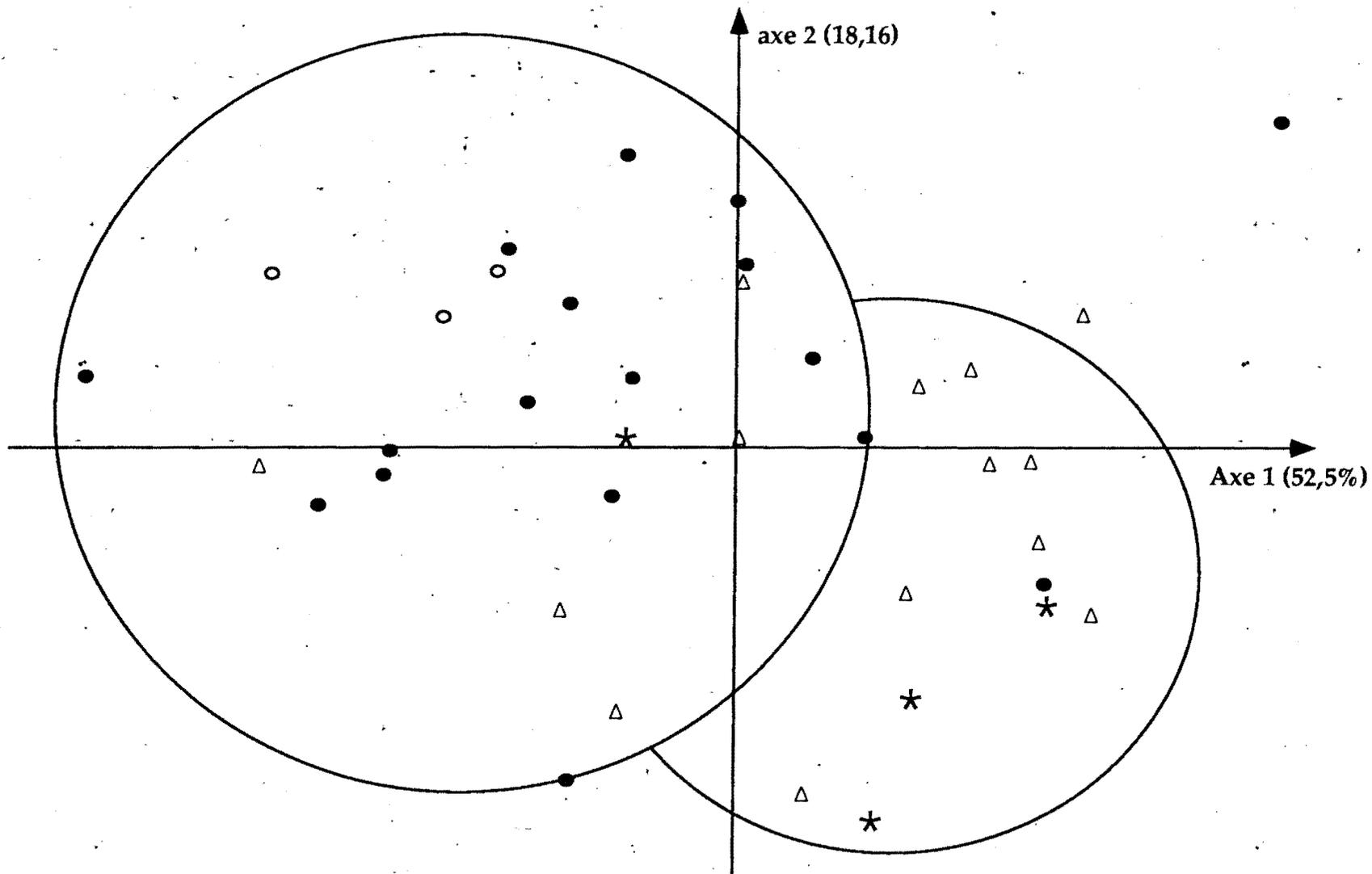


Figure 2: Plan 1 (axes 1 et 2) de l'ACP réalisée à partir des données morphologiques

- Forastero Haut Amazoniens
- Forastero Bas Amazoniens
- \* Criollo
- △ Trinitario

poids de cabosse vide ( $r=0,95$ ), le poids de cabosse et son diamètre (0,88), entre le diamètre de la fève et la longueur de fève (0,83).

Les analyses de variance ainsi que les comparaisons des moyennes faites sur les caractères précédents en considérant chaque population comme un facteur, sont présentées dans les tableaux 7 et 8. Tous les tests sont significatifs à 1 pour 10 000.

Au niveau des cabosses, les poids diffèrent dans les quatre populations considérées: les cabosses les plus grosses sont celles des Trinitario et celles des Forastero Bas Amazoniens sont les plus petites. Par contre, il n'y a pas de différence entre le nombre de fèves par cabosses des Forastero Haut Amazoniens, des Criollo et des Trinitario. Seuls les Forastero Bas Amazoniens peuvent être caractérisés par un nombre élevé de fèves dans la cabosse. La forme des cabosses mesurée par le rapport de la longueur sur le diamètre fait apparaître trois formes différentes. Celle des Forastero Bas Amazoniens, qui est la moins allongée, celle des Forastero Haut Amazoniens et celle des Criollo et Trinitario, qui est la plus allongée. Les Forastero Bas Amazoniens possèdent des cabosses au cortex fin contrairement au Criollo, Trinitario et Forastero Haut Amazoniens dont les cabosses sont épaisses. Au niveau des fèves, la taille des fèves diffère entre populations et les plus grosses fèves sont rencontrées chez les Criollo. Les fèves des Forastero ont la même forme, c'est à dire allongées et aplaties alors que celles des Trinitario et des Criollo sont dodues. Ce sont les Criollo qui possèdent les fèves les plus dodues. Qualitativement, l'on peut décrire les populations en disant que les Forastero Bas Amazoniens possèdent de petites cabosses peu allongées à cortex fin contenant beaucoup de fèves allongées et aplaties. Les Forastero Haut Amazoniens ont des cabosses de taille moyenne à cortex épais contenant peu de fèves allongées et aplaties. Les Criollo et les Trinitario ont de grosses cabosses allongées à cortex épais contenant peu de fèves grosses et dodues. Si l'on quantifie un caractère donné dans les différents groupes, on observe alors une variation continue de ce caractère entre les groupes .

L'ACP (Figure 2) permet l'analyse simultanée des caractères et permet de mieux visualiser les relations entre populations. Les axes les plus discriminants sont l'axe 1 avec 52,5 % de l'inertie totale et l'axe 2 avec 18,16% de la variabilité . Les variables qui contribuent le plus à la construction du premier axe sont le poids de cabosse, le poids de cabosse vide et le diamètre de cabosse: l'axe 1 peut être assimilé à l'axe des cabosses. L'axe 2 est constitué par le nombre de fèves par cabosse, et la taille des fèves. L'axe 3 quant à lui représente 11 % de la variabilité et est constitué par l'épaisseur du cortex. Ainsi l'axe 1 oppose les clones Forastero aux petites cabosses aux Trinitario et Criollo. Globalement, les Forastero Haut Amazoniens et les Forastero Bas Amazoniens sont regroupés d'une part et les

Criollo et les Trinitario sont regroupés d'autre part (Figure 2). Sur le plan 2 constitué par les axes 1 et 3, les regroupements sont du même type que ceux sur le plan 1, c'est à dire une séparation des Forastero des Criollo.

Dans cette étude où certaines populations ont des effectifs faibles, la séparation des différents groupes est moins nette que l'étude de Engels (1985). Bien que moins efficaces que les marqueurs moléculaires, les marqueurs morphologiques restent tout de même un outil précieux pour structurer globalement la diversité. Ils ne nécessitent pas beaucoup d'investissements mais par contre ils demandent beaucoup de temps. De plus comme nous l'avons montré, il n'est pas toujours facile d'obtenir l'échantillon qu'il faut pour estimer certains caractères.

## 2-DIVERSITÉ RÉVÉLÉE PAR LES MARQUEURS RAPD

### A- MISES AU POINT MÉTHODOLOGIQUES DE LA TECHNIQUE DE RAPD

Plusieurs facteurs influencent la réussite de l'amplification génique *in vitro* de l'ADN comme vu précédemment. Afin d'obtenir un maximum de robustesse et de répétabilité, certains facteurs ont été testés pour déterminer les conditions optimales d'amplification.

#### Effet de la qualité de l'ADN sur les produits d'amplification

Un des problèmes rencontrés au cours des études RFLP sur le cacaoyer est la qualité de l'ADN extrait des feuilles séchées ou fraîches. Les feuilles de cacaoyer comme chez certaines plantes contiennent une quantité importante de polyphénols et de polysaccharides qui ne sont pas complètement éliminés lors des extractions. Pour obtenir de l'ADN propice aux analyses RFLP, il faut une étape finale de purification au chlorure de césium, aboutissant à l'obtention de faibles quantités d'ADN (Laurent *et al.*, 1993). Ces extractions rallongent considérablement la durée des analyses. Pour des analyses RAPD faites chez d'autres espèces, il ne semblerait pas nécessaire d'obtenir une très grande pureté de l'ADN. Néanmoins nous avons testé le facteur qualité de l'ADN:

L'ADN matrice à partir duquel les réactions d'amplifications ont été réalisées a été extrait par deux méthodes.

-Méthode Dellaporta modifiée (voir annexe)

-Méthode Dellaporta suivie d'une purification par ultracentrifugation en présence de Chlorure de césium.

Trois clones (H1258, H1380, H1258 x H1380) et deux amorces (D02 et D05) ont été utilisés avec plusieurs concentrations d'ADN. Nous n'avons pas observé

de différence entre les deux qualité d'ADN au niveau de la qualité des bandes ou de leur quantité.

#### Effet de la quantité d'ADN

Les concentrations d'ADN dans les solutions ont été estimées par plusieurs méthodes. Une première estimation a été faite avec un spectrophotomètre ou un fluorimètre puis un gel a été réalisé avec en témoin des concentrations connues de lambda. Huit concentrations d'ADN extrait par la méthode Dellaporta modifiée ont été testées sur deux clones P7 et H1380 avec deux primers, D02 et D05 et de la Taq polymérase provenant de deux fabricants différents. Ces concentration sont: 0,5 ng; 2,5 ng; 5 ng; 25 ng; 50 ng; 125 ng; 250 ng; 500 ng. Les résultats variaient selon les clones et selon les concentrations: pour le clone H1380 à partir de 50 ng et au delà nous n'avions pas de produits d'amplification. Avec le clone P7, les produits d'amplifications de l'amorce D02 étaient stables pour toutes les concentrations de 0,5 ng à 500 ng par contre ceux de l'amorce D05 étaient stables à partir de 5 ng. Dans notre cas, les quantités d'ADN matrice trop faibles ont provoqué une instabilité des produits d'amplification. La quantité d'ADN matrice a donc été fixée à 2,5 et 5 ng par réaction de 25 µl.

#### Effet de la qualité et de la quantité de la Taq polymérase

Les enzymes Taq testées provenaient de chez Boehringer, Promega, Amersham, Beckman et USB. Les amplifications réalisées avec ces enzymes donnaient sensiblement les mêmes résultats pour toutes ces enzymes sauf pour l'enzyme USB qui ne donnait pas de produits d'amplification du tout. Quelques différences sont apparues au niveau de la qualité des bandes. La comparaison entre Taq Promega et Taq Boehringier a donné les mêmes résultats pour les bandes majeures et quelques bandes mineures.

Deux concentrations d'ADN sur deux clones et 5 concentrations de Taq polymérase ont été testées: 0,5 U; 1 U; 1,5 U; 2 U; 3 U (U = unité) par réaction de 25 µl.

Avec 3 unités d'enzymes, un bruit de fond important était obtenu alors qu'avec 1,5 et 2 unités le bruit de fond était moindre. Avec 0,5 et 1 unité, tous les fragments n'étaient pas amplifiés. Les deux types d'enzymes donnaient globalement les mêmes résultats. Aussi, 1,5 unité d'enzyme ont-elle été choisie pour la suite des analyses. La Taq promega a été retenue pour la suite des travaux à cause de son coût plus faible et de son efficacité.

#### Effet de la quantité de Magnésium

Au départ, l'enzyme de Promega choisi pour la suite des analyses était livré avec un tampon de réaction contenant 1,5 mM de MgCl<sub>2</sub>. Puis la solution de magnésium a été livré séparément. C'est alors que nous avons testé quatre

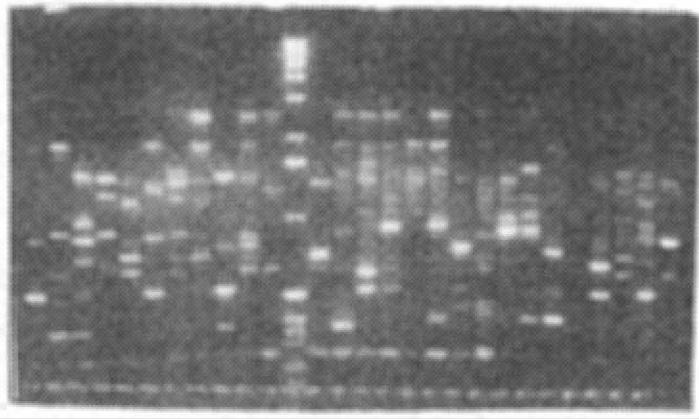


Photo 4a: Profil complexe obtenu avec le primer D12 sur quelques génotypes de cacaoyers

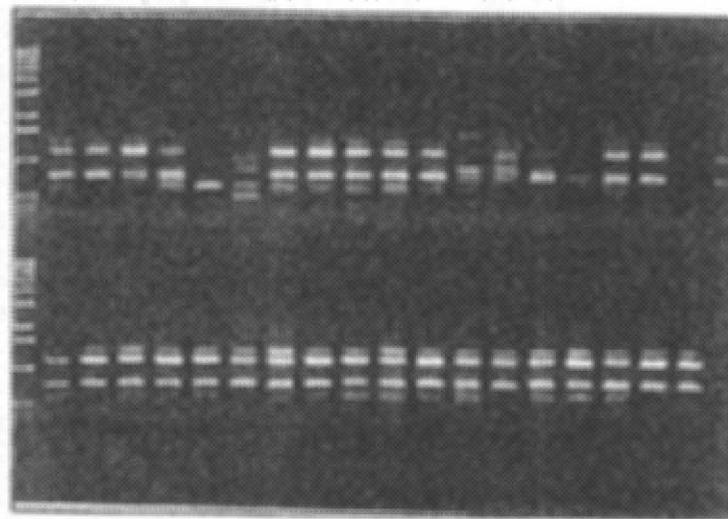


Photo 4b: Profil simple obtenu avec le primer D08 sur quelques génotypes de cacaoyers

concentrations de  $MgCl_2$  : 1,5 mM; 2mM; 2,5 mM, 3 mM avec une amorce (D15) sur cinq clones:

Avec 1,5 mM de  $MgCl_2$ , certaines bandes étaient absentes chez certains clones et avec 3 mM de  $MgCl_2$ , chez certains clones un important bruit de fond apparaissait. Aussi la concentration de  $MgCl_2$  a-t-elle été fixée à 2 mM par réaction.

#### **Effet de la température d'hybridation**

Les températures d'hybridation testées étaient 25°C, 36°C, 40°C, 45°C et 50°C. Il n'y a pas eu d'amplification à 50°C, et à 45 °C le nombre de produits d'amplification obtenus était considérablement réduit. De 40°C à 25°C de nombreux produits étaient amplifiés indiquant une diminution de spécificité avec la diminution de température. A 36°C, il y avait suffisamment de produits et peu de bruit de fond. La température de 36°C préconisée par Williams *et al.* (1990) a donc été maintenue.

#### **Répétabilité**

Lorsque toutes les conditions d'amplifications ont été mises au point, les réactions ont été reproduites plusieurs fois afin de tester leur répétabilité. Certaines amorces ont engendré des produits peu répétables et ceci, indépendamment de l'importance quantitative des bandes obtenues: des bandes mineures aussi bien que des bandes majeures pouvaient s'avérer instables. Cette situation n'était apparemment pas due à un faible pourcentage en GC car toutes les amorces utilisées avaient un pourcentage en GC supérieur à 50 %. La concentration d'ADN adaptée aux amplifications réalisées est de 2,5 ng pour un volume de réaction de 25 µl mais pour certains génotypes il a fallu doubler cette quantité. Il semble que même avec la purification réalisée au cours de l'extraction de l'ADN, certains génotypes gardent une forte quantité d'impuretés qui influe sur le dosage. Certaines amorces ont donné un nombre assez limité de fragments et d'autres par contre ont généré de très nombreux fragments qui étaient à la limite du lisible ( photo 4 ).

### **B- NATURE DES FRAGMENTS AMPLIFIÉS**

Plusieurs fragments amplifiés ont été isolés et hybridés sur l'ADN de cacao digéré par plusieurs enzymes de restriction. Plusieurs types de profils d'hybridation ont été obtenus (photo 5): un ensemble de fragments RAPD utilisés comme sondes ont généré des profils d'hybridation à bandes nettes, sans bruit de fond. Lorsqu'au maximum trois bandes étaient révélées par génotype, les sondes RAPD correspondantes étaient classées comme sondes uniques. Lorsque les autoradiographies révélaient un smear, les sondes ont été classées comme



répétées dispersées et si le nombre de signaux était supérieur à trois mais distinguables et dénombrables, les sondes étaient classés comme faiblement répétées. Un certain nombre de sondes ont donné naissance à des signaux avec un important bruit de fond. Une même amorce pouvait générer aussi bien des fragments uniques que des fragments répétés. Au total, 12 produits d'amplifications ont été considérés comme uniques, 12 comme hautement répétés et 12 comme faiblement répétés et quelques fragments sont restés indéterminés.

### C- FRÉQUENCE DES BANDES RAPD DANS LES POPULATIONS

La taille des fragments amplifiés variait de 0,3 kb à 2 kb (Tableau 9). Des différences de fréquences de bandes entre populations ont été observées. Aucune bande n'était vraiment spécifique d'une population donnée. Cependant, la fréquence de certaines bandes était tellement faible dans certaines populations qu'elles pourraient être considérées comme absentes si l'on considère que ces faibles fréquences peuvent être dues à des problèmes de mauvaise classification. Ainsi, les fragments L05/0,7 et 0,9 peuvent être considérés comme spécifiques aux Forastero et les fragments P03/1,5 et M20/0,6 comme spécifiques aux Criollo et Trinitario. Le test d'homogénéité de la présence des bandes dans les différentes populations s'est révélé positif pour certaines bandes mais essentiellement négatif pour la majorité des bandes car de nombreuses différences ont été trouvées.

	FBA	Criollo	Trinitario
FHA	15	16	18
FBA		21	14
Criollo			16

Tableau 10 : Nombre de bandes RAPD ayant des fréquences différentes entre populations (FHA= Forastero Haut Amazoniens, FBA= Forastero Bas Amazoniens)

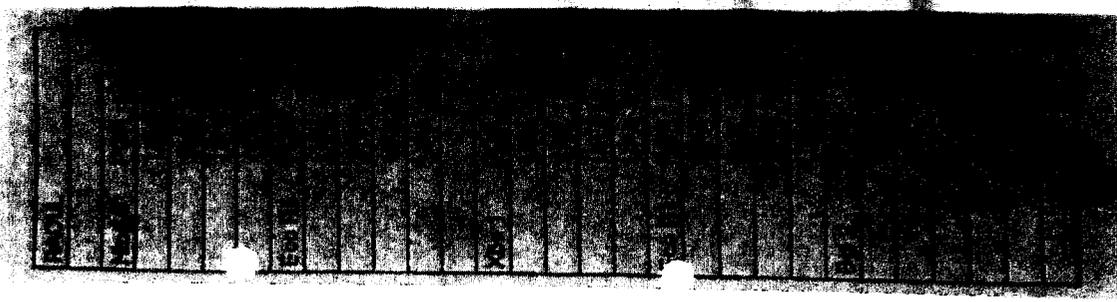
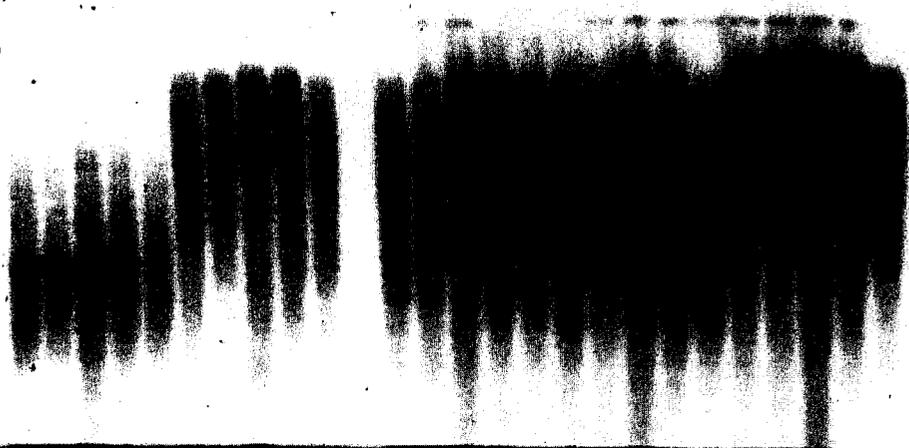
Les comparaisons faites entre les groupes pris deux à deux ont donné les résultats présentés dans le tableau 10. Par ailleurs 23 bandes RAPD ont des fréquences différentes entre les populations Forastero et Criollo (tableau 9 et figure 3). Ces bandes sont répétées ou uniques. Parmi les 17 bandes de nature connue à fréquences différentes entre Forastero et Criollo, 13 bandes sont répétées et quatre sont uniques. Par contre parmi les 26 bandes à fréquence semblable, 17 sont de nature connue et il y a autant de bandes uniques (8) que de bandes répétées (9).

Fragment RAPD	nature	Pourcentage de présence dans les populations				Niveau de signification des comparaisons deux à deux					
		BA	C	HA	T	BA-C	BA-HA	BA-T	C-HA	C-T	HA-T
rOPD15/0.5	-	8	19	40	4	ns	**	ns	ns	ns	**
rOPD15/1.0	-	24	38	76	46	ns	**	ns	**	ns	*
rOPD15/1.2	U	84	92	88	74	ns	ns	ns	ns	ns	ns
rOPD15/1.4	-	72	73	68	70	ns	ns	ns	ns	ns	ns
rOPD15/1.8	U	16	92	36	67	**	ns	**	**	*	*
rOPK07/0.3	U	27	41	42	61	ns	ns	**	ns	ns	ns
rOPK07/0.5	R	8	78	38	64	**	*	**	**	ns	ns
rOPK08/0.3	U	17	60	29	46	**	ns	ns	*	ns	ns
rOPK08/1.5	-	21	13	29	15	ns	ns	ns	ns	ns	ns
rOPK09/0.6	FR	100	84	84	93	*	*	ns	ns	ns	ns
rOPK09/1.3	FR	36	32	40	52	ns	ns	ns	ns	ns	ns
rOPK13/0.4	R	85	69	75	74	ns	ns	ns	ns	ns	ns
rOPK13/1.5	R	31	77	33	67	**	ns	*	**	ns	*
rOPK15/0.9	FR	35	78	71	29	**	*	ns	ns	**	**
rOPK15/1.4	I	23	74	71	14	**	**	ns	ns	**	**
rOPL02/0.3	U	83	39	33	92	**	**	ns	ns	**	**
rOPL02/1.0	R	15	35	25	24	ns	ns	ns	ns	ns	ns
rOPL02/1.4	U	65	96	92	64	**	**	ns	ns	**	**
rOPL05/0.7	-	35	4	68	18	**	*	ns	**	ns	**
rOPL05/0.9	U	15	0	20	7	*	ns	ns	*	ns	ns
rOPL11/0.6	R	28	88	22	60	**	ns	*	**	*	*
rOPL11/0.8	R	16	88	22	56	**	ns	**	**	**	*
rOPL11/0.9	-	76	56	78	65	ns	ns	ns	ns	ns	ns
rOPL19/0.4	FR	48	96	82	74	**	*	ns	ns	*	ns
rOPL19/0.5	U	28	50	68	48	ns	**	ns	ns	ns	ns
rOPL19/0.8	I	88	96	83	81	ns	ns	ns	ns	ns	ns
rOPM17/0.6	U	81	65	56	82	ns	ns	ns	ns	ns	*
rOPM17/1.1	U	81	62	80	32	ns	ns	**	ns	ns	**
rOPM17/2.1	I	73	85	84	61	ns	ns	ns	ns	*	*
rOPM20/0.5	FR	64	96	87	88	**	ns	*	ns	ns	ns
rOPM20/0.6	FR	12	74	13	33	**	ns	ns	**	**	ns
rOPM20/1.8	FR	80	96	78	58	ns	ns	ns	ns	**	ns
rOPO04/0.9	-	38	57	45	70	ns	ns	*	ns	ns	ns
rOPO04/1.0	-	75	57	95	80	ns	ns	ns	**	ns	ns
rOPO04/1.4	-	83	86	85	70	ns	ns	ns	ns	ns	ns
rOPO13/0.8	FR	40	55	58	54	ns	ns	ns	ns	ns	ns
rOPO13/0.9	R	86	91	88	96	ns	ns	ns	ns	ns	ns
rOPO13/1.4	U	82	64	42	92	ns	**	ns	ns	*	**
rOPO15/0.5	FR	15	96	67	57	**	**	**	**	**	ns
rOPO15/0.6	U	19	26	25	50	ns	ns	*	ns	*	*
rOPO15/2.0	FR	15	24	4	39	ns	ns	ns	*	ns	**
rOPP03/0.8	-	48	23	58	14	*	ns	**	**	ns	**
rOPP03/1.5	I	77	4	8	54	**	**	*	ns	**	**
rOPP14/0.4	R	19	92	17	59	**	ns	**	**	**	**
rOPP14/0.6	R	27	92	35	52	**	ns	ns	**	**	ns
rOPP14/0.9	R	23	29	30	41	ns	ns	ns	ns	ns	ns
rOPP16/0.6	R	42	89	76	82	**	*	**	ns	ns	ns
rOPQ15/0.3	I	84	77	67	89	ns	ns	ns	ns	ns	ns
rOPQ15/0.5	R	56	54	46	52	ns	ns	ns	ns	ns	ns

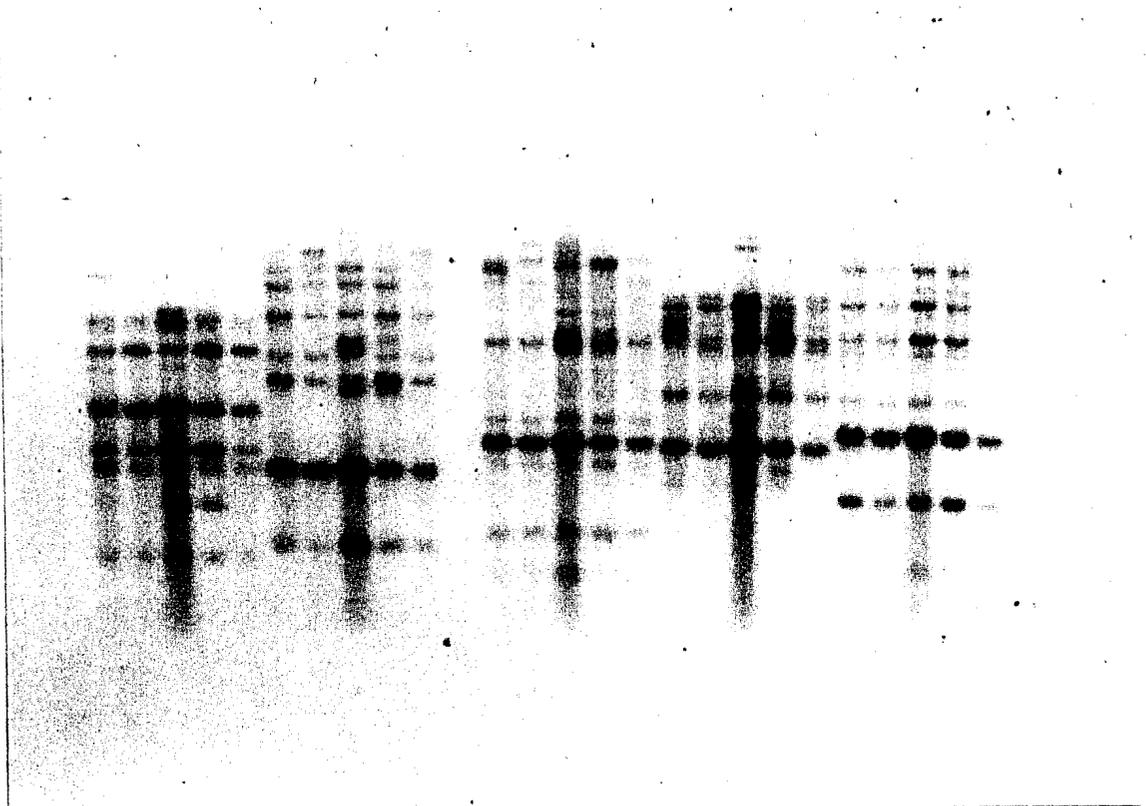
**Tableau 9:** Fréquence des bandes RAPD dans les populations et comparaisons de fréquences par le test du Chi-deux: HA= Forastero Haut Amazonien: BA= Forastero Bas Amazonien: C=Criollo T= Trinitario ns= non significatif \* significatif à 0,05. \*\* significatif à 0,01. U= sonde unique, R=sonde répétée, FR= sonde faiblement répétée, I= nature indéterminée.



**Photo 5a: Profil d'hybridation d'une bande RAPD (rOPM17/0.6) correspondant à une séquence unique amplifiée**



**Photo 5b: Profil d'hybridation d'une bande RAPD (rOPK13/0.4) correspondant à une séquence répétée amplifiée**



**Photo 5c: Profil d'hybridation d'une bande RAPD (rOPO15/0.8)  
correspondant à une séquence moyennement répétée amplifiée**

## **D- ORGANISATION DE LA DIVERSITÉ GÉNÉTIQUE RÉVÉLÉE PAR LES RAPD**

### **Analyse factorielle des correspondances**

Seuls les 48 individus ayant des données complètes ont participé à la construction des axes factoriels (individus actifs). Les autres génotypes ont été représentés dans les différents plans en tant qu'individus supplémentaires. Le plan 1 de l'AFC représente 24 % de la variabilité totale (Figure 4). Il sépare distinctement les Forastero Bas Amazoniens, des Forastero Haut Amazoniens sur l'axe 1 et les Criollo sur l'axe 2. La majorité des Trinitario occupe une position intermédiaire entre les Criollo et les Forastero Bas Amazoniens. La plupart des Trinitario américains, comportant probablement un fond génétique Criollo plus important, se retrouve proche des Criollo sur l'axe 2 ; d'autres, de nature plus hybride sans doute, occupent une position réellement médiane entre les Forastero et les Criollo. Quant aux Trinitario Africains, ils sont intermédiaires entre les Criollo et les Forastero Bas Amazoniens. Au sein des Criollo, quelques génotypes, ZEA206, Hernandez 212, CHUAO49 et OC61 se distinguent par une position extrême. Au sein des Forastero Bas Amazoniens, les génotypes Nacional(n), ECNR (h), EET59(l), ERJOH(g) et SF23 se retrouvent parmi les Forastero Haut Amazoniens, et les Guyanais (e) se retrouvent entre les Forastero Haut Amazoniens et les Forastero Bas Amazoniens. Deux génotypes Forastero, IFC4 et P32 A se trouvent en dehors des groupes constitués tandis que SPA5 et SPA17 se comportent davantage comme des Trinitario sur le plan de AFC. Les groupes Criollo, Forastero Haut Amazoniens, et Forastero Bas Amazoniens présentent sur le plan 1 le même degré de variabilité alors que les Trinitario semblent plus variables.

Le plan 1 de l'AFC réalisée à partir des données sur les bandes amplifiées répétées constitue 38, % de la variabilité (Figure 5a). Une séparation des groupes Forastero et Criollo est obtenue sur l'axe 1. On observe par ailleurs une grande variabilité des deux groupes. Les Trinitario sur l'axe 1 se regroupent soit avec les Forastero, soit avec les Criollo. Au niveau des séquences uniques amplifiées, une image différente de la variabilité est observée (Figure 5b). En effet sur le plan 1 de l'AFC représentant 33% de la variabilité, aucune distinction des groupes ne peut être mise en évidence. Les 12 séquences répétées amplifiées des RAPD ont ainsi mieux structuré la variabilité que les 12 séquences uniques amplifiées.

### **Classification des génotypes**

#### **Classification sur tableau de distances**

La figure 6 représente le phénogramme obtenu à partir des indices de Czekanowski. A un niveau élevé de la classification on sépare trois groupes: un

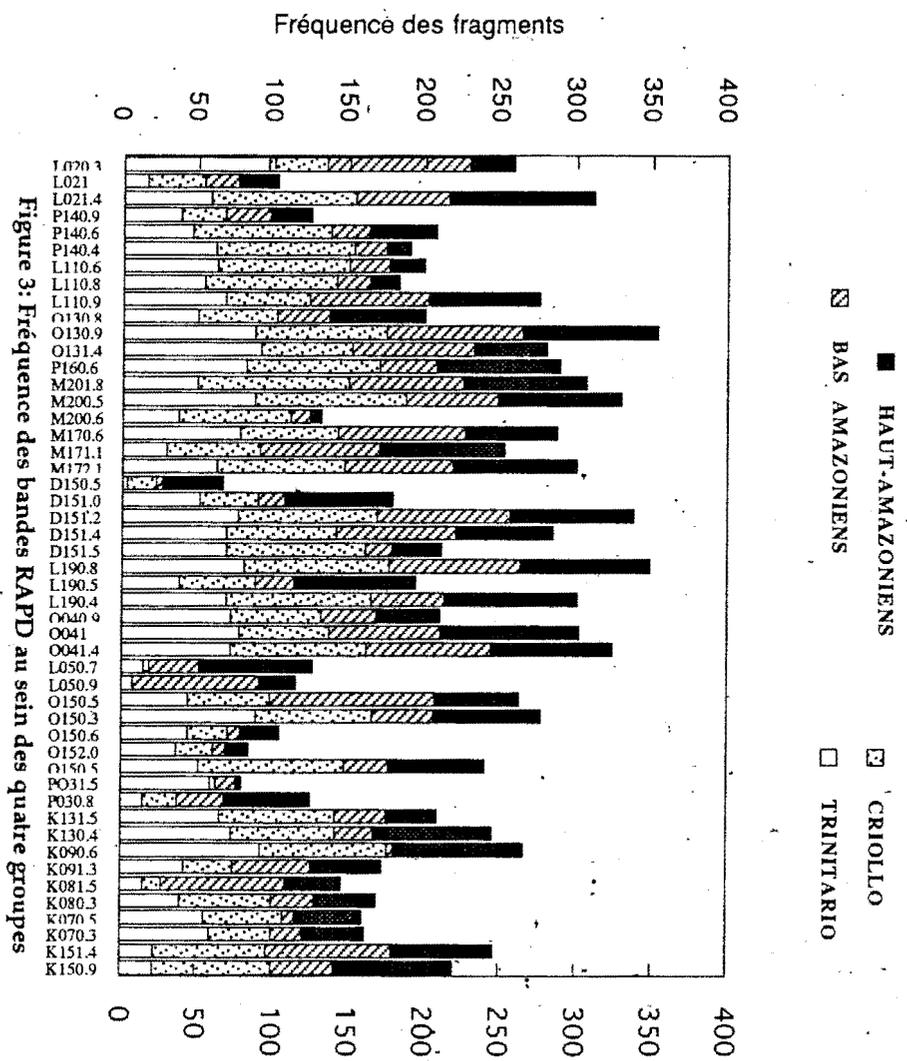


Figure 3: Fréquence des bandes RAPD au sein des quatre groupes

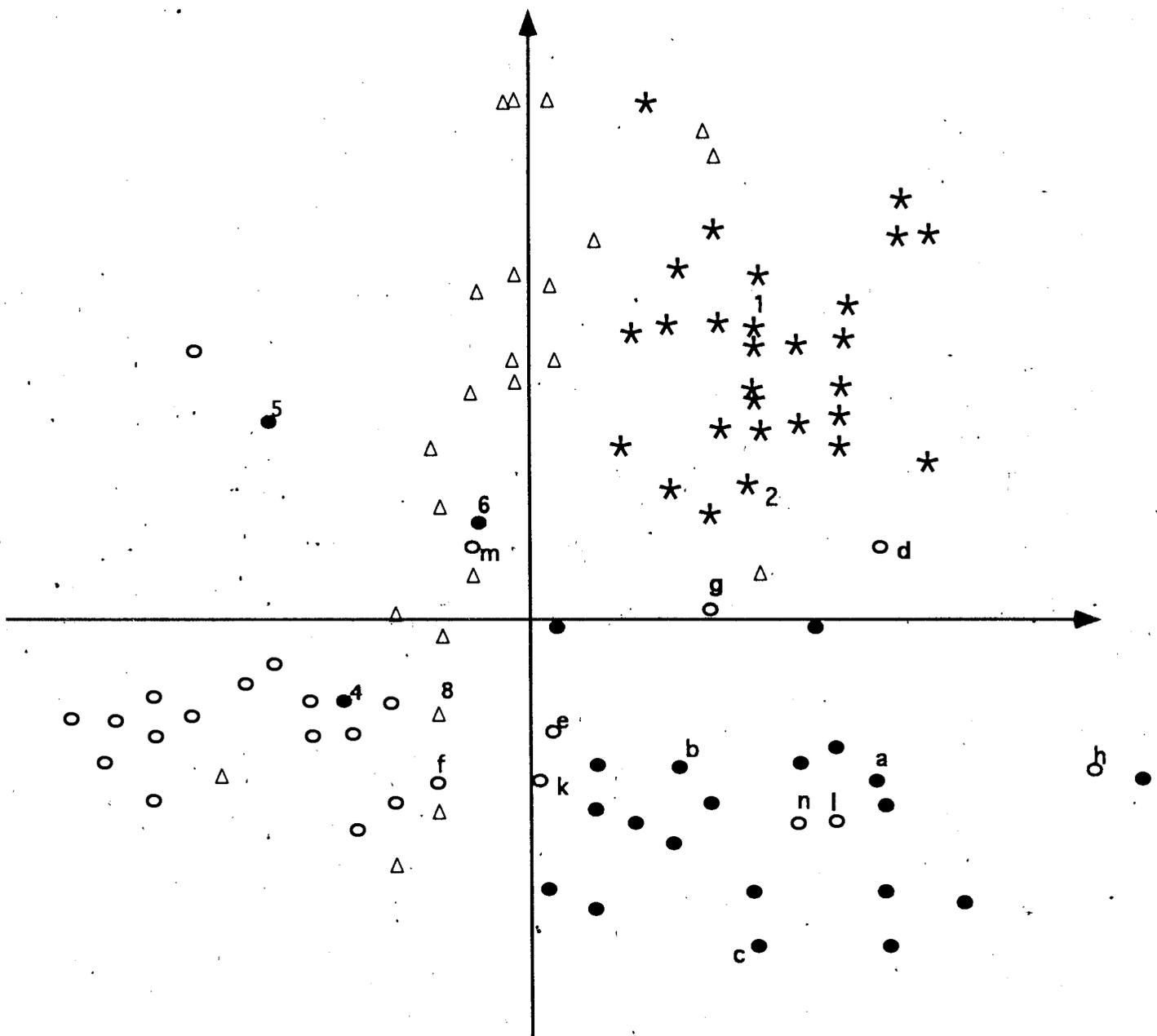


Figure 4: Plan 1 (24%) de l'AFC sur les données RAPD

- Forastero Haut Amazoniens
- Forastero Bas Amazoniens
- ★ Criollo
- △ Trinitario

- |            |          |           |             |
|------------|----------|-----------|-------------|
| 1- PV2     | 7- IFC4  | e- GU349  | k-ERJOH6    |
| 2- LAF3    | 8- MT1   | f- GU144  | l- EET59    |
| 3- SPEC185 | a- SCA6  | g-ERJOH5  | m-ERJOH1    |
| 4- P2      | b- SCA9  | h- ECNR   | n- nacional |
| 5- P32A    | c- SCA12 | i- SPA5   |             |
| 6- SPA17   | d- SF23  | j- ERJOH4 |             |

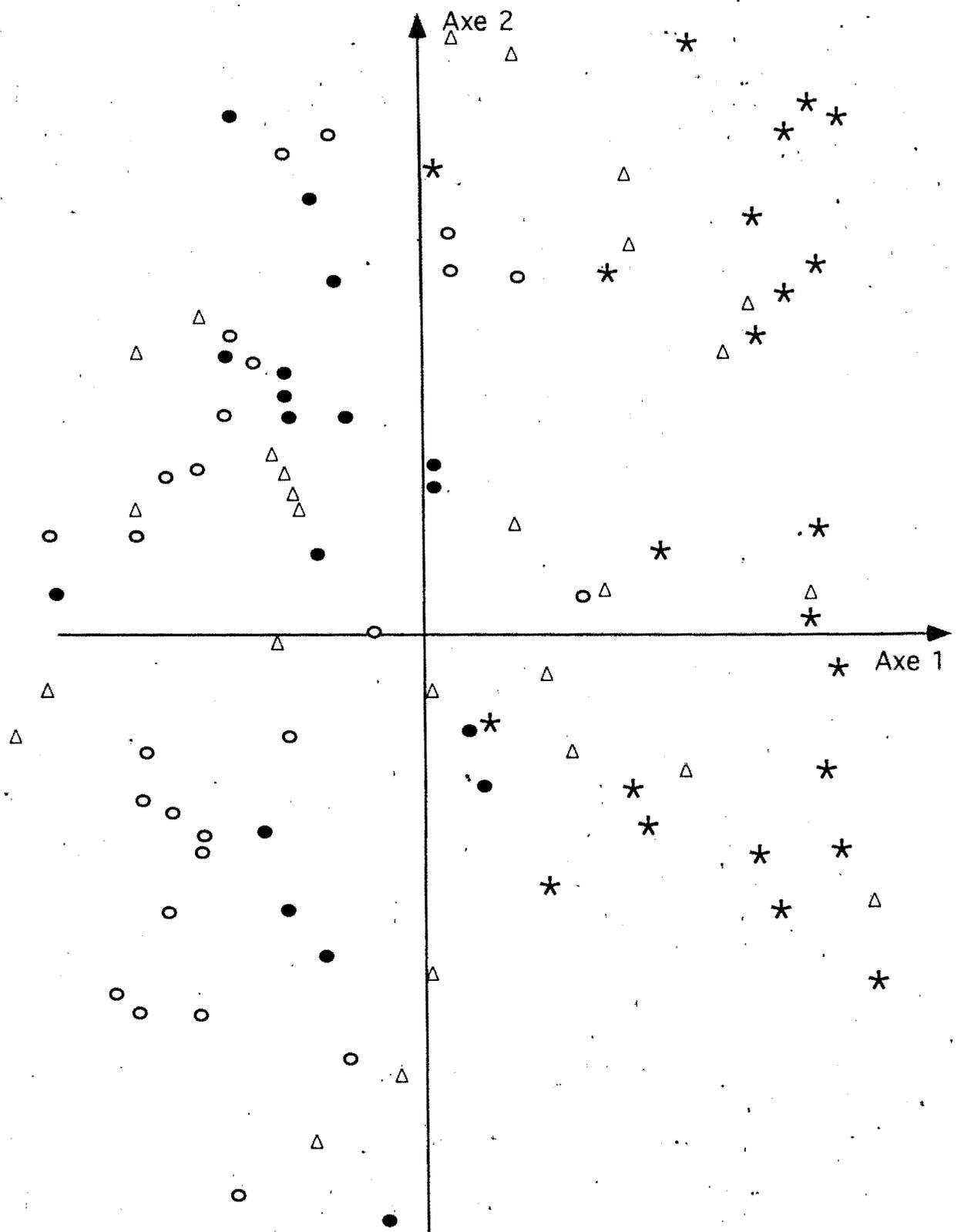


Figure 5a: Plan 1 de l'AFC(38% de variabilité) basée sur 12 sequences répétées amplifiées isolées à partir des bandes RAPD:

- Forastero Haut Amazoniens
- Forastero Bas Amazoniens
- △ Trinitario
- \* Criollo

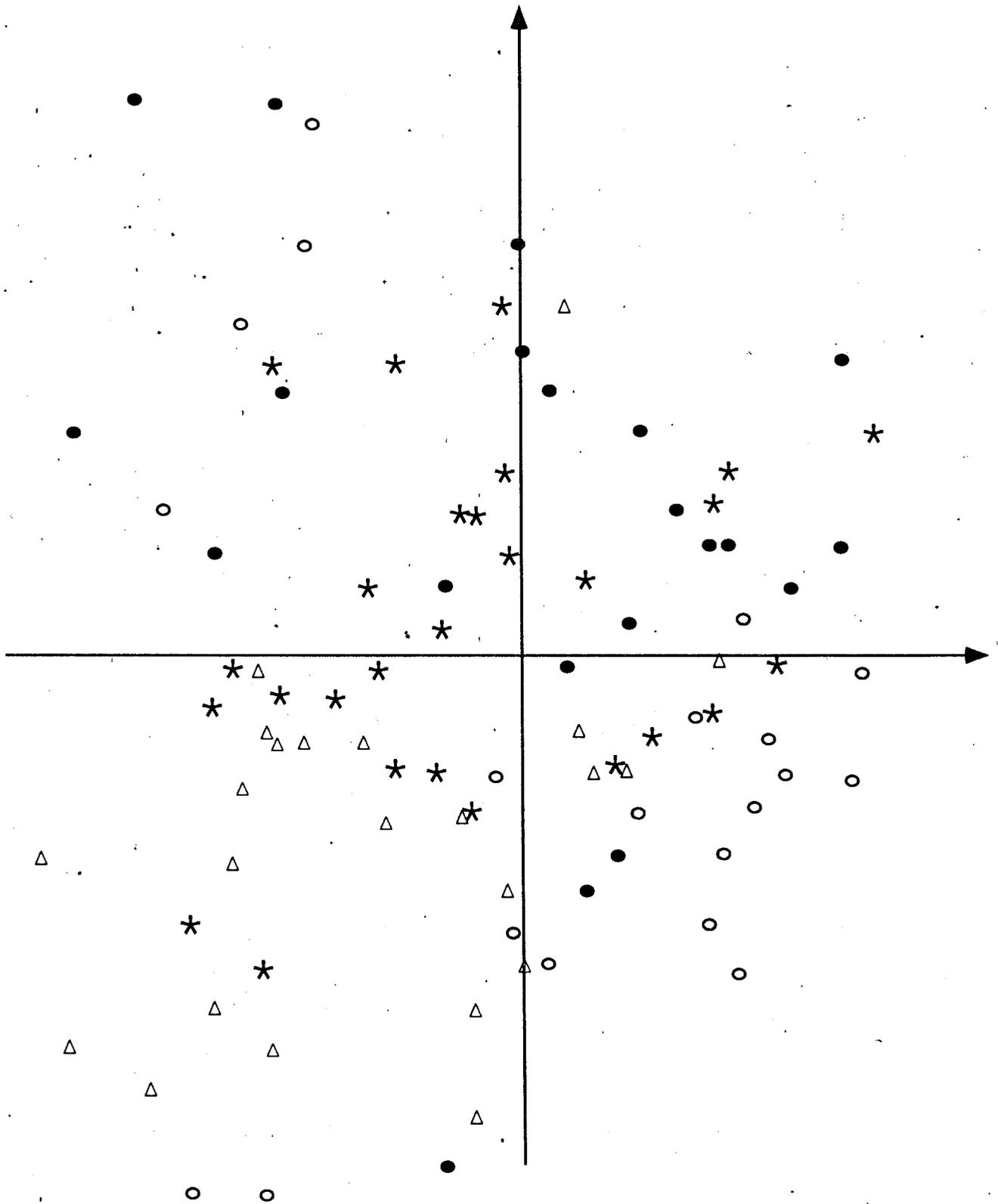
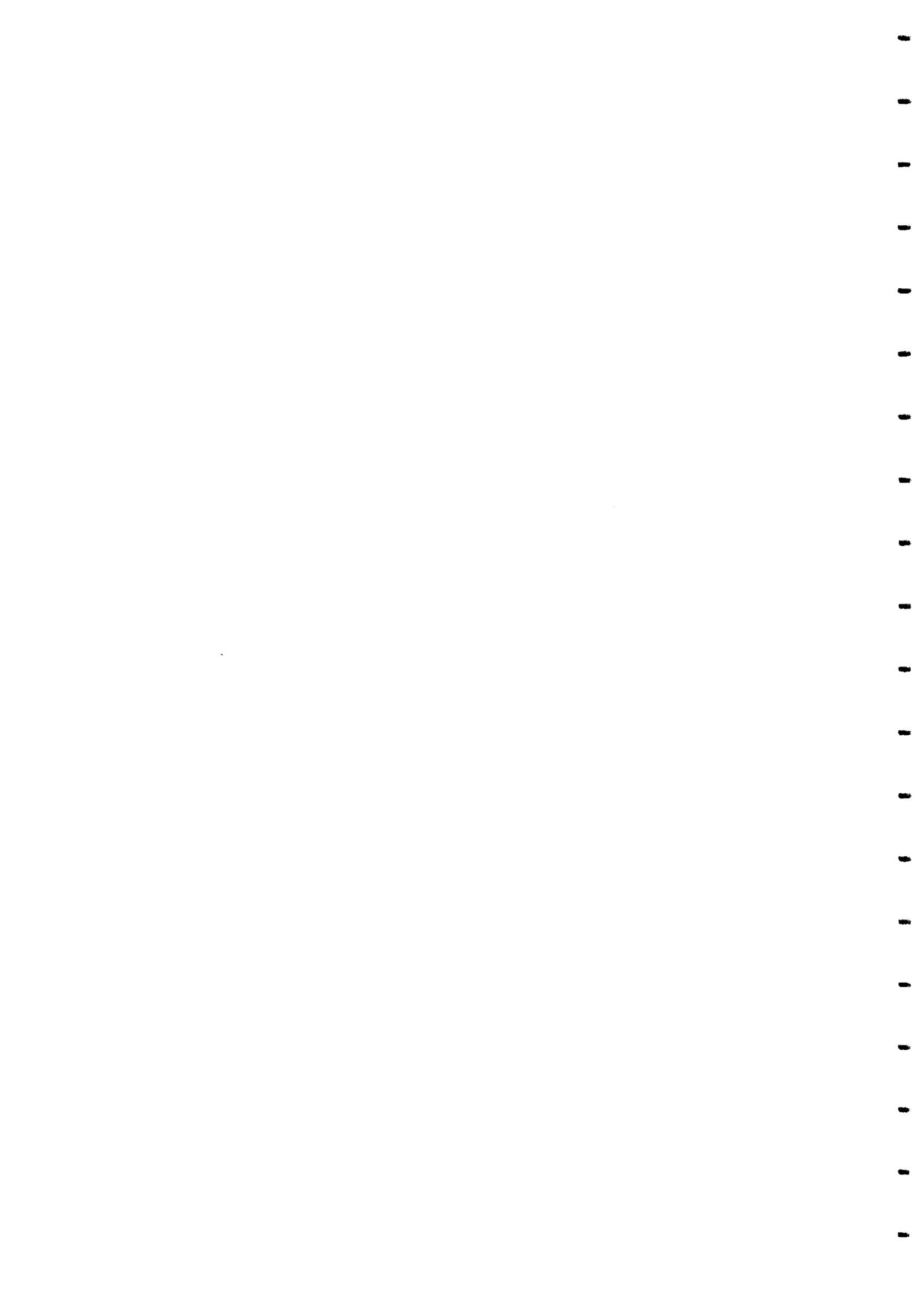


Figure 5b: Plan 1 de l'AFC(33% de variabilité) faite à partir des 12 sequences uniques amplifiées isolées à partir des bandes RAPD

- Forastero Haut Amazoniens
- Forastero Bas Amazoniens
- △ Trinitario
- \* Criollo



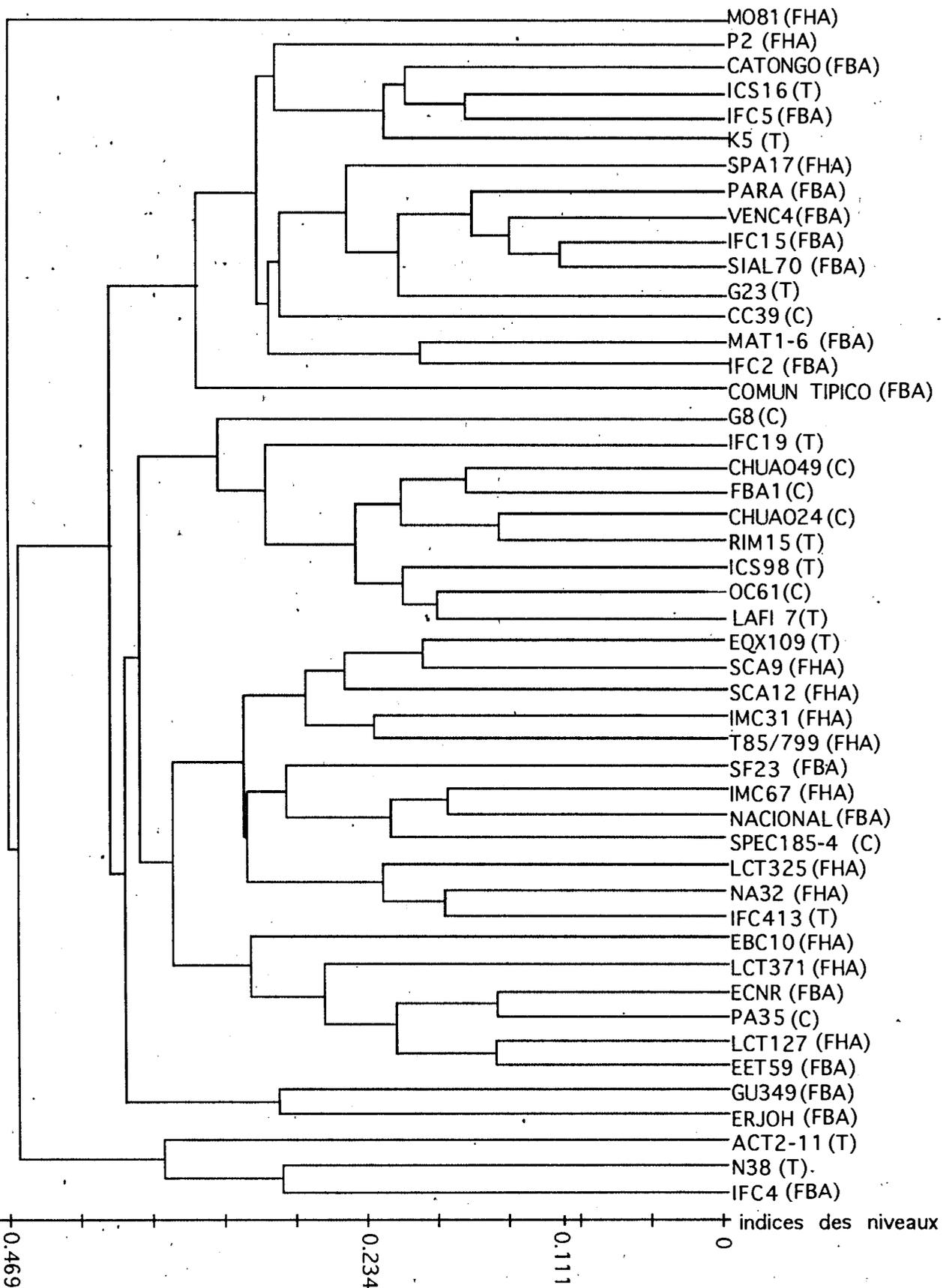


Figure 6: Phénogramme obtenu au moyen des indices de CZEKANOWSKI sur les données RAPD

FHA=Forastero Haut Amazoniens

FBA=Forastero Bas Amazoniens

T=Trinitario

C=Criollo

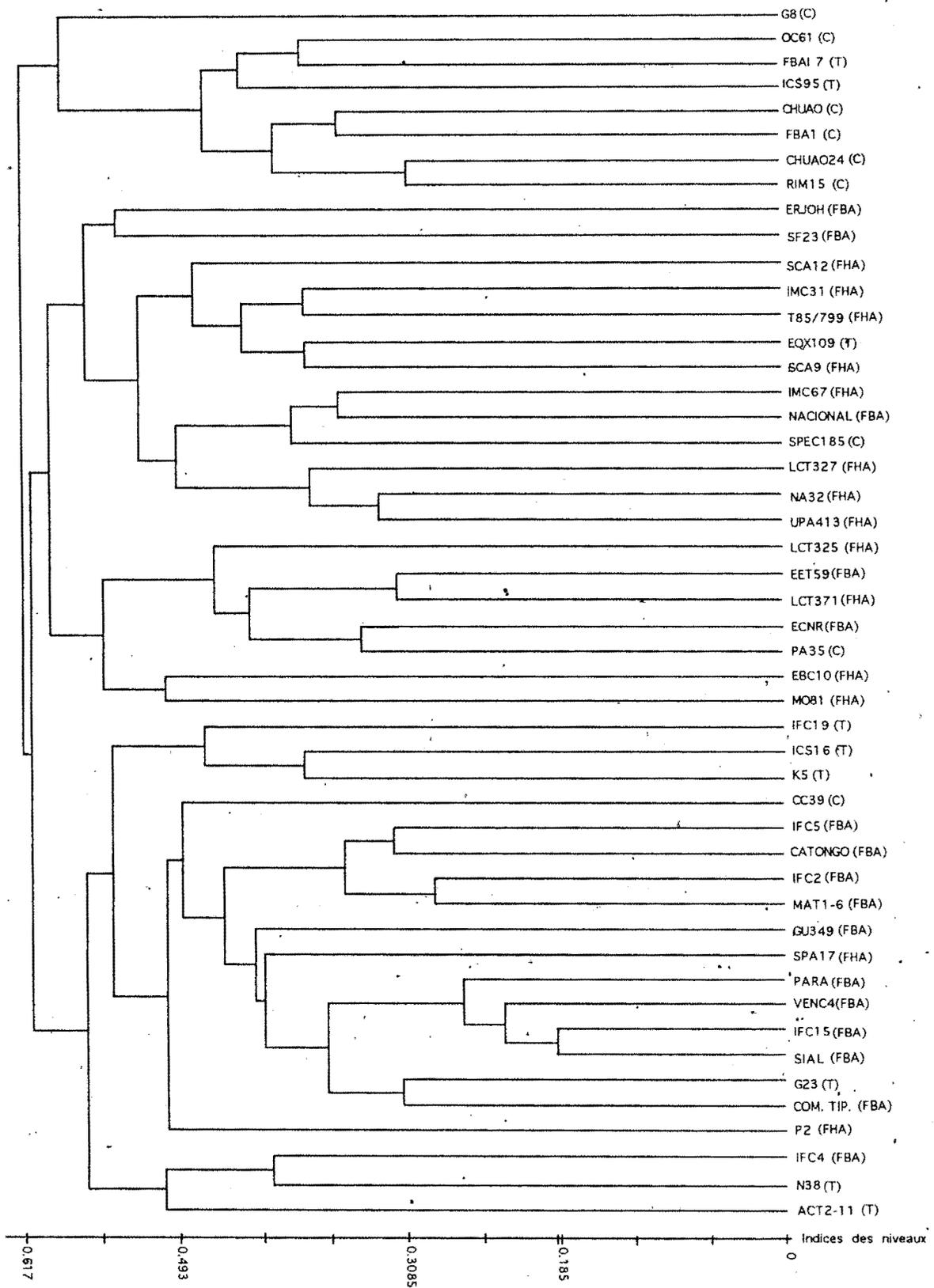


Figure7: Phénogramme obtenu au moyen des indices de ROGERS ET TANIMOTO sur les données RAPD de 48 individus

T=Trinitario      FHA=Forastero Haut Amazoniens  
 C=Criollo        FBA=Forastero Bas Amazoniens

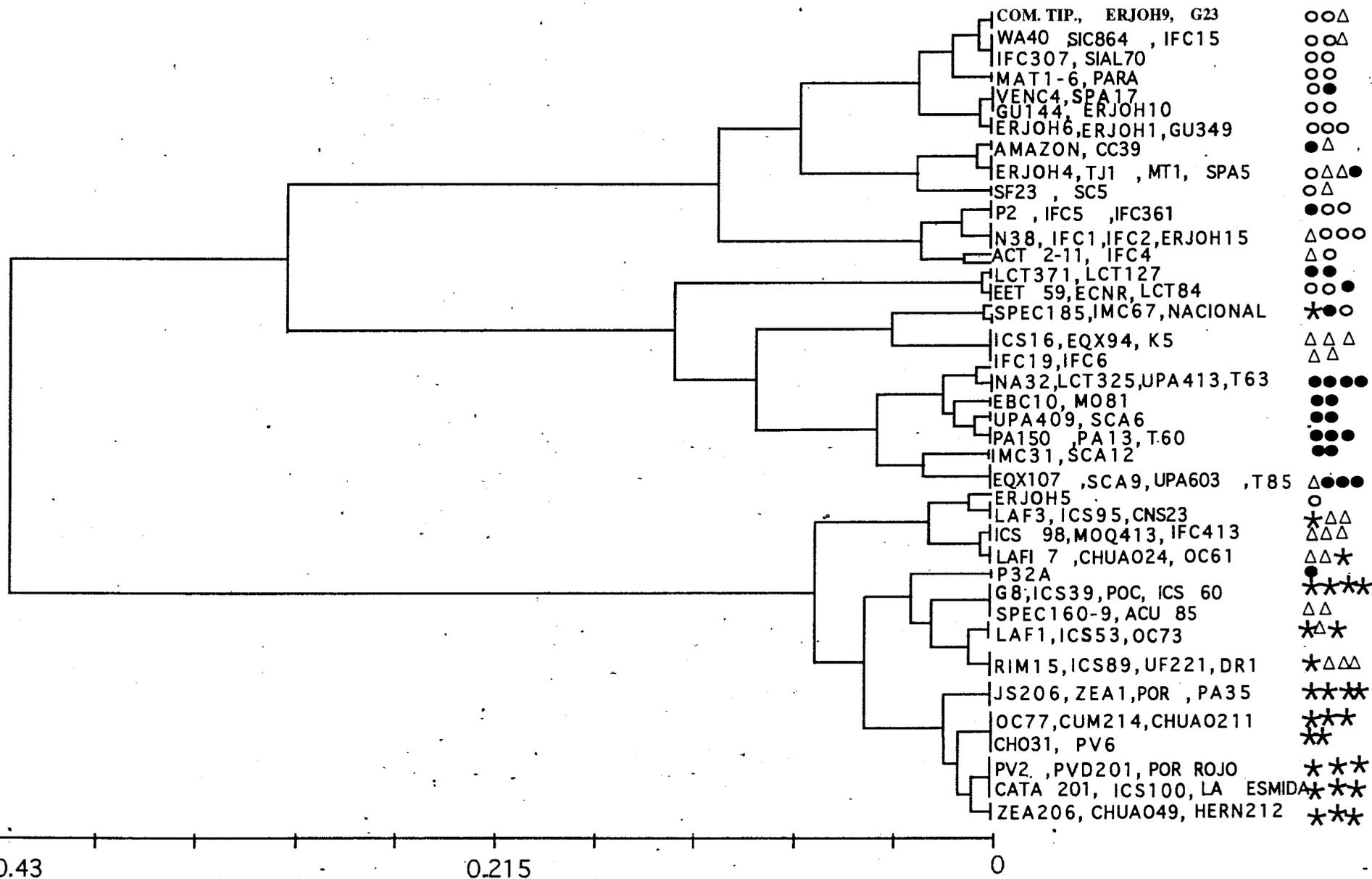


Figure 8: Phénogramme obtenu à partir des coordonnées des sept premiers axes de l'AFC portant sur l'ensemble des bandes RAPD.

★ Criollo ● Forastero Haut Amazoniens ○ Forastero Bas Amazoniens △ Trinitario

formé essentiellement de Forastero Haut Amazoniens et de quelques géotypes d'Équateur, de Guyane et du Brésil considérés comme Forastero Bas Amazoniens, un autre groupe réunit les Forastero Bas Amazoniens et quelques Trinitario qui leur sont très proches et enfin un troisième groupe est composé de Criollo et de Trinitario. Trois géotypes (N38, IFC4 et ACT2-11) se distinguent par leur regroupement à l'écart des autres.

La figure 7 présente le phénogramme obtenu à partir des indices de Rogers et Tanimoto. Les indices varient de 0,185 à 0,617, l'inertie totale est donc plus étendue que celle obtenue par les distances de Czekanowski. Très tôt, il y a scission entre les Forastero et les Criollo. Par la suite, ce sont les Forastero Haut et Bas Amazoniens qui se séparent, donnant deux lignées distinctes. Les Trinitario ne constituent pas véritablement un groupe au sein du phénogramme car ils sont associés soit aux Criollo soit aux Forastero. Au sein des Criollo les indices varient de 0,367 à 0,586 soit une amplitude de 0,219. A l'intérieur des Forastero Haut Amazoniens, les indices vont de 0,310 à 0,587 soit une étendue de 0,277 alors chez les Forastero Bas Amazoniens la variation est de 0,185 à 0,500 soit une étendue de 0,315.

Bien que les géotypes ACT2-11, IFC4 et N38 soient intégrés dans le groupe des Forastero Bas Amazoniens, ils forment une lignée distincte. La comparaison des deux phénogrammes montre une concordance dans les classifications. Quels que soient les indices utilisés, trois grands groupes sont obtenus. Les trois géotypes, N38, IFC4, et ACT2-11 semblent particuliers. Parmi les Forastero Haut Amazoniens, on retrouve des géotypes classés comme Forastero Bas Amazoniens (Nacional, EET59, SF23 etc.). La somme des indices de Czekanowski représenté par le phénogramme est inférieure à celle constitués des indices de Rogers et Tanimoto. Dans les deux situations, les Criollo semblent plus homogènes que les Forastero Haut et Bas Amazoniens, mais il faut noter que les Criollo dans cette analyse sont sous représentés.

#### *Classification à partir des coordonnées des individus sur l'AFC*

Les coordonnées des individus prises sur les sept premiers axes représentant 58,39 % de l'inertie totale, ont été utilisées pour cette analyse. Ainsi, un nombre important d'individus ont pu être analysés et classés et une bonne représentativité des groupes a pu être obtenue. Le dendrogramme obtenu (figure 8) sépare les types Criollo des types Forastero. La branche des Criollo comporte trois lignées constituées par les Criollo provenant essentiellement du Vénézuéla qui semblent très homogènes, les Criollo plus variables proches des Trinitario originaires d'Indonésie et de Costa Rica et enfin les Trinitario. La branche des Forastero Haut Amazoniens semble très variable et une grande part de cette

variabilité a été apportée par des génotypes LCT-EEN. En effet, on retrouve quelques génotypes de ce groupe parmi les Forastero Bas Amazoniens. Les Forastero Bas Amazoniens semblent aussi variables et intègrent les génotypes guyanais. Les clones d'Equateur classés comme Forastero Bas Amazoniens semblent plus proches des Forastero Haut Amazoniens.

#### *Comparaison des différents types de regroupements*

Il est à noter que quel que soit le type de classification utilisé, trois groupes sont obtenus: Criollo, Forastero Bas Amazoniens et Forastero Haut Amazoniens. Alors que le dendrogramme obtenu par les indices de Czekanowski distingue d'emblée les trois grandes populations, le phénogramme obtenu à l'aide des indices de Rogers et Tanimoto ainsi que celui basé sur les coordonnées de l'AFC mettent en évidence l'origine commune des Forastero Haut et Bas Amazoniens avant de les subdiviser en deux groupes. Du fait de leur caractère hybride les Trinitario sont rarement mis en évidence en tant que groupe quel que soit le type de classification mais sont associés soit aux Forastero soit aux Criollo. Contrairement aux classifications basées sur des indices de distances, la classification à partir des coordonnées des axes factoriels est la plus conforme à la classification classique. En outre, elle permet d'intégrer dans l'analyse un nombre important d'individus.

### **3- COMPARAISON DE LA DIVERSITÉ RÉVÉLÉE PAR LES CARACTÈRES MORPHOLOGIQUES ET CELLE RÉVÉLÉE PAR LES RAPD**

La structure globale de la diversité génétique révélée par les RAPD est différente de celle révélée par les marqueurs morphologiques. En effet, alors que les marqueurs RAPD discriminent les différents groupes de cacaoyers, les marqueurs morphologique ne discriminent que deux groupes: Trinitario et Criollo d'une part et Forastero d'autre part, car au niveau des caractères morphologiques les Criollo sont sous représentés. Lorsque les séquences répétées RAPD sont considérées, alors davantage de similarités sont trouvées. En fait les séquences répétées RAPD et les marqueurs morphologiques ne discriminent que les groupes dont les différences sont les plus marquées comme c'est le cas entre les Forastero et Criollo. Quant à la diversité révélée par les séquences uniques amplifiées, elle est différente de celle révélée par les marqueurs morphologiques car elle ne sépare pas réellement les différents groupes mais cela est peut être du au nombre de marqueurs dans cette étude.

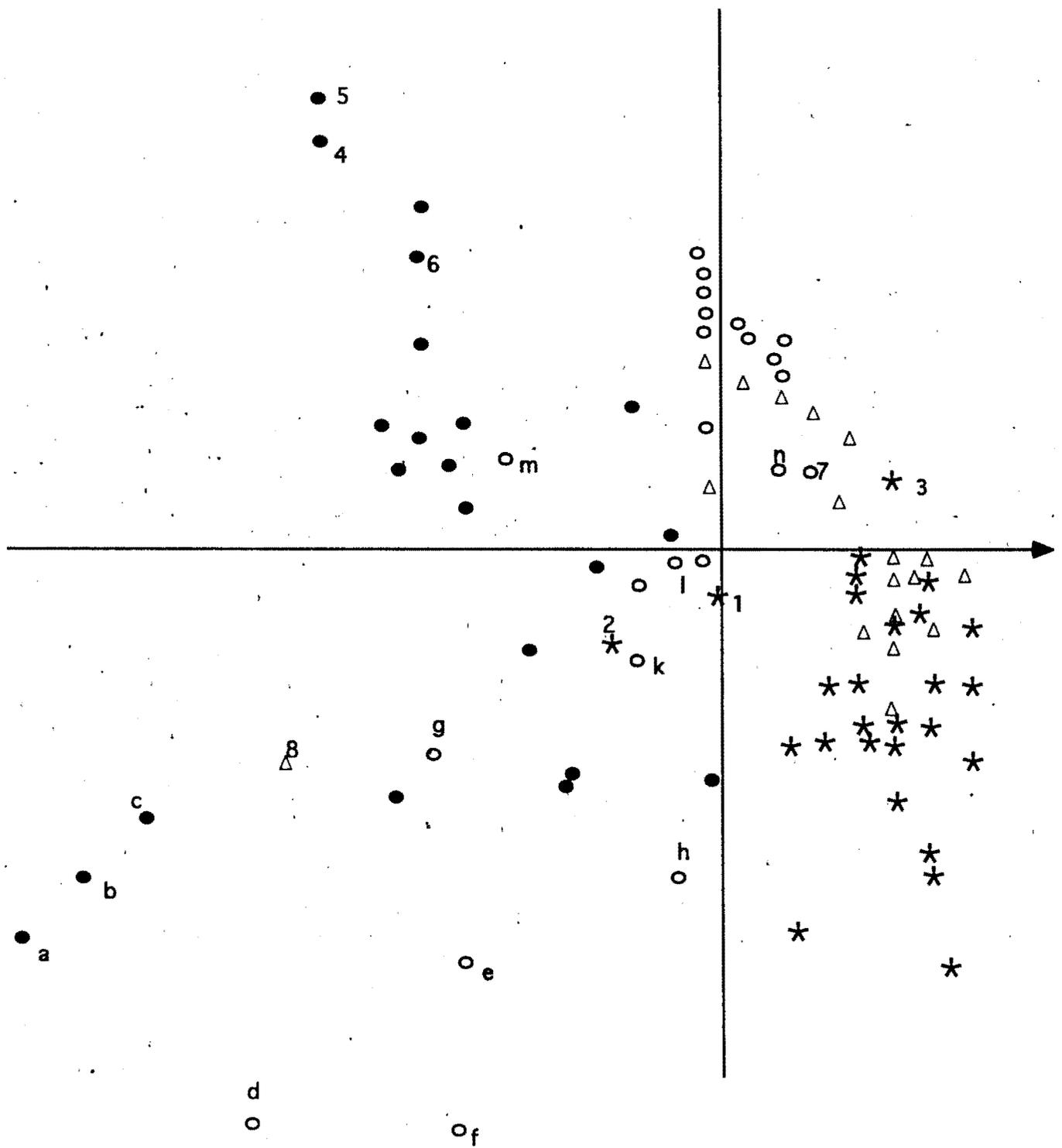


Figure 9 a: Plan 1 (31%) de l'AFC sur les données cDNA

- Forastero Haut Amazoniens
- Forastero Bas Amazoniens
- ★ Criollo
- △ Trinitario

- |            |          |           |             |
|------------|----------|-----------|-------------|
| 1- PV2     | 7- IFC4  | e- GU349  | k- ERJOH6   |
| 2- LAF3    | 8- MT1   | f- GU144  | l- EET59    |
| 3- SPEC185 | a- SCA6  | g- ERJOH5 | m- ERJOH1   |
| 4- P2      | b- SCA9  | h- ECNR   | n- nacional |
| 5- P32A    | c- SCA12 | i- SPA5   |             |
| 6- SPA17   | d- SF23  | j- ERJOH4 |             |

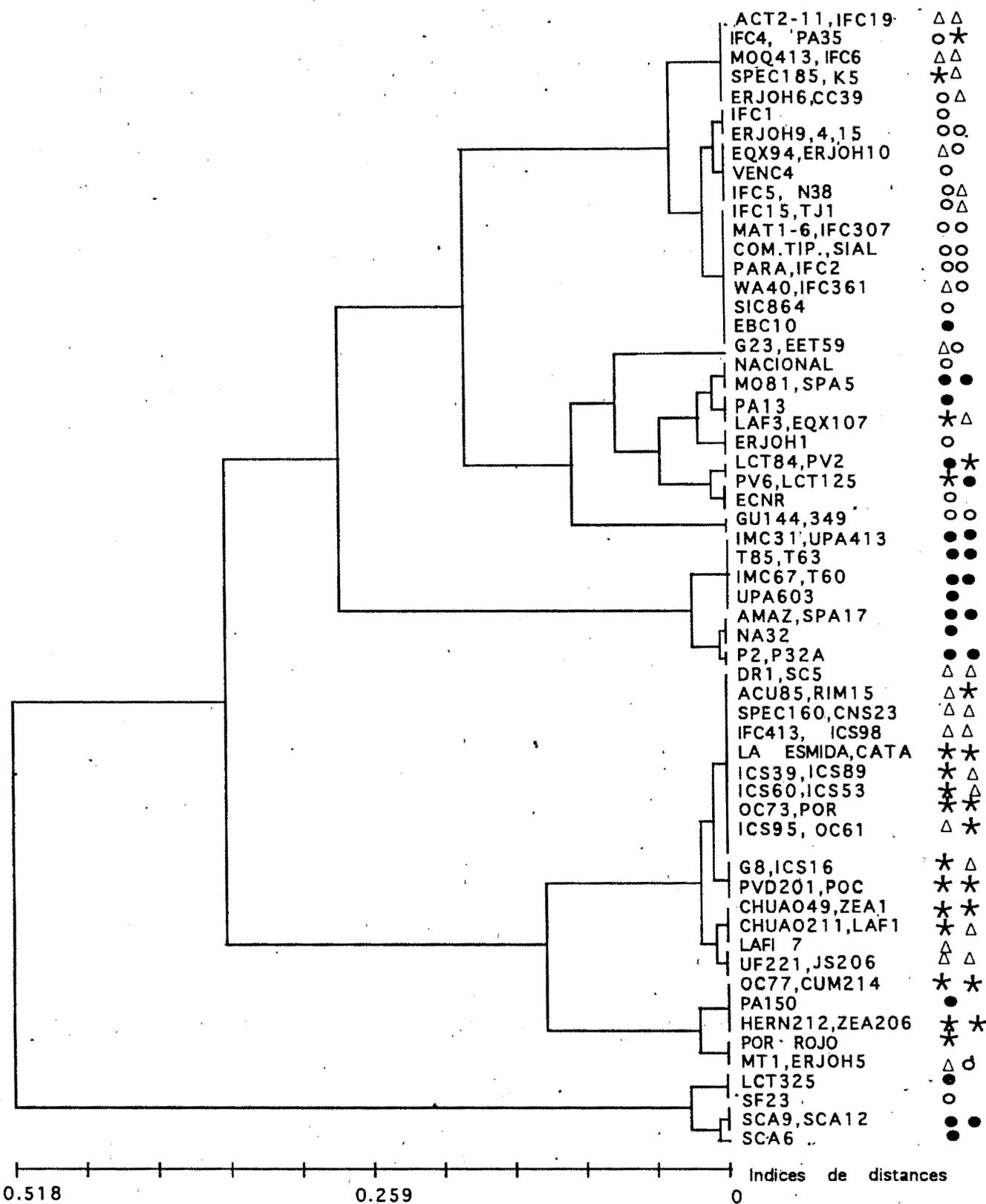


Figure 9b: Phénogramme obtenu à partir des coordonnées des sept premiers axes de l'AFC portant sur les données RFLP/cDNA

- ★ Criollo
- Forastero Haut Amazoniens
- Forastero Bas Amazoniens
- △ Trinitario

#### 4- COMPARAISON RFLP/RAPD

Les résultats de cette comparaison sont décrits dans l'article à paraître dans *Heredity*:

COMPARATIVE GENETIC DIVERSITY STUDIES OF *THEOBROMA CACAO* L. USING RFLP AND RAPD MARKERS par J. A. K. N'goran , V. Laurent, A. M. Risterucci and C. Lanaud.

Les principales conclusions auxquelles nous avons abouti sont:

- Les RAPD aussi bien que les RFLP /cDNA structurent mieux la diversité de l'espèce que les isozymes et les marqueurs morphologiques (figure 8 et figure 9).
- Ils ont permis d'identifier des groupes à partir des clones étudiés- Ils structurent les cacaoyers en Forastero Haut Amazoniens, Forastero Bas Amazoniens et Criollo. Ils révèlent de la variabilité dans les trois groupes.
- Ils représentent un outil de classification sur la base de ces groupes
- Les RFLP et les RAPD ont mis en évidence quelques cas probables de mauvaise classification : les Forastero Bas Amazoniens d'Equateur sont soit des hybrides , soit des Forastero Haut Amazoniens mais ne se classent pas parmi les Forastero Bas Amazoniens.
- Bien que certains Trinitario soient difficiles à séparer des Criollo, nous avons mis en évidence un ensemble de génotypes Criollo suffisamment distinct des Trinitario. Il proviennent en général du Vénézuéla dans notre analyse.
- Des génotypes particuliers ont été mis en évidence . Certains l'avaient aussi été par les RFLP.

## COMPARATIVE GENETIC DIVERSITY STUDIES OF *THEOBROMA CACAO* L. USING RFLP AND RAPD MARKERS.

J. A. K. N'goran \*, V. Laurent, A. M. Risterucci and C. Lanaud

Address: Laboratoire Biotrop /CIRAD (Centre de Coopération Internationale en Recherches Agronomiques pour le Développement) BP 5035, 34032 Montpellier cedex, France

Fax: 19 33 67615792

Lanaud C. : Corresponding author

Key words: *Theobroma cacao* L., genetic diversity, RAPD, RFLP.

### ABSTRACT

The genetic diversity of 106 genotypes of *Theobroma cacao* L. was assessed using RFLP and RAPD methods. Thirty one cDNA probes and 19 primers generated respectively 87 and 49 polymorphic bands that were submitted to factorial and cluster analyses. Both RFLP and RAPD bands separated the individuals into three distinctive groups. Some genotypes did not correspond to the currently accepted classification and are considered to have been originally misclassified. The nature of RAPD bands was investigated by Southern analysis. There were as many highly repeated sequences as single copy sequences but the two types of sequences did not structure the variability in the same way. Similar results were obtained with RFLPs and RAPDs with regard to the genetic structure of *Theobroma cacao* L. Therefore, RAPDs appear to be as efficient as RFLPs for classifying cocoa genotypes.

### INTRODUCTION

Based on morphological characteristics and geographic distribution, *Theobroma cacao* L. has been subdivided into three main groups namely Forastero, Criollo and Trinitario. Forastero clones have originated in the Amazonian region of South America and possess green pods with predominantly purple seeds. Most cultivated clones are Forastero which represent over 80 % of the world cocoa production. Forastero have been further subdivided into Upper Amazon and Lower Amazon due to their geographic locations and to some distinctive characteristics (e.g. precocity and disease resistance). Criollo were the first domesticated cocoas and have originated in Central America and in the north of South America. They possess large white or rosy beans that give a desirable chocolate although they have poor agronomic value. Trinitario are hybrids between Criollo and Forastero. The great variability

of pods and seeds makes it difficult to distinguish cocoa groups. Moreover, lack of information on a large number of clones limits their use in breeding programs.

Although morphological characters have allowed Engels (1986) to discriminate Criollo from Forastero within a pool of 294 genotypes, they are affected by environmental effects and are difficult to measure due to continuous variation. In 1987, Lanaud used six enzymatic systems representing 9 loci to assess the genetic diversity of 350 genotypes of cacao. She uncovered great variability among Forastero clones as was encountered by Pound (1945) during his collecting trips in the Upper Amazon region, but the small number of enzymatic markers available did not allow the fine discrimination of groups as described by Cheesman (1944). The availability of new, abundant and neutral genetic markers has allowed some progress in the distinction of groups. Laurent et al (1993a, 1993b, in press) used three types of RFLP probes to assess the genetic diversity of 203 genotypes. Ribosomal nuclear DNA probes distinguished Criollo, American Trinitario and Forastero. Mitochondrial probes revealed considerable variability among Criollo clones and cDNA probes confirmed the original structuring into Forastero and Criollo. However, the RFLP technique is labour intensive and the use of radio-elements makes it difficult to extend its use to developing countries. RAPDs on the other hand are easier to develop and handle (Williams et al, 1990) and have proved to be valuable for a number of purposes. They have been used for establishing genetic linkage maps (Carlson et al, 1991; Fauré et al, 1993), for fingerprinting (Fukuoka et al. 1992; Wilde et al. 1992) and for diversity studies (Kazan et al, 1993 ; Stiles et al. 1993). Although RAPDs are simple, their results are sometimes controversial due to lack of repeatability of some amplification products. In this study, we compare the structure of the genetic variability obtained by RAPD and RFLP methods, to test the general reliability of RAPD technique in cocoa in order to apply it for cocoa germplasm characterization as a possible alternative to RFLPs.

## **MATERIALS AND METHODS**

### **Plant Material**

A total of 101 genotypes of *Theobroma cacao* L. belonging to various groups and origins were collected from different research stations as dried leaves and used for RFLP studies (Table 1). One hundred and six accessions (Table 1) were used for RAPD analyses (101 common to both studies).

### **cDNA isolation**

cDNA probes were synthesized from fresh seed tissue using Pharmacia kits. They were ligated into pUC18 and used to transform strain DH5 $\alpha$  of *E. coli*. Plasmids were isolated by minipreparation and inserts were amplified by PCR and extracted from low melting gels.

### **Total DNA extraction and RFLP procedures**

Total DNA was extracted from dried leaves as described by Laurent et al. (1993a, 1993b). Five  $\mu$ g of total DNA were digested with four restriction enzymes (*EcoRI*, *XbaI*, *EcoRV*, *HindIII*) and transferred to a nylon membrane (hybond N<sup>+</sup>). Thirty one cDNA probes were labeled by random priming with <sup>32</sup>PdCTP and used to probe blots containing digested total DNA. Prehybridizations and hybridizations were performed overnight at 42°C in a solution containing 50% formamide 6xSSC, 5xDenhart, 0.5% SDS, 25  $\mu$ g/ml herring sperm DNA. Autoradiographs were exposed at -80°C for a week.

### **RAPD procedures**

Each amplification was performed in a reaction volume of 25  $\mu$ l containing 1x Promega reaction buffer, 0.2  $\mu$ M of primer, 1.5 mM MgCl<sub>2</sub>, 200  $\mu$ M of each of dNTPs, 5 ng of total DNA and 1.5 unit of Taq polymerase (Promega). Primers from Operon Technologies (Kits D, K, L, M, P, Q) previously screened in a mapping project were used. Amplifications were carried out in a Techne PHC2 thermocycler. After an initial denaturation step at 94°C for 5 min, 43 cycles of 1 min at 93°C, 50 sec at 36°C, 2 min at 72 °C were performed followed by a final incubation phase of 10 min at 72 °C. Amplification products were then analyzed by electrophoresis in 1.4 % agarose gels run at 100 V for 3 to 4 hours in a TBE buffer and stained with ethidium bromide. Repeatability tests and primer screening were conducted on a sample of 19 genotypes belonging to different origins. Nineteen primers generated 49 stable and readable bands that were scored for their presence or absence.

The nature of 39 polymorphic RAPD products was investigated by Southern analysis. Amplification product leftovers were allowed to migrate overnight in 1.2 % low-melting agarose gels. They were extracted from gels and hybridized to blots containing total DNA of five clones digested with five restriction enzymes (*BglIII*, *EcoRI*, *XbaI*, *EcoRV*, *HindIII*) as described in the RFLP procedures.

### **Statistical analyses**

RAPD or RFLP data were scored as presence or absence of bands in each genotype. A chi square test was used to compare the frequencies of RAPD bands in populations. Data were then subjected to factorial analyses of correspondences (FAC) (Benzecri, 1973). Cluster analyses were performed on the coordinates of

factorial axes and phenograms were built by the average method (UPGMA). After investigation of RAPD bands, single copy and highly repeated sequences were analyzed separately. All statistical analyses were performed using ADDAD software (Addad, 1985).

## RESULTS

### A) cDNA analysis

Thirty one cDNA probes produced 87 polymorphic fragments that were scored for their presence or absence in the genotypes.

The first plane of the factorial analysis (Fig 1) represents 31.2 % of the total variability. It distinguishes Upper Amazon Forastero, Criollo and Lower Amazon Forastero. Forastero clones exhibit a wider range of variability compared to Criollo clones. Trinitario clones cluster either with Criollo or with Lower Amazon Forastero. In the Forastero group, the Scavina genotypes (a, b, c) known as good parents in crosses lie outside the pool of Upper Amazon clones. Wild French Guyana clones (e, f), some Ecuadorian clones (h, l) and some genotypes from Brazil (g, k, m) are not included within the Lower Amazon Forastero group. Criollo genotypes, except some (1,2,3) are concentrated in the lower right quarter of the plane.

A phenogram based on coordinates of factorial analysis identified Upper Amazon Forastero, lower Amazon Forastero and Criollo (data not shown). Trinitario clustered either with Lower Amazon Forastero or Criollo clones and a small cluster comprises the Upper Amazon Scavina genotypes.

### B) RAPD analysis

#### Nature of RAPD products in cocoa

Southern analysis of RAPD products revealed three types of patterns: 12 amplification products used as probes gave one to three bands on blots and were therefore considered as single copy sequences, 12 other products generated a smear on blots and were assimilated to dispersed highly repeated sequences and a third type of product induced more than three discrete bands with or without background smears. This last type could not be unambiguously assigned. Some amplification products could not be extracted alone from gels whatever the duration of migration because they were too close to other bands.

#### Level of Polymorphism and comparison of RAPD products in populations

The 19 primers generated 1 to 5 polymorphic fragments. The size of bands ranges from 0.3 to 2 kb. Some bands occur more often in some populations. A chi square test was performed on data for uniformity of band presence in the

populations (Table 2). Significant differences in frequencies were revealed for some bands although no population specific band was obtained. A pairwise comparison of frequencies in populations revealed that Lower Amazon Forastero clones are close to Upper Amazon Forastero and to Trinitario. Trinitario are also related to Criollo as was expected from current classification.

#### Diversity organization

##### *a) Bulk polymorphic RAPD products analysis*

The first plane of the FAC using all polymorphic bands without any distinction of their nature represents 24.3 % of the total variability (Fig 2). The first axis discriminates Forastero into Upper Amazon and Lower Amazon, while the second axis distinguishes Forastero from Criollo. Trinitario clones spread between Lower Amazon Forastero and Criollo. Most American Trinitario, probably with a strong Criollo background, cluster with Criollo and African Trinitario are characterized by a median position between Criollo and Lower Amazon Forastero. Some Ecuadorian clones (h, n, l), Guyana populations (e) and some Brazilian clones (g, k), usually classified as Lower Amazon Forastero, cluster however with Upper Amazon Forastero. Some Forastero genotypes (5, 6, 7, m) are characterized by a median position between Criollo and Forastero. Upper Amazon Scavina clones (a, b, c) are well integrated within their group while the clone P2' (4) is included in the Lower Amazon Forastero group. Forastero and Trinitario show a wider range of variability than Criollo. Cluster analysis as well as FAC distinguishes Upper Amazon Forastero, Lower Amazon Forastero and Criollo but did not identify most Trinitario as a separate group. With the cluster analysis, Trinitario clones group together either with Criollo or with Lower Amazon Forastero (data not shown).

*b) Analysis of single copy and repeated sequences derived from RAPD products* The first plane of the FAC on the 12 single copy RAPD bands represents 33 % of the variability. Forastero and Criollo could not be identified as separate groups (data not shown). Repeated sequences, on the other hand, structured the diversity better. The first plane, which accounts for 38.6 % of the total variability, (Fig 3) distinguishes clearly Criollo and Forastero. No discrimination could however be made between Upper and Lower Amazon Forastero clones. Trinitario are mixed up either with Forastero or with Criollo.

## DISCUSSION AND CONCLUSION

RAPD fragments are DNA sequences amplified by a polymerase. They could cover the entire genome by revealing coding or non coding regions, repeated or single copy sequences. The number of repeated sequences is correlated

with the size of the genome (Vedel and Delseny, 1987; Lapitan, 1992), and Lanaud et al. (1992) and Figueira et al. (1992) showed that the cocoa tree has a small genome of 0.4 pg per haploid cell. One could, therefore, reasonably expect a relatively low percentage of repeated sequences in the cocoa genome. However, in the present study, more repeated sequences (high or low repeats) were found than single copy sequences. Repeated sequences may have been preferentially amplified by the RAPD technique. Repetitive and single copy sequences do not reflect the same evolutionary processes. According to Delseny et al. (1988), repetitive sequences could evolve more rapidly than single copy sequences by amplification and transposition. Amplified single copy probes were expected to structure the variability in the same way as cDNA sequences but, probably due to the number of sequences analyzed, different features were observed. Amplified single copy sequence probes in this experiment gave similar results to enzymatic markers (Lanaud, 1987), revealing a wide range of variability among Forastero group without a clear distinction between Forastero and Criollo. Amplified repeated sequence probes on the other hand gave the same results as morphological characters by separating Forastero from Criollo. Twelve repeated sequence probes were enough to discriminate the two major groups as did 39 morphological characters, cDNA and rDNA probes. Those repeated sequences seem to be well distributed over the genome and are different from rDNA which is a specific repeated sequence. In cocoa, the analyses of cDNA, rDNA, morphological characters and dispersed repeated RAPD sequences separated Criollo from Forastero and probably reflect ancient events that took place prior to the separation of Forastero into two subgroups. Forastero and Criollo genotypes might have first differentiated independently on each side of the Andean barrier as suggested by Cuatrecasas (1964). Later, mutational, transpositional and amplification events allowed a further genetic discrimination of Upper Amazon Forastero from Lower Amazon Forastero as shown by RFLP (Laurent et al. in press) and by RAPD analyses. Trinitario clones have a hybrid origin, and are difficult to separate from Criollo or Lower Amazon clones most of the time. Only RAPD analysis has differentiated some of them.

RAPDs in cocoa seem to be a valuable tool to classify genotypes and are as useful as other complex methods of investigation. They have made it possible to gain access to several events of the domestication of *Theobroma cacao* L., particularly recent events such as the creation of Trinitario. Figueira et al. (1993) and Lerceteau (1992) reported that they could not group genotypes according to the existing classification using RFLP and RAPD markers. The differences in the results may have arisen from the smaller size and the constitution of the population analysed.

Some genotypes are apparently misclassified by both types of molecular markers: ECNR(h), EET59(l) and SF23(d). This is not surprising for some Ecuadorian genotypes called "Nacional". Ecuador nacional cocoa comprises genotypes with plump whitish beans usually classified as Lower Amazon clones (Soria, 1970) or Criollo clones (Enriquez, 1992). From these analyses based on two types of molecular markers, strong evidence exists that those clones are more related to Upper Amazon or Trinitario clones.

The cocoa clones from Brazil are shown by RAPD and RFLP/cDNA to be heterogeneous. Some show little variability and cluster with african Amelonado clones whose ancestors they are supposed to be. Others like Erjoh genotypes collected in the western part of Brazil are more diverse. Erjoh clones, on the plane of the factorial analysis spread in between Upper Amazon clones, to which they are close geographically, and Lower Amazon Forastero. Some types behave like Trinitario and are located in between Forastero and Criollo.

The observed diversity within Brazilian wild germplasm is in accordance with the findings of Marcio et al. (1987) who collected plant material in the Rondonia area: those clones fit into a complex of diverse but genetically related clones.

Although a rather global agreement exists between RFLP and RAPD techniques for cocoa diversity studies, some particular results are obtained from each one. In the factorial analysis of cDNA data, Scavina genotypes are somewhat isolated from the other members of the Upper Amazon group whereas in the RAPD analysis they fit well in the group. P32A (5) and IFC4 (7) appear distinct on RAPD analysis, but are quite integrated in their usual respective group of classification in the cDNA analysis. These genotypes should be studied in more detail.

Based on these results, RAPD is an appropriate technique for studying the diversity of cocoa populations as RAPD analyses are fast and well suited for surveying large samples of genotypes like those held in international genebanks.

**Acknowledgments:** We thank all Cocoa Research Stations for providing us with plant material. We are grateful to Dr. Bartley and Dr. Lockwood for information on the origins of clones.

## RÉFÉRENCES

- Addad. 1985. Manuel de référence. Association pour le Développement et la Diffusion de l'Analyse des données. Paris.
- Benzecri, J. P. 1973. L'analyse des données. tome 2: L'analyse des correspondances. Dunod (eds)Paris.
- Carlson, J. E, L. K. Tulsieram, J. C. Glaubitz, V. W. K. Luk, C. Kauffeldt, and R. Rutledge. 1991. Segregation of random amplified DNA markers in F1 progeny of conifers. *Theor. Appl. Genet.* 83: 194-200.
- Cheesman, E.E. 1944. Notes on the nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture* 21:144-159.
- Cuatrecasas, J. 1964. Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Bulletin of the United States National Museum, Smithsonian Institution(Washington)* 35:379-614.
- Delseny, M., F. Grellet, D. Tremousaygue, M. Raynal and F. Panabières. 1988. Structure, évolution et expression de l'ADN nucléaire. *Bull. Soc. Bot. Fr.*135, *Actual. bot.*(2), 23-38.
- Engels, J.-M. M. 1986. The systematic description of cacao clones and its significance for taxonomy and plant breeding. Ph.D. thesis, Wageningen.
- Enriquez, G.A.. 1992. Characteristics of cacao "nacional" of Ecuador. *Internation. Workshop on Conservation, Characterisation and Utilisation of cocoa genetic resources in the 21st century,september13-17:269-278*. Trinidad.
- Fauré S., Noyer J.L., Horry J.P., Bakry F., Lanaud C., D. Gonzalez D. Leon.1993. A molecular marker-based linkage map of diploid bananas (*Musa acuminata*). *Theor Appl Genet.* 87 (4): 517-526.
- Figueira, A. J. Janick, P. Goldsbrough. 1992. Genome size and DNA polymorphism in *Theobroma cacao*. *J. of Amer. Soc. Hort. Sci.* 17:673-677.
- Figueira, A. J. Janick, M. Levy and P. Goldsbrough. 1993. Reexamining the classification of *Theobroma cacao* L. using molecular markers. *XII Intern. Cocoa Res. conference* , July 18-24 , Yamoussoukro, Cote d'Ivoire.
- Fukuoka, S. K. Hosak, O. Kamijima. 1992. Use of random amplified polymorphic DNAs (RAPDs) for identification of rice accessions. *Jpn. J. Genet* 67: 243-252.
- Kazan, K., J. M. Manners, and D. F. Cameron. 1993. Genetic variation in agronomically important species of *Stylosantes* determined using random amplified polymorphic DNA markers. *Theor. Appl. Genet.* 85: 882-888.

- Lanaud, C. 1987. Nouvelles données sur la biologie du cacaoyer (*Theobroma cacao* L.). Diversité des populations, système d'incompatibilité, haploïdes spontanés. Leurs conséquences pour l'amélioration génétique de cette espèce. Doctorat d'état, Paris, XI.
- Lanaud, C., P. Hamon et C. Duperray. 1992. Estimation of the nuclear DNA content of *Theobroma cacao* L. by flow cytometry. *Café, Cacao, Thé* 36: 3-8.
- Lapitan, N. L. V. 1992. Organization and evolution of higher plant nuclear genomes. *Genome* 35: 171-181.
- Laurent V., Risterucci A.M., C. Lanaud. 1993a. Chloroplast and mitochondrial DNA diversity in *Theobroma cacao*. *Theor Appl Genet.* 87 (1-2): 81-88.
- Laurent V., Risterucci A.M., C. Lanaud. 1993b. Variability for nuclear ribosomal genes within *Theobroma cacao*. *Heredity* 71 (1): 96-103.
- Laurent, V., Risterucci A.M. and C. Lanaud. Genetic diversity in cocoa revealed by cDNA probes. *Theor. Appl. Genet.* in press
- Lerceteau, E. 1992. Evaluation de la diversité génétique de *Theobroma cacao* par les techniques de détection d'ADN polymorphe amplifié au hasard (RAPD) et de polymorphisme de longueur des fragments de restriction de l'ADN (RFLP). Rapport de stage du DEA de Biologie moléculaire et Cellulaire végétale. Université Paris XI.
- Marcio, C., V. C. De Almeida, C. F. Gomes de Almeida. 1987. Coleta de cacau silvestre no estado de Rondonia, Brasil. *Revista Theobromae* 17(2): 65-92.
- Pound, F.J. 1945. A note on the cocoa populations of South America. *Rep. proc. Cocoa Conf.* 131-133.
- Soria, J.V.. 1970. Principal varieties of cocoa cultivated in tropical America. *Cocoa Growers' Bulletin* 19:12-21.
- Stiles, J. I., C. Lemme, S. Sondur, M. B. Morshidi, R. Manshardt. 1993. Using random amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars. *Theor. Appl. Genet.* 85: 697-701.
- Vedél, F. and M. Delseny. 1987. Repetitvity and variability of higher plant genomes. *Plant Physiol. Biochem.* 25 (2),191-210.
- Wilde, J. R. Waugh, W. Powell. 1992. Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. *Theor. Appl. Genet.* 83:871-877.
- Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nuc. Ac. Res.* 18: 6531-6535.

LOCUS	Fréquences alléliques	
	E1(N=180)	E2(N=102)
cTcCIR9	n=179	n=99
allèle 1	0.34	0.34
allèle 2	0.66	0.66
Diversité	0.45	0.45
cTcCIR25	n=180	n=100
allèle 1	0.04	0.05
allèle 2	0.96	0.95
Diversité	0.08	0.10
cTcCIR22	n=179	n=100
allèle 1	0.06	0.06
allèle 2	0.94	0.94
Diversité	0.11	0.11
cTcCCIR12	n=180	n=99
allèle 1	0.56	0.56
allèle 2	0.44	0.44
Diversité	0.49	0.49
cTcCIR2	n=176	n=96
allèle 1	0.50	0.47
allèle 2	0.50	0.53
Diversité	0.50	0.50
cTcCIR3	n=171	n=94
allèle 1	0.19	0.20
allèle 2	0.67	0.53
allèle 3	0.14	0.27
Diversité	0.53	0.61
cTcCIR29	n=172	n=97
allèle 1	0.88	0.87
allèle 2	0.12	0.13
Diversité	0.21	0.23
cTcCIR21	n=170	n=88
allèle 1	0.48	0.45
allèle 2	0.52	0.55
Diversité	0.50	0.50

LOCUS	Fréquences alléliques	
	E1(N=180)	E2(N=102)
cTcCIR33	n=162	n=86
allèle 1	0.85	0.87
allèle 2	0.15	0.13
Diversité	0.26	0.23
cTcCIR42	n=164	n=85
allèle 1	0.81	0.80
allèle 2	0.19	0.20
Diversité	0.31	0.32
cTcCIR26	n=164	n=88
allèle 1	0.72	0.71
allèle 2	0.28	0.29
Diversité	0.40	0.41
cTcCIR68	n=155	n=88
allèle 1	0.53	0.50
allèle 2	0.43	0.45
allèle 3	0.04	0.05
Diversité	0.53	0.55
cTcCIR39	n=158	n=85
allèle 1	0.70	0.69
allèle 2	0.28	0.28
allèle 3	0.02	0.03
Diversité	0.43	0.44
cTcCIR72	n=155	n=92
allèle 1	0.73	0.69
allèle 2	0.27	0.31
Diversité	0.39	0.43
cTcCIR28	n=160	n=87
allèle 1	0.95	0.94
allèle 2	0.05	0.06
Diversité	0.10	0.11
cTcCIR80	n=143	n=80
allèle 1	0.13	0.14
allèle 2	0.87	0.86
Diversité	0.23	0.24

Tableau 11: Fréquences alléliques et indices de diversité au sein des échantillons utilisés pour les analyses RFLP (E1) et pour les analyses RAPD (E2)  
n=nombre de génotypes

LOCUS	Fréquences alléliques	
	E1(N=180)	E2(N=102)
cTcCIR6	n=149	n=77
allèle 1	0.76	0.81
allèle 2	0.24	0.19
Diversité	0.36	0.31
cTcCIR30	n=170	n=90
allèle 1	0.55	0.50
allèle 2	0.42	0.44
allèle 3	0.03	0.06
Diversité	0.52	0.55
cTcCIR83	n=146	83.00
allèle 1	0.54	0.57
allèle 2	0.46	0.43
Diversité	0.50	0.49
cTcCIR71	n=146	n=83
allèle 1	0.39	0.33
allèle 2	0.51	0.60
allèle 3	0.02	0.02
allèle 4	0.08	0.05
Diversité	0.58	0.53
cTcCIR15	n=140	n=79
allèle 1	0.12	0.16
allèle 2	0.88	0.84
Diversité	0.21	0.27
cTcCIR31	n=132	n=70
allèle 1	0.45	0.52
allèle 2	0.50	0.39
allèle 3	0.05	0.09
Diversité	0.57	0.57

LOCUS	Fréquences alléliques	
	E1(N=180)	E2(N=102)
cTcCIR5	n=137	n=71
allèle 1	0.92	0.96
allèle 2	0.08	0.04
Diversité	0.15	0.08
cTcCIR35	n=155	n=87
allèle 1	0.05	0.06
allèle 2	0.95	0.94
Diversité	0.10	0.11
cTcCIR17	n=133	n=70
allèle 1	0.91	0.90
allèle 2	0.09	0.10
Diversité	0.16	0.18
cTcCIR43	n=124	n=61
allèle 1	0.33	0.35
allèle 2	0.25	0.20
allèle 3	0.42	0.45
Diversité	0.76	0.64
cTcCIR32	n=176	n=93
allèle 1	0.31	0.32
allèle 2	0.67	0.65
allèle 3	0.00	0.01
allèle 4	0.02	0.02
Diversité	0.47	0.47

Diversité globale      0,37      0,37

Tableau 11 (fin)

## 5- PARAMÈTRES GÉNÉTIQUES DE LA DIVERSITÉ RÉVÉLÉE PAR LES RFLP

### A- DIVERSITÉ GLOBALE

Pour les 27 locus cDNA analysés, le tableau 11 montre les fréquences alléliques au sein d'un échantillon de 181 individus analysés par les RFLP et au sein de celui de 101 individus utilisés pour les analyses RAPD. Le nombre d'individus par locus varie de  $n=124$  à 180 d'un côté et de 61 à 100 à cause des données manquantes. La comparaison des fréquences alléliques au sein des deux groupes d'échantillons montre peu de différences, indiquant que le sous échantillon de 101 individus est représentatif de l'espèce. Si on s'accorde sur la définition d'un locus polymorphe en disant qu'il s'agit d'un locus avec au moins deux allèles et dont l'allèle le plus rare a une fréquence supérieure ou égale à 0,05, alors le pourcentage de locus polymorphe est de 96,30. Ce pourcentage ne tient pas compte des locus monomorphes sur les cinq génotypes qui ont servi au tri des couples enzyme-sondes. Le nombre moyen d'allèles par locus est de 2,30 (S.E.=0,12) et le taux d'hétérozygotie moyen ou diversité de Nei est de 0,37 (S.E.=0,03; 0,157 avec isozyme). Le test global d'équilibre de Hardy Weinberg ainsi que le test spécifique par locus montrent une déviation significative. Seuls 4 locus sur 27 sont en équilibre au sein des cacaoyers analysés. Ce déséquilibre peut être attribué à plusieurs causes dont les types de croisements en cours au sein de la population et la subdivision en sous populations, mais aussi la sélection ou la migration.

### B- DIVERSITÉ SELON LES ORIGINES GÉOGRAPHIQUES

Le tableau 12 donne les valeurs par locus des fréquences alléliques, de la diversité de Nei et du taux d'hétérozygotie observée dans chaque groupe génétique. Au sein des Criollo, deux allèles sont fixés: l'allèle 1 de cTcCIR33 et allèle 2 de cTcCIR22 alors qu'aucun ne l'est au sein des Forastero. Certains allèles ont des fréquences faibles dans les deux populations. Dix neuf % des allèles au sein des Forastero et 22 % au sein des Criollo ont une fréquence inférieure ou égale à 0,1. D'autres allèles sont présents dans une population et pas dans l'autre. Cependant cette situation n'est pas fréquente et les allèles spécifiques des deux groupes sont peu nombreux : 4 allèles peuvent être considérés comme spécifiques aux Criollo: Ce sont l'allèle 1 de cTcCIR9, l'allèle 2 de cTcCIR42, l'allèle 2 de cTcCIR5 et l'allèle 3 de cTcCIR68 et trois sont spécifiques aux Forastero: l'allèle 1 de cTcCIR22, l'allèle 1 de cTcCIR3 et l'allèle 2 de cTcCIR33. Le nombre moyen d'allèles par locus et par population, le pourcentage de locus polymorphes, la

Locus	Forastero	Criollo	Locus	Forastero	Criollo
cTcCIR9	N=71	N=40	cTcCIR30	N=66	N=40
allèle 1	0.06	0.57	allèle 1	0.61	0.41
allèle 2	0.94	0.43	allèle 2	0.33	0.59
H	0.11	0.50	allèle 3	0.06	0.00
Ho	0.00	0.00	H	0.52	0.49
cTcCIR25	N=72	N=40	Ho	0.32	0.58
allèle 1	0.06	0.03	cTcCIR33	N=57	N=40
allèle 2	0.94	0.97	allèle 1	0.75	1.00
H	0.11	0.05	allèle 2	0.25	0.00
Ho	0.08	0.05	H	0.37	0.00
cTcCIR22	N=71	N=40	Ho	0.18	0.00
allèle 1	0.13	0.00			CR
allèle 2	0.87	1.00	cTcCIR42	N=65	N=35
H	0.22	0.00	allèle 1	0.90	0.67
Ho	0.09	0.00	allèle 2	0.09	0.33
cTcCIR12	N=72	N=40	allèle 3	0.01	0.00
allèle 1	0.62	0.50	H	0.18	0.45
allèle 2	0.38	0.50	Ho	0.08	0.20
H	0.48	0.51	cTcCIR26	N=65	N=36
Ho	0.26	0.45	allèle 1	0.89	0.47
cTcCIR2	N=71	N=38	allèle 2	0.11	0.53
allèle 1	0.63	0.38	H	0.19	0.51
allèle 2	0.37	0.62	Ho	0.12	0.56
H	0.47	0.48	cTcCIR68	N=57	N=36
Ho	0.34	0.55	allèle 1	0.79	0.22
cTcCIR3	N=70	N=35	allèle 2	0.21	0.67
allèle 1	0.34	0.06	allèle 3	0.00	0.11
allèle 2	0.54	0.57	H	0.33	0.50
allèle 3	0.12	0.37	Ho	0.32	0.19
H	0.58	0.54	cTcCIR39	N=64	N=31
Ho	0.29	0.17	allèle 1	0.47	0.84
cTcCIR29	N=68	N=39	allèle 2	0.48	0.16
allèle 1	0.79	0.94	allèle 3	0.05	0.00
allèle 2	0.21	0.06	H	0.56	0.28
H	0.34	0.12	Ho	0.16	0.19
Ho	0.25	0.13	cTcCIR72	N=63	N=37
cTcCIR21	N=70	N=35	allèle 1	0.78	0.70
allèle 1	0.48	0.66	allèle 2	0.22	0.30
allèle 2	0.22	0.34	H	0.35	0.42
H	0.50	0.46	Ho	0.22	0.32
Ho	0.30	0.34			

Tableau 12: Fréquences alléliques et indices de diversité par locus au sein des populations Forastero et Criollo.

H= Diversité de Nei Ho= Hétérozygotie observée

N= nombre de génotypes

E.S. erreur standard

Locus	Forastero	Criollo	Locus	Forastero	Criollo
cTcCIR28	N=63	N=36	cTcCIR31	N=52	N=34
allèle 1	0.96	0.90	allèle 1	0.39	0.35
allèle 2	0.04	0.10	allèle 2	0.51	0.60
H	0.08	0.18	allèle 3	0.10	0.05
Ho	0.05	0.14	H	0.58	0.52
cTcCIR80	N=61	N=26	Ho	0.14	0.41
allèle 1	0.20	0.10	cTcCIR5	N=42	N=37
allèle 2	0.80	0.90	allèle 1	0.87	0.95
H	0.32	0.18	allèle 2	0.13	0.05
Ho	0.16	0.12	H	0.23	0.10
cTcCIR35	N=59	N=36	Ho	0.26	0.11
allèle 1	0.07	0.04	cTcCIR17	N=45	N=33
allèle 2	0.93	0.96	allèle 1	0.99	0.77
H	0.13	0.08	allèle 2	0.01	0.23
Ho	0.07	0.03	H	0.02	0.36
cTcCIR6	N=54	N=36	Ho	0.02	0.15
allèle 1	0.44	0.94	cTcCIR43	N=45	N=29
allèle 2	0.56	0.06	allèle 1	0.03	0.69
H	0.50	0.11	allèle 2	0.38	0.17
Ho	0.26	0.06	allèle 3	0.59	0.14
cTcCIR83	N=59	N=36	H	0.52	0.48
allèle 1	0.70	0.38	Ho	0.07	0.48
allèle 2	0.30	0.62	cTcCIR32	N=71	N=39
Diversité	0.42	0.48	allèle 1	0.15	0.55
cTcCIR71	N=54	N=34	allèle 2	0.84	0.45
allèle 1	0.37	0.53	allèle 3	0.01	0.00
allèle 2	0.44	0.41	H	0.27	0.50
allèle 3	0.01	0.04	Ho	0.24	0.69
allèle 4	0.18	0.02			
H	0.64	0.56	H globale	0.35	0.33
Ho	0.26	0.38	E.S.	0.03	0.04
cTcCIR15	N=52	N=36	Ho globale	0.18	0.25
allèle 1	0.26	0.05	E.S.	0.02	0.04
allèle 2	0.74	0.94			
allèle 3	0.00	0.01			
H	0.39	0.11			
Ho	0.10	0.11			

Tableau 12(fin)

Locus	Fis-Forastero	Fis-Criollo	Fst
cTcCIR9	1	1	0.51
cTcCIR25	0.21	-0.01	0
cTcCIR22	0.62	-	0.08
cTcCIR12	0.45	0.11	0.01
cTcCIR2	0.28	-0.16	0.11
cTcCIR3	0.51	0.68	0.1
cTcCIR29	0.26	-0.05	0.7
cTcCIR21	0.4	0.25	0.5
cTcCIR30	0.39	-0.17	0.09
cTcCIR33	0.53	-	0.2
cTcCIR42	0.58	0.56	0.15
cTcCIR26	0.37	-0.1	0.36
cTcCIR68	0.06	0.61	0.4
cTcCIR39	0.72	0.3	0.19
cTcCIR72	0.36	0.24	0
cTcCIR28	0.38	0.22	0.01
cTcCIR80	0.49	0.35	0.01
cTcCIR35	0.47	0.66	-0.01
cTcCIR6	0.48	0.48	0.41
cTcCIR83	0.4	0.036	0.18
cTcCIR71	0.6	0.31	0.02
cTcCIR15	0.75	-0.03	0.12
cTcCIR31	0.77	0.21	0
cTcCIR5	-0.14	-0.04	0.02
cTcCIR17	0	0.58	0.2
cTcCIR43	0.87	0	0.39
cTcCIR32	0.1	-0.39	0.31
moyenne	0.48	0.25	0.18
écart type	0.05	0.08	0.04
I.C. à 95%	0,39-0,57	0,09-0,41	0,12-0,25

**Tableau 13 Valeurs des paramètres F au sein des Forastero et des Criollo**

I.C= Intervalle de confiance

diversité de Nei ainsi que les Fis sont donnés dans le tableau 13. Les deux groupes génétiques ne sont pas très différents au niveau de la richesse allélique car un seul allèle est absent au sein des Forastero et cinq au sein des Criollo par rapport à la population non subdivisée qui compte 64 allèles. De même, au niveau de la diversité de Nei, les deux aires géographiques représentées par les deux populations Forastero et Criollo sont comparables. Le test d'équilibre de Hardy Weinberg indique un déséquilibre de liaison pour la majorité des locus. En effet, 18 locus sur 26 observés, présentent un déséquilibre au sein de la population Forastero et 10 locus sur 25 au sein des Criollo. Le test de différenciation par comparaison simultanée des populations deux à deux, indique une différence significative entre les deux origines génétiques. En effet 16 locus sur 27 présentent une différenciation de type géographique. Les trois types d'indice de fixation de Wright sont présentés dans le tableau 13 pour chaque locus et pour chaque population. Le tableau 14 récapitule les données par population.

	Forastero	Criollo
N moyen d'allèles par locus (E.S.)	2,33 (0,11)	2,19 (0,12)
*% locus polymorphes	92,59	85,19
Diversité, H (E.S.)	0,35(0,03)	0,33 (0,04)
Heterozygotie observée	0,18(0,02)	0,25(0,04)
Fis	0,48( s=0,04)	0,25(s=0,08)
I.C. Fis	0,39-0,57	0,09-0,41
$t_{eq}$	0,76	0,94

**Tableau 14:** Valeurs moyennes des paramètres génétiques des Forastero et des Criollo: H= Diversité de Nei, I.C.= Intervalle de confiance,  $t_{eq}$ = taux d'allogamie estimée à l'équilibre de H.W., E.S= erreur standard, s= écart type.

\* Polymorphisme qui ne tient pas compte des locus non polymorphes sur les génotypes ayant servi au tri des couples enzyme -sonde.

Les valeurs de Fis par locus des Forastero varie de -0,14 à 1 avec une moyenne de 0,48. Pour la population Criollo, Fis varie de -0,39 à 1 avec une moyenne de 0,25. Les valeurs moyennes fortes de Fis sont observées au niveau des deux populations indiquant une forte structuration intrapopulation. Les valeurs de Fst entre les deux populations varient de 0,00 à 0,51 avec une moyenne de 0,18. L'intervalle de confiance à 95% du Fst obtenu par rééchantillonnage (bootstrap) sur les locus varie de 0,11 à 0,25 et ne contient pas la valeur 0

Locus	HA	BA	CR	TR
cTcCIR9	N=37	N=34	N=40	N=59
allèle 1	0.05	0.06	0.58	0.51
allèle 2	0.95	0.94	0.42	0.49
Diversité	0.10	0.11	0.49	0.50
cTcCIR25	N=37	N=35	N=40	N=59
allèle 1	0.09	0.01	0.03	0.03
allèle 2	0.91	0.99	0.97	0.97
Diversité	0.16	0.03	0.06	0.05
cTcCIR22	N=36	N=35	N=40	N=59
allèle 1	0.14	0.11	0.00	0.01
allèle 2	0.86	0.89	1.00	0.99
Diversité	0.24	0.19	0.00	0.02
cTcCCIR12	N=37	N=35	N=40	N=59
allèle 1	0.41	0.84	0.50	0.59
allèle 2	0.59	0.16	0.50	0.41
Diversité	0.49	0.26	0.51	0.48
cTcCIR2	N=37	N=34	N=38	N=58
allèle 1	0.57	0.71	0.38	0.49
allèle 2	0.43	0.29	0.62	0.51
Diversité	0.49	0.42	0.48	0.50
cTcCIR3	N=38	N=32	N=35	N=57
allèle 1	0.55	0.09	0.06	0.03
allèle 2	0.26	0.88	0.57	0.94
allèle 3	0.19	0.03	0.37	0.03
Diversité	0.60	0.23	0.54	0.12
cTcCIR29	N=34	N=34	N=39	N=57
allèle 1	0.63	0.94	0.94	0.95
allèle 2	0.37	0.06	0.06	0.05
Diversité	0.47	0.11	0.12	0.10
cTcCIR21	N=36	N=34	N=35	N=57
allèle 1	0.62	0.32	0.66	0.37
allèle 2	0.38	0.68	0.34	0.63
Diversité	0.47	0.44	0.46	0.47
cTcCIR30	N=33	N=33	N=40	N=56
allèle 1	0.45	0.77	0.41	0.60
allèle 2	0.50	0.15	0.59	0.38
allèle 3	0.05	0.08	0.00	0.02
Diversité	0.55	0.38	0.49	0.50
cTcCIR33	N=29	N=28	N=40	N=58
allèle 1	0.55	0.96	1.00	0.95
allèle 2	0.45	0.04	0.00	0.05
Diversité	0.50	0.07	0.00	0.10

**Tableau 16:** Fréquences alléliques et diversités géniques au sein des Forastero Haut Amazoniens(HA), Bas Amazoniens(BA) des Criollo(CR) et des Trinitario (TR): N= nombre de génotypes

Locus	HA	BA	CR	TR
cTcCIR42	N=33	N=32	N=35	N=55
allèle 1	0.94	0.86	0.67	0.76
allèle 2	0.05	0.14	0.33	0.24
allèle 3	0.01	0.00	0.00	0.00
Diversité	0.12	0.25	0.45	0.36
cTcCIR26	N=35	N=30	N=36	N=55
allèle 1	0.84	0.95	0.47	0.65
allèle 2	0.16	0.05	0.53	0.35
Diversité	0.27	0.10	0.50	0.46
cTcCIR68	29.00	28.00	35.00	54.00
allèle 1	0.79	0.78	0.22	0.47
allèle 2	0.21	0.22	0.68	0.47
allèle 3	0.00	0.00	0.11	0.06
Diversité	0.33	0.34	0.48	0.56
cTcCIR39	N=33	N=31	N=31	N=55
allèle 1	0.17	0.79	0.84	0.90
allèle 2	0.77	0.16	0.16	0.09
allèle 3	0.06	0.05	0.00	0.01
Diversité	0.38	0.35	0.27	0.18
cTcCIR72	N=32	N=31	N=37	N=47
allèle 1	0.73	0.82	0.70	0.72
allèle 2	0.27	0.18	0.30	0.28
Diversité	0.39	0.30	0.42	0.42
cTcCIR28	N=33	N=30	N=36	N=53
allèle 1	0.94	0.98	0.90	0.95
allèle 2	0.06	0.02	0.10	0.05
Diversité	0.11	0.03	0.18	0.10
cTcCIR80	N=29	N=32	N=26	N=53
allèle 1	0.36	0.05	0.10	0.07
allèle 2	0.64	0.95	0.90	0.93
Diversité	0.47	0.09	0.18	0.13
cTcCIR35	N=32	N=32	N=36	N=52
allèle 1	0.13	0.00	0.04	0.03
allèle 2	0.87	1.00	0.96	0.97
Diversité	0.22	0.00	0.08	0.06
cTcCIR6	N=29	N=25	N=36	N=53
allèle 1	0.34	0.56	0.94	0.94
allèle 2	0.66	0.44	0.06	0.06
Diversité	0.46	0.50	0.11	0.11
cTcCIR83	N=27	N=32	N=36	N=45
allèle 1	0.57	0.81	0.38	0.51
allèle 2	0.43	0.19	0.62	0.49
Diversité	0.50	0.31	0.47	0.51

Tableau 16 (suite)

Locus	HA	BA	CR	TR
cTcCIR71	N=31	N=23	N=34	N=52
allèle 1	0.50	0.20	0.54	0.33
allèle 2	0.18	0.80	0.41	0.62
allèle 3	0.02	0.00	0.04	0.01
allèle 4	0.30	0.00	0.01	0.04
Diversité	0.63	0.32	0.54	0.51
cTcCIR15	N=22	N=30	N=36	N=44
allèle 1	0.45	0.12	0.04	0.02
allèle 2	0.55	0.88	0.95	0.98
allèle 3	0.00	0.00	0.01	0.00
Diversité	0.51	0.21	0.10	0.04
cTcCIR31	N=24	N=28	N=34	N=40
allèle 1	0.15	0.61	0.35	0.64
allèle 2	0.73	0.32	0.60	0.34
allèle 3	0.12	0.07	0.05	0.02
Diversité	0.43	0.53	0.52	0.48
cTcCIR5	N=15	N=27	N=37	N=53
allèle 1	0.73	0.94	0.95	0.98
allèle 2	0.27	0.06	0.05	0.02
Diversité	0.39	0.11	0.10	0.04
cTcCIR17	N=15	N=22	N=33	N=53
allèle 1	1.00	0.98	0.77	0.92
allèle 2	0.00	0.02	0.23	0.08
Diversité	0.00	0.04	0.35	0.14
cTcCIR43	N=26	N=19	N=29	N=54
allèle 1	0.02	0.05	0.69	0.42
allèle 2	0.50	0.21	0.17	0.18
allèle 3	0.48	0.74	0.14	0.40
Diversité	0.52	0.42	0.48	0.64
cTcCIR32	N=37	N=34	N=39	N=57
allèle 1	0.14	0.16	0.62	0.32
allèle 2	0.84	0.75	0.38	0.68
allèle 3	0.02	0.09	0.00	0.00
Diversité	0.27	0.40	0.47	0.44
Diversité moyenne	0.38	0.24	0.33	0.30
Erreur	0.03	0.03	0.04	0.04

Tableau 16 (fin)

indiquant une différenciation spatiale importante des deux populations de cacaoyers. Les distances génétiques de Nei et de Rogers observée entre les deux populations sont de 0,13 et 0,29.

### C-DIVERSITÉ SELON LES GROUPES MORPHOGÉOGRAPHIQUES.

Les caractéristiques de chaque population sont données dans le tableau 15 et les informations par locus sont données dans le tableau 16. Le nombre d'allèles observés par locus varie peu d'un groupe à l'autre. En effet, ces valeurs ne sont pas statistiquement différentes. Le groupe le plus diversifié est celui des Forastero Haut Amazoniens dont la diversité correspond à celle observée dans la population non subdivisée.

	HA	BA	CR	TR
N moyen d'allèles par locus	2,3 (0,12)	2,15 (0,09)	2,19 (0,12)	2,3 (0,10)
% locus polymorphes	96,30	77,78	85,19	77,78
Diversité, H (ES)	0,38 (0,03)	0,24 (0,03)	0,33 (0,04)	0,30 (0,04)
Hmin-Hmax	0,1-0,63	0,029-0,5	0-0,55	0,09-0,64
Hétérozygotie observée	0,23(0,03)	0,126(0,02)	0,25(0,04)	0,3(0,05)
Min-max	0-0,53	0-0,32	0-0,69	0-0,67
Fis	0,38	0,46	0,25	-0,01
I.C. de Fis	0,26-0,51	0,33-0,58	0,09-0,41	-0,16-0,16
$t_{eq}$	0,85	0,79	0,94	1

**Tableau 15:** Valeurs moyennes des paramètres génétiques des Forastero Haut Amazoniens(HA), des Forastero Bas Amazoniens(BA), des Trinitario(TR) et des Criollo: H=Diversité de Nei, I.C.= Intervalle de confiance,  $t_{eq}$ = taux d'allogamie estimée à l'équilibre de H.W., E.S= erreur standard, s= écart type.\* Polymorphisme qui ne tient pas compte des locus non polymorphes sur les génotypes ayant servi au tri des couples enzyme -sonde.

Les Forastero Bas Amazoniens ont la plus faible diversité. Celle des Trinitario est comparable à celle des Criollo. Très peu de locus peuvent être dits spécifiques d'un groupe donné. L'allèle 1 du locus cTcCIR9 a une fréquence assez importante chez les Criollo et les Trinitario alors qu'il est presque inexistant chez les Haut et Forastero Bas Amazoniens. L'allèle 1 du locus cTcCIR3, l'allèle 2 du locus cTcCIR33 et l'allèle 4 du locus cTcCIR71 se rencontrent plus fréquemment au sein des Forastero Haut Amazoniens: ils ont de faibles fréquences dans les

		HA	BA	CR
BA	Comparaisons *	15	-	-
	Fst	0,20(s=0,04)	-	-
	I. C. à 95% de Fst	0,13-0,26	-	-
	Distance de Nei	0,13	-	-
	Distance MRD	0,29	-	-
CR	Comparaisons *	22	19	-
	Fst	0,23(s=0,04)	0,24(s=0,03)	-
	I. C. à 95% de Fst	0,15-0,31	0,17-0,29	-
	Distance de Nei	0,19	0,14	-
	Distance MRD	0,34	0,31	-
TR	Comparaisons *	20	15	11
	Fst	0,23(s=0,05)	0,11(s=0,03)	0,06(s=0,02)
	I. C. à 95% de Fst	0,15-0,32	0,061-0,16	0,03-0,09
	Distance de Nei	0,17	0,052	0,035
	Distance MRD	0,33	0,19	0,15

**Tableau 17:** Récapitulatif des comparaisons des quatre populations prises deux à deux (I.C. Intervalle de confiance, s= écart type, \*=significatif à 5 %.)

Locus	Fis-HA	Fis-BA	Fis-CR	Fis-TR
cTcCIR9	1	0.94	1	0.95
cTcCIR25	0.22	0	-0.01	-0.03
cTcCIR22	0.54	0.68	-	-0.02
cTcCIR12	0.45	0.03	0.11	-0.19
cTcCIR2	0.02	0.56	-0.16	-0.35
cTcCIR3	0.3	0.44	0.68	0.2
cTcCIR29	0.19	-0.05	-0.05	0.26
cTcCIR21	0.36	0.32	0.25	-0.19
cTcCIR30	0.23	0.43	-0.17	-0.4
cTcCIR33	0.32	0.89	-	-0.06
cTcCIR42	0.48	0.6	0.56	0.14
cTcCIR26	0.47	-0.04	-0.1	-0.33
cTcCIR68	-0.03	0.15	0.61	0.16
cTcCIR39	0.44	0.7	0.3	0.44
cTcCIR72	0.45	0.2	0.24	-0.05
cTcCIR28	0.48	-0.007	0.22	-0.06
cTcCIR80	0.49	-0.03	0.35	-0.07
cTcCIR35	0.44	0	0.66	-0.04
cTcCIR6	0.1	0.82	0.48	0.23
cTcCIR83	0.33	0.38	0.036	0.25
cTcCIR71	0.34	0.84	0.31	0.03
cTcCIR15	1	0.22	-0.03	-0.02
cTcCIR31	0.72	0.72	0.21	-0.41
cTcCIR5	-0.33	-0.04	-0.04	-0.03
cTcCIR17	-	0	0.58	0.62
cTcCIR43	0.93	0.75	0	-0.02
cTcCIR32	0.28	-0.18	-0.39	-0.4
moyenne	0.38	0.46	0.25	-0.01
écart type	0.06	0.07	0.08	0.09
I.C.	0,26-0,5	0,33-0,58	0,09-0,41	0,16-0,16

**Tableau 18:** Valeurs des Fis par locus au sein des quatre populations  
HA= Forastero Haut Amazonien BA= Forastero Bas Amazonien  
TR= Trinitario CR= Criollo I.C= Intervalle de confiance

Locus	Fst-HA-BA	Fst-HA-CR	Fst-HA-TR	Fst-BA-CR	Fst-BA-TR	Fst-CR-TR
cTcCIR9	-0.03	0.45	0.37	0.44	0.34	-0.01
cTcCIR25	0.044	0.03	0.02	-0.01	0	-0.01
cTcCIR22	-0.02	0.13	0.1	0.1	0.01	0
cTcCIR12	0.32	0	0.05	0.22	0.12	0
cTcCIR2	0.02	0.05	0.005	0.17	0.08	0.02
cTcCIR3	0.4	0.23	0.53	0.19	0.01	0.3
cTcCIR29	0.24	0.23	0.27	-0.013	-0.01	-0.01
cTcCIR21	0.15	-0.016	0.11	0.18	0	0.06
cTcCIR30	0.18	-0.003	0.02	0.26	0.08	-0.03
cTcCIR33	0.36	0.47	0.37	0.11	-0.013	0
cTcCIR42	0.02	0.19	0.1	0.07	0.13	0.05
cTcCIR26	0.04	0.25	0.09	0.41	0.21	0.07
cTcCIR68	-0.02	0.38	0.14	0.37	0.13	0
cTcCIR39	0.5	0.55	0.64	-0.02	0.02	-0.01
cTcCIR72	0	-0.017	-0.01	0.02	0.01	0
cTcCIR28	0	-0.01	-0.013	0.04	0	-
cTcCIR80	0.25	0.16	0.23	-0.002	0	0
cTcCIR35	0.09	0.02	0.04	0.01	0	-0.01
cTcCIR6	0.06	0.57	0.58	0.34	0.37	-0.01
cTcCIR83	0.1	0.06	-0.01	0.31	0.16	0.02
cTcCIR71	0.35	0.09	0.21	0.21	0.03	0.06
cTcCIR15	0.23	0.37	0.46	0.14	0.06	0
cTcCIR31	0.26	0.04	0.29	0.1	-0.013	0.12
cTcCIR5	0.16	0.18	0.29	-0.16	0	0
cTcCIR17	0	0.16	0.03	0.13	0	0.07
cTcCIR43	0.1	0.39	0.17	0.44	0.16	0.09
cTcCIR32	-0.1	0.3	0.1	0.27	0.08	0.07
moyenne	0.2	0.23	0.23	0.24	0.11	0.06
écart type	0.04	0.04	0.04	0.03	0.03	0.01
I.C.	0,12-0,26	0,22-0,41	0,15-0,32	0,17-0,29	0,06-0,16	0,03-0,09

**Tableau 19: Valeurs du paramètre Fst entre les populations prises deux à deux**

HA= Forastero Haut Amazonien

BA= Forastero Bas Amazonien

CR= Criollo

Tr= Trinitario

I.C. intervalle de confiance

autres groupes. L'allèle 2 du locus cTcCIR6 ne se rencontre pratiquement que chez les Forastero Haut et les Forastero Bas Amazoniens. Des allèles rares se rencontrent dans toutes les populations et certains sont fixés ou sont en voie de fixation. Cinq locus sont fixés chez les Forastero Bas Amazoniens et les Criollo contre seulement un chez les Trinitario et un chez les Forastero Haut Amazoniens. Le test de Hardy Weinberg indique une situation d'équilibre pour neuf locus chez les Forastero Haut Amazoniens, 10 chez les Forastero Bas Amazoniens, 14 chez les Criollo et 16 chez les Trinitario. Les résultats des tests de différenciation des populations deux à deux (162 comparaisons) au niveau des fréquences alléliques sont présentés dans le tableau récapitulatif (Tableau 17) où sont consignés les nombres de différences significatives entre groupes pris deux à deux. Ces résultats indiquent que les Trinitario sont très proches des Criollo tandis que les Forastero Haut Amazoniens sont plus proches des Forastero Bas Amazoniens.

Les différents indices de fixation par locus et par population sont donnés dans le tableau 18. Les valeurs de  $F_{is}$  par population sont de 0,38 pour les Forastero Haut Amazoniens, de 0,48 pour les Forastero Bas Amazoniens, de 0,25 pour les Criollo, et de 0 pour les Trinitario. Le  $F_{st}$  global vaut 0,18 mais est variable de 0,06 à 0,24 (tableau 19) lorsque les populations sont comparées deux à deux. Aucun intervalle de confiance de  $F_{st}$  (tableau 19) ne contient la valeur 0 indiquant que les valeurs de  $F_{st}$  sont significatives. Cependant, la valeur du  $F_{st}$  entre Criollo et Trinitario est très faible indiquant peu de différenciation entre ces deux groupes, comme l'indiquent par ailleurs les distances génétiques. Par contre, de fortes valeurs de  $F_{st}$  sont obtenues pour toutes les autres comparaisons. Quelle que soit la méthode de comparaison utilisée (tableau 17), les Trinitario sont proches des Criollo mais aussi des Forastero Bas Amazoniens. Ces derniers sont proches des Forastero Haut Amazoniens mais en sont suffisamment éloignés pour présenter une différenciation importante comme l'indique la valeur de  $F_{st}$ . Les groupes les plus éloignés sont les Forastero Haut Amazoniens et les Criollo. Les relations entre les différents groupes sont décrites dans le dendrogramme de la figure 10.

### III- DISCUSSION

#### A- DIVERSITÉ RÉVÉLÉE PAR LES RAPD

Contrairement aux cDNA et rDNA qui correspondent à des séquences particulières du génome, les fragments obtenus par l'amplification aléatoire de l'ADN ne sont pas spécifiques d'une région donnée du génome. L'amplification

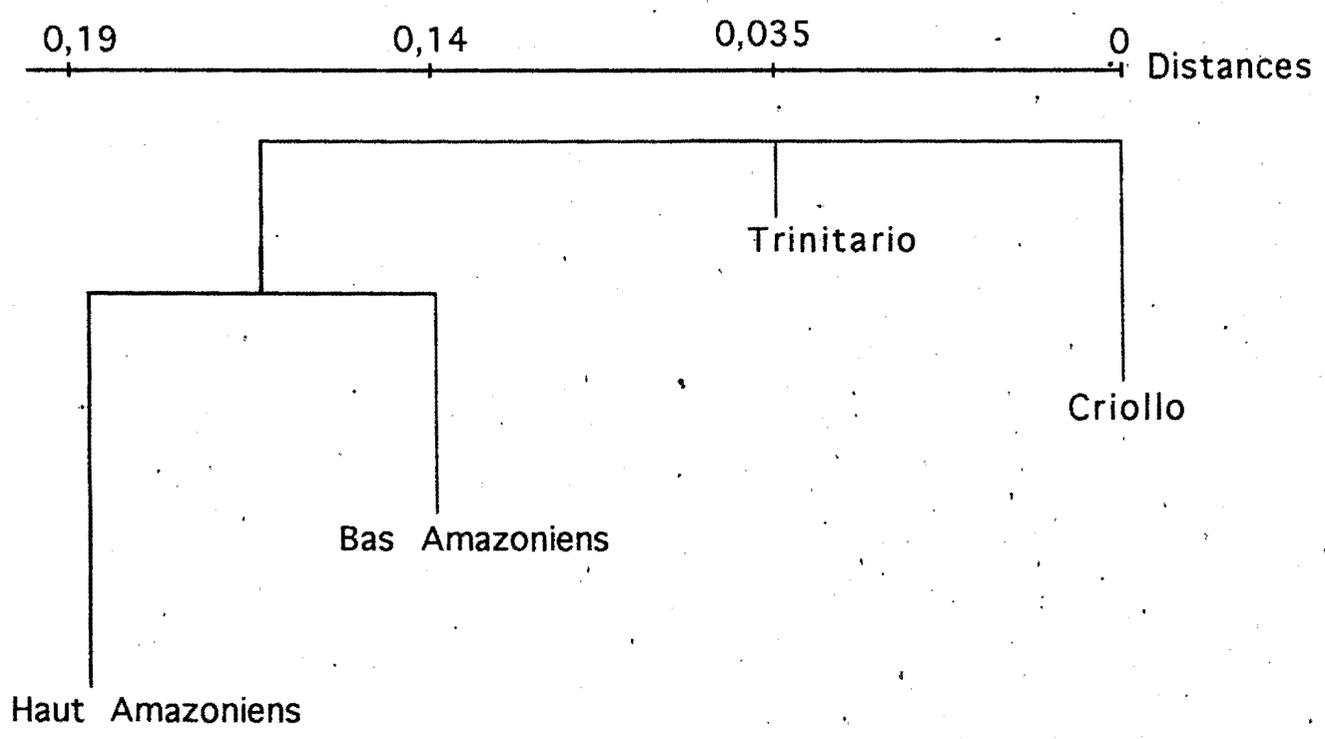


Figure 10: Dendrogramme obtenu à partir des distances de Nei

étant aléatoire, c'est un échantillonnage aléatoire des séquences du génome qui est réalisé. Ces séquences peuvent être uniques ou répétées et peuvent correspondre à des régions codantes ou non codantes du génome. Les séquences uniques du génomes ont servi à des études de diversité par RFLP (Paik-Ro *et al.*, 1992, Besse *et al.*, 1994). Bien souvent, elles ont pu détecter plus de polymorphisme que les isozymes ou les marqueurs morphologiques du fait de leur nombre élevé et de leur répartition variée dans le génome. De ce fait, les classifications obtenues à l'aide de ces séquences sont souvent plus précises que celles obtenues par les isozymes. Quant aux séquences répétées, il en existe de plusieurs types au sein du génome qui sont utiles à des degrés divers dans les études de diversité (Thomas *et al.*, 1993; Besse *et al.*, 1993). Les fragments RAPD peuvent contenir un échantillon varié de toutes les séquences qui existent au sein du génome, il semble donc possible d'obtenir des informations variées sur les différents types de séquences présents dans le génome et ces marqueurs apparaissent donc intéressants pour avoir une image globale du génome qui comprend des séquences uniques et des séquences répétées. Sur le cacaoyer, espèce à petit génome, davantage de séquences répétées ont été amplifiées que de séquences uniques. Ce phénomène pourrait être dû à la technique d'amplification qui serait alors sélective. La banque génomique PstI construite au laboratoire Agetrop à partir de feuilles adultes a révélé un faible taux de séquences répétées (Laurent, com. pers), cependant il faut aussi noter que PstI est une enzyme sensible à la méthylation et de ce fait enrichit considérablement les banques génomiques en séquences uniques.

Une proportion élevée de bandes amplifiées ont les mêmes fréquences entre Forastero et Criollo et les différences observées sont plus grandes entre Forastero et Criollo qu'entre Forastero Bas Amazoniens et Forastero Haut Amazoniens. La majorité des bandes dont les fréquences sont différentes entre Forastero et Criollo sont des séquences répétées. L'analyse de ces séquences répétées a permis d'emblée de différencier les Criollo des Forastero. Chez le cacaoyer, la mise en place de séquences répétées a probablement contribué à différencier les types Forastero et Criollo. Les séquences répétées du génome sont connues pour se mettre en place de manière postérieure aux séquences uniques et contribuent à des différenciations interspécifiques (Flavell *et al.* 1974, Delseny *et al.*, 1988). Les résultats obtenus tendent à conforter l'hypothèse de Pittier et de Cuatrecasas (1964) de deux origines séparées pour les deux groupes originaux que sont les Forastero et les Criollo. Les RAPD ont permis ainsi d'accéder facilement au polymorphisme des séquences répétées dispersées jusqu'alors difficilement observables.

Au niveau de l'AFC, une bonne séparation de toutes les origines morphogéographiques a été obtenue y compris une différenciation nette entre

Forastéro Haut Amazoniens et Bas Amazoniens qui n'avait pas été mise en évidence jusqu'à présent sur le cacaoyer à partir des différents outils de l'étude de la diversité. Les Forastero Haut Amazoniens sont les plus variables de tous et ceci correspond à ce qui a été observé aussi bien avec les marqueurs morphologiques (Pound, 1938,1945), qu'avec les isozymes et récemment avec les sondes RFLP (Laurent *et al.*, 1993). En effet, du fait de l'autoincompatibilité des génotypes Forastero Haut Amazoniens, un brassage perpétuel est réalisé qui contribue ainsi à la création de nouvelles combinaisons d'allèles favorables au maintien ou même à l'élargissement de la variabilité. Un regroupement de certains génotypes en provenance d'Equateur et du Brésil, répertoriés comme Forastero Bas Amazoniens avec les Forastero Haut Amazoniens a été obtenu et n'est pas fortuit. Le Brésil partage sur sa côte Ouest, des frontières avec l'Equateur et le Pérou où les prospections de Pound ont eu lieu et qui ont révélé une grande diversité des Forastero Haut Amazoniens et la présence de nombreux Trinitario. Il semble probable qu'au niveau de ces frontières se trouve une zone hybride au sein de laquelle se rencontrent plusieurs formes. En effet, les génotypes brésiliens prospectés le long des frontières avec l'Equateur et le Pérou (Abud Fonseca, com. pers.) contiennent selon nos résultats, des formes proches des Forastero Haut Amazoniens (cas de Erjoh5 et Erjoh6) des formes Trinitario (Erjoh1) mais aussi des formes Forastero Basses Amazoniennes (Erjoh4, Erjoh15). Les formes indigènes sauvages d'Equateur sont essentiellement des Forastero Haut Amazoniens, qui ont été petit à petit remplacées par des Trinitario (Soria, 1970). Cependant des populations particulières telles le "cacao nacional" caractérisées par Enriquez (1992) typiques d'Equateur sont souvent classées parmi les Forastero Bas Amazoniens. Or les génotypes ECNR, nacional et EET59 incluent dans cette étude et appartenant à la variété nacional d'Equateur ne sont pas des Forastero Bas Amazoniens, selon nos résultats. Ce sont probablement des hybrides entre les variétés Cacao nacional, indigène d'Equateur ou Trinitario introduit et les Forastero Haut Amazoniens avec un fond génétique Forastero Haut Amazonien plus important. De plus au niveau morphologique, ils sembleraient proches des Criollo (Enriquez, 1992). Apparemment, il y a eu une mauvaise classification de la variété "nacional", due à la diversité des origines de cacaoyers en Equateur.

Moins variables que les Forastero Haut Amazoniens alors qu'ils sont souvent autoincompatibles, les Criollo pouvant être considérés comme tels sont en fait assez homogènes. Cependant la difficulté de les différencier des Trinitario et leur regroupement avec ces derniers aboutit à faire des Criollo un groupe variable. Ce sont les premiers cacaoyers domestiqués mais leurs faibles capacités agronomiques les ont petit à petit fait remplacer dans les plantations par des Trinitario avec lesquels ils ont des gènes en commun et de forts liens de parenté,

comme le montre le dendrogramme (Figure 8). Même si on ne trouve presque plus de criollo à l'état sauvage, il est possible de trouver des Trinitario qui leur sont très proches mais possédant une vigueur végétative suffisante pour en faire des géniteurs en vue d'une amélioration variétale. En outre, la découverte d'une importante variabilité au niveau de l'ADN cytoplasmique (Laurent *et al.* 1993) des Criollo rend possible une plus grande utilisation d'individus Criollo dans les programmes d'amélioration génétique. La variabilité des Trinitario étudiés est importante. Ayant une origine hybride, l'identité des Trinitario en tant que groupe ne peut être établie. On les retrouve avec les Forastero Bas Amazoniens, mais surtout avec les Criollo. Il faut envisager leur intégration au sein des Criollo. Pour cela, l'évaluation de la diversité d'un nombre plus important individus Criollo et Trinitario est nécessaire pour déterminer les génotypes les plus proches des Criollo. Les Forastero Haut Amazoniens sont très variables quel que soit le type d'analyse. Ceci est à mettre en relation avec le système de reproduction au sein de ce groupe qui est essentiellement allogame par la présence d'un système d'incompatibilité qui limite les autofécondations et avec le fait que l'essentiel du matériel Forastero Haut Amazonien est constitué de génotypes sauvages ou semi sauvage. Au sein des Forastero Bas Amazoniens, l'autofécondation est la règle et il existe une certaine consanguinité au sein de ce groupe qui apparaît dès lors homogène.

## B- PARAMÈTRES GÉNÉTIQUES DE LA DIVERSITÉ

Les locus cDNA analysés chez le cacaoyer possédaient généralement deux allèles et le nombre maximum d'allèles par locus était de 4. Souvent lorsque les locus avaient plus de deux allèles, les autres avaient de très faibles fréquences dans la population. Ce nombre d'allèles détectés par les RFLP est plus important que ceux des locus isozymes sur le cacaoyer (Lanaud, 1987). La richesse allélique du cacaoyer est comparable à celle de certaines espèces tropicales comme *Psychotria faxlucens* (Perez-Nasser *et al.*, 1993) mais est moins importante que celle de l'hévéa qui présente en moyenne quatre allèles par locus génomique (Besse *et al.*, 1994) ou de celle du chêne qui possède aussi en moyenne quatre allèles par locus isozyme (Bacilieri *et al.*, 1994). Sur les 64 allèles détectés au sein de la population non subdivisée, 63 sont retrouvés au sein des Forastero, et 59 au sein des Criollo. La faible différence en nombre d'allèles détectés entre la population non subdivisée et les deux groupes génétiques montre que tous les deux sont représentatifs de l'espèce *Theobroma cacao*. Alors que le taux de polymorphisme le plus élevé observé par les isozymes est de 66 % (Lanaud 1986), le polymorphisme des locus RFLP varie de 77 % à 96 %, même si au niveau des RFLP un tri préalable des sondes à été fait dont la portée est limitée par le nombre

de génotypes utilisés pour le tri. C'est d'ailleurs au niveau du polymorphisme détecté que les deux groupes Forastero et Criollo apparaissent différents. Ceci suggère qu'il y a davantage de génotypes au sein des Forastero qu'au sein des Criollo et corroborerait l'hypothèse que l'Amérique du Sud est le centre de diversité de l'espèce. Cependant, rien ne permet de dire que les génotypes Forastero d'Amérique du Sud sont, à l'origine des génotypes Criollo d'Amérique centrale comme l'a considéré Cheesman (1944). L'indice  $F_{st}$  de Wright entre les deux populations est différent de zéro et indique une réelle différenciation génétique entre ces deux groupes. Un nombre important de locus présente des déséquilibres au sein des deux populations. La valeur moyenne de  $F_{is}$  au sein des Forastero est plus élevée que celle au sein des Criollo. De plus, alors que par locus il y a une tendance à un déficit des hétérozygotes au sein des Forastero, au sein des Criollo c'est plutôt un excès en hétérozygotes qui est souvent observé. L'excès en hétérozygotes est souvent relié à un avantage sélectif de ceux-ci par rapport aux homozygotes (Jelinski et Cheliak, 1992). Des phénomènes différents entraînant des déséquilibres au sein des locus analysés seraient donc en oeuvre au sein des Forastero et des Criollo. Au sein des Criollo, la valeur moyenne du  $F_{is}$  ne reflète probablement pas l'ampleur réelle du déséquilibre du fait que les valeurs positives et négatives aux locus s'annulent. Certains facteurs oeuvrent en faveur des hétérozygotes alors que d'autres ont un effet contraire. Ce phénomène a déjà été observé chez *Populus tremuloides* où au niveau d'un locus lié à l'adaptabilité à un environnement, les hétérozygotes étaient sélectionnés alors qu'aux autres locus c'étaient plutôt les homozygotes qui l'étaient (Jelinski et Cheliak, 1992). Les Criollo sont les premiers cacaoyers à avoir été domestiqués par les Maya et n'existent pratiquement plus à l'état sauvage. Leur culture est donc plus ancienne que celle des autres cacaoyers. Les Criollo ont sûrement subi des pressions de sélection fortes de la part de l'homme qui n'a continué à cultiver que les cacaoyers qui répondaient à ses exigences. De ce fait, une part importante du  $F_{is}$ , celle qui entraîne un excès d'hétérozygotes est probablement due à l'effet de la sélection pour la vigueur et d'autres caractères agronomiques. En culture, les fécondations croisées sont favorisées par le regroupement sur une surface donnée de plusieurs génotypes. Ceci entraîne l'apparition de nouvelles combinaisons qui sont sélectionnées à l'état hétérozygote. Une autre part du  $F_{is}$  au sein des Criollo peut être due au mode de reproduction. En effet, les Criollo sont quelquefois autogames et un certain nombre de locus peuvent présenter une consanguinité importante. Cependant, pour les locus analysés et à l'équilibre, un taux important d'hétérozygotes a été détecté.

Au sein des Forastero, le  $F_{is}$  est plus important qu'au sein des Criollo, et un déficit en hétérozygotes est observé. Dans ce cas, la structuration en sous

populations semble être la cause car comme nous l'avons vu précédemment, les Forastero sont constitués de plusieurs origines, souvent séparées par des distances géographiques importantes .

Lorsque l'on considère quatre populations au lieu de deux, aucune d'elles ne contient autant d'allèles que la population non subdivisée et bien que des petites différences apparaissent entre populations, elles ne sont pas importantes. Les Forastero Haut Amazoniens qui présentent la même diversité que celle de la population non subdivisée et qui sont beaucoup plus polymorphes que les autres populations, apparaissent comme les plus représentatifs de la population globale. Ce résultat est en accord avec ceux de Lanaud (1987) qui montrent par les isozymes, une grande diversité des Forastero Haut Amazoniens. Les Criollo présentent la même diversité que les Forastero Haut Amazoniens mais sont moins polymorphes que ces derniers.

Les valeurs du Fis par locus au sein des Trinitario font apparaître des excès en hétérozygotes plus importants que ceux des Criollo pour lesquels le Fis global est positif. Une grande variation de Fis indique une situation hétérogène pour chaque locus aboutissant à l'annulation des effets au niveau du Fis global. Les Trinitario étant des hybrides entre Criollo et Forastero Bas Amazoniens, la situation au niveau de leurs locus est variable. Certains locus peuvent subir des effets de consanguinité à l'instar des Forastero Bas Amazoniens, d'autres, comme chez le parent Criollo subiraient des pressions en faveur du maintien de l'hétérozygotie.

Au sein des Forastero Haut et Bas Amazoniens, peu de locus présentent un excès en hétérozygotes et de plus lorsqu'il y a excès, les valeurs de Fis sont très proches de zéro.

Au sein des cacaoyers, au moins deux phénomènes ont donc lieu: l'un à lieu au sein des Forastero et aboutit à des déficits en hétérozygotes et l'autre est présent au sein des Criollo et des Trinitario qui ont des excès en hétérozygotes pour certains locus. Chez les Forastero Bas Amazoniens, la présence d'un déficit en hétérozygotes peut être expliquée par le mode de reproduction. L'autofécondation qui règne dans ce groupe, conduit au cours du temps à la consanguinité qui est par ailleurs responsable de la réduction du nombre de génotypes. Au sein des Forastero Haut Amazoniens, il existe un système d'autoincompatibilité qui favorise l'allogamie (Knight et Rogers, 1955) et qui devrait en principe contribuer à maintenir une importante variabilité au niveau des génotypes. Ce système d'autoincompatibilité n'empêche pourtant pas des autofécondations en présence simultanée d'allopollen et d'autopollen (Lanaud, 1987; Falque, 1993). Il est donc possible qu'une part du déficit soit due à la consanguinité. En effet, lorsque l'on utilise la formule de Nei et Syakudo (1958)

pour estimer le taux d'allogamie, les Forastero Haut Amazoniens ont un taux non négligeable d'autogamie. Chez *Psychotria faxlucens*, arbre tropical allogame autoincompatible, les taux d'allofécondation pour deux populations analysées par Pérez-Nasser *et al*, 1993, sont de 100% mais ces populations ne présentaient pas de déficit en hétérozygotes. Le déficit en hétérozygotes chez les Forastero Haut Amazoniens pourrait aussi être expliqué par la dérive génétique et la structuration de ce groupe en diverses sous populations. En effet, contrairement aux Criollo qui sont domestiqués depuis des siècles, les premiers cacaoyers Forastero Haut Amazoniens ont été découverts à l'état sauvage en 1932 par Pound et le matériel végétal disponible est souvent sauvage ou a subi un ou deux croisements contrôlés suivis d'une sélection. Ce matériel sauvage est souvent composé de populations généralement isolées collectées en des endroits différents et de ce fait échangeant peu ou pas de gènes. Chaque peuplement peut comporter un nombre limité d'individus favorisant ainsi un phénomène de dérive au sein de chacun d'eux.

Les valeurs de  $F_{st}$  entre populations prises deux à deux montrent une différenciation entre Forastero Haut Amazoniens et Bas Amazoniens. Ce résultat explique pourquoi les croisements contrôlés entre ces deux populations aboutissent à l'obtention de descendances hybrides manifestant de l'hétérosis. Par contre la tendance à considérer que les Criollo et les Trinitario sont deux groupes génétiques très différents devrait être abandonnée car ils ne le sont pas génétiquement. Il semble en effet que les Trinitario aient intégré davantage de gènes en provenance des Criollo par rapport à ceux des Forastero Bas Amazoniens et de ce fait sont plus proches des Criollo.

#### IV - CONCLUSION

Les résultats obtenus au cours de ces travaux sur l'analyse génétique de la diversité génétique par les RAPD, l'analyse des paramètres génétiques et par des travaux précédents permettent de discuter de la classification du cacaoyer. La classification en deux sous groupes, FORASTERO et CRIOLLO nous semble en accord avec les résultats des études de diversité génétique entreprises jusqu'à présent, basées sur des caractères morphologiques et sur les marqueurs moléculaires. Il existe donc des bases génétiques à cette distinction qui devrait être maintenue, ces deux groupes devraient être assimilés à deux complexes génétiques intraspécifiques : les éléments à l'intérieur d'un groupe sont plus ressemblants entre eux que ceux entre deux groupes. Cependant des éléments nouveaux viennent affiner cette distinction: les études de diversité au moyen des marqueurs RAPD et les analyses des paramètres génétiques de la diversité révélée

par les RFLP permettent d'envisager dorénavant une subdivision des Forastero en deux sous ensembles : FORASTERO HAUT AMAZONIENS et FORASTERO BAS AMAZONIENS non seulement sur leur aire de répartition mais aussi sur des bases moléculaires. Ces deux sous groupes sont suffisamment proches pour ne pas pouvoir être différenciés sur la base de caractères morphologiques de cabosses et de fèves mais au niveau de la structure de leur ADN des différences importantes ont pu être décelées. Ainsi leurs fréquences alléliques sont différentes au niveau des locus cDNA et au niveau de la structuration inter et intragroupe. Les programmes de sélection à travers le monde ont permis de déceler des différences entre ces deux groupes par l'intermédiaire de la performance de leurs hybrides. Au niveau de la vigueur, la grande variabilité des Forastero Haut Amazoniens fait que même les croisements intragroupes donnent naissance à des hybrides performants. Le groupe des Trinitario connu pour être hybride entre Bas Amazoniens et Criollo ont intégré davantage de gènes Criollo. Il faut rappeler que les premiers Trinitario sont apparus à Trinidad où avaient auparavant été introduits des Criollo environ deux siècles plus tôt. Les flux de gènes ont été si importants entre les deux populations qu'elles ont fini par ne plus pouvoir être différenciées au niveau des fréquences alléliques. Aussi, nous pensons qu'il ne faut pas considérer les Trinitario comme un groupe génétique différencié des autres. Il semble que leur création soit trop récente pour leur avoir permis de développer des caractères distinctifs originaux. On peut considérer qu'il y a trois grands groupes de cacaoyer qui sont les Forastero Haut Amazoniens, les Forastero Bas Amazoniens et les Criollo. Des études de diversité plus approfondies au sein de chaque grand groupe génétique ainsi défini seraient souhaitables afin de pouvoir mettre en évidence et définir les subdivisions qui existent.

Par suite des diverses prospections menées au Pérou et en Equateur et à la faveur de la variabilité qui y est rencontrée, la haute vallée de l'Amazonie est considérée comme le centre de diversité du cacaoyer. Toutefois les études de la structure génétique révèlent certains faits: les cacaoyers d'Amérique du Sud, c'est à dire les Forastero ne sont pas plus diversifiés que ceux d'Amérique centrale que sont les Criollo mais sont plus polymorphes. Ils sont donc plus variables que ceux d'Amérique du centre qui ne manquent pas non plus de variabilité. La forte différenciation des locus selon les origines géographiques et la similarité des diversités géniques supportent l'hypothèse de deux origines séparées des cacaoyers.

# CHAPITRE IV : CARTOGRAPHIE GENETIQUE DU CACAOYER PAR LES MARQUEURS RAPD ET LOCALISATION DE QTLs

## I - INTRODUCTION

La cartographie génétique d'un génome correspond à son repérage régulier par des marqueurs. Des travaux plus anciens de cartographie ont utilisé des marqueurs morphologiques et isozymes mais les cartes génétiques obtenues étaient alors peu denses car les marqueurs morphologiques et isozymes sont peu nombreux. Avec l'avènement des marqueurs moléculaires de type RFLP ou RAPD, dont le nombre est théoriquement illimité, la cartographie génétique a pris un nouvel essor et de plus en plus de cartes denses sont établies chez les végétaux (Echt *et al*, 1994; Hemmat *et al*, 1994; Reiter *et al*, 1992). De développement plus récent que les marqueurs RFLP, les marqueurs RAPD sont aussi très nombreux. Ils ont permis d'élaborer des cartes de densité élevée chez certaines plantes (Carlson *et al*, 1991; Rieseberg *et al*, 1993).

Les cartes génétiques permettent de comprendre l'organisation des génomes. La comparaison de cartes génétiques entre espèces apparentées permet d'appréhender les phénomènes évolutifs ayant abouti à la spéciation. Ainsi, la comparaison des cartes génétiques du maïs, du sorgho et de la canne à sucre a permis de révéler de nombreuses zones chromosomiques conservées entre ces espèces (Lu *et al*, 1994; Grivet *et al*, 1994). Les cartes génétiques sont aussi le point de départ d'une sélection assistée par marqueurs. La démarche consiste à la localisation de QTL (QTL = Quantitative Trait Loci), locus impliqués dans la variation des caractères quantitatifs. Les caractères quantitatifs sont souvent déterminés par plusieurs locus. Classiquement, ces caractères sont analysés par des méthodes statistiques faisant intervenir des moyennes, des variances génétiques et des estimations d'héritabilité (Falconer, 1974). Ces méthodes permettent d'estimer les effets globaux de locus mais pas les effets individuels de chaque locus intervenant dans la construction d'un caractère quantitatif donné. Les cartes génétiques denses du génome permettent de disposer de marqueurs régulièrement répartis sur le génome pouvant avoir des corrélations étroites avec des QTL qui peuvent ainsi être localisés. Elles permettent de développer une stratégie de sélection assistée par marqueurs comme cela a été le cas chez la tomate.

La détection de locus impliqués dans la construction des caractères quantitatifs est basée sur l'analyse des associations de ces caractères avec des marqueurs (de Vienne, 1990). Dans la pratique, la population d'individus devant servir à la recherche de QTL est caractérisée au niveau des marqueurs et également au niveau des caractères quantitatifs. Pour chaque marqueur, plusieurs génotypes sont possibles selon le type de population utilisée. Les individus ayant le même génotype au niveau d'un marqueur sont regroupés et les moyennes des groupes ainsi obtenus sont comparées. En l'absence de QTL, les espérances de ces moyennes sont égales. La mise en évidence d'une différence entre moyennes conduit à la présomption de la présence d'un QTL au voisinage du marqueur.

La première partie de ce chapitre rapporte les essais de cartographie génétique du cacaoyer faits au moyen des RAPD. Dans une deuxième partie, nous abordons la cartographie de locus affectant quelques caractères quantitatifs du cacaoyer.

## **II - CARTOGRAPHIE GENETIQUE PAR LES MARQUEURS RAPD**

### **A- MATERIEL ET METHODES**

#### **1- MATÉRIEL VÉGÉTAL**

Le choix de la descendance à cartographier a été guidé par plusieurs considérations:

Le fait que les marqueurs RAPD soient dominants est une limite quant au choix des populations de cartographie utilisables. Si une population de type F2 est choisie, les estimations de la fréquence de recombinaison sont peu précises et pour augmenter la précision il faut un nombre important d'individus (Allard, 1956). Pour les RAPD les populations backcross sont adéquates pour les analyses mais il n'était pas possible de développer rapidement une vraie descendance backcross de cacaoyers, aussi des populations pseudobackcross ont-elles été utilisées. En effet, certains cacaoyers étant allogames et d'autres autogames, il est possible de trouver des individus présentant un niveau d'hétérozygotie ou d'homozygotie élevé qui peuvent servir de parents de descendance pseudobackcross. Ces descendance sont d'autant plus intéressantes qu'elles sont au champ depuis un certain nombre d'années et permettent d'obtenir des données sur des caractères agronomiquement intéressants. La descendance UPA409 x POR a été choisie sur la base du polymorphisme enzymatique des deux parents permettant aussi de révéler du polymorphisme dans la descendance.

### **CLONE UTILISÉ COMME PARENT FEMELLE : UPA409**

C'est un clone Forastero Haut Amazonien issu d'un croisement frère-soeur dont les ancêtres sont IMC60 et NA34 prospectés par Pound (1945) dans le bassin Amazonien en Équateur et au Pérou. En Côte d'Ivoire, UPA409 a été utilisé comme parent femelle dans de nombreux croisements en vue de la création variétale. C'est un cacaoyer vigoureux, autoincompatible, à développement végétatif rapide. Ses cabosses sont de taille moyenne et ses fèves sont moyennes et violettes. Au niveau de la production, c'est un bon clone mais aussi un bon géniteur. Il est moyennement tolérant à la pourriture brune des cabosses et peu sensible aux attaques des insectes ravageurs que sont les mirides. Six locus enzymatiques y sont à l'état homozygote et un à l'état hétérozygote.

### **CLONE UTILISÉ COMME PARENT MÂLE: POR**

C'est un clone Criollo de la variété Porcelana qui donne un chocolat très apprécié. POR est peu vigoureux mais a de grosses cabosses dont les fèves sont grosses et dodues. La couleur des fèves varie du blanc au rose. Il a été peu utilisé en croisement en Côte d'Ivoire car son introduction y est récente. Il est sensible aux mirides et à la pourriture brune des cabosses. Six locus enzymatiques y sont à l'état hétérozygote et un à l'état homozygote.

La descendance UPA409 x POR est une des meilleures familles sélectionnées pour la production au cours de ces dernières années et en confirmation dans des essais multilocaux en Côte d'Ivoire.

Soixante dix sept individus de cette descendance ont été utilisés pour la cartographie moléculaire et sont basés à Bingerville en Côte d'Ivoire.

## **2- RÉACTIONS D'AMPLIFICATION**

Elles ont été mises au point selon la description donnée dans la première partie de ce document. Le tri des primers polymorphes a été réalisé d'abord sur les parents et 10 individus de la descendance, et ensuite sur les parents et toute la descendance. Ainsi 12 individus ont été testés deux fois. Les lots d'amorces A, D, F, K, L, M, N, O, P, Q, S d'Operon Technologies ont été utilisés.

## **3- ANALYSES GÉNÉTIQUES**

Les fragments amplifiés ont été notés pour leur absence ou leur présence au sein des parents et dans la descendance. Pour chaque fragment polymorphe qui ségrégeait dans la descendance, un test de  $\chi^2$  a été réalisé pour tester l'hypothèse de ségrégation mendélienne du marqueur en se basant sur les effectifs théoriques devant être observés dans le cas d'un backcross. L'analyse de cartographie a été réalisée au moyen du logiciel MAPMAKER (Lander *et al.* 1987). MAPMAKER

utilise la méthode du maximum de vraisemblance comme base des tests et des calculs. Le test de liaison entre marqueurs est réalisé sous les hypothèses de liaison et d'indépendance. On définit le LODscore comme le logarithme décimal du rapport des vraisemblances, des données selon la liaison et en cas d'indépendance. On définit un seuil de LODscore au-dessus duquel l'hypothèse d'indépendance est rejetée. En matière de cartographie, ce seuil est généralement de trois et signifie que l'hypothèse de liaison est 1000 fois plus vraisemblable que celle d'indépendance.

Soit  $n$  le nombre d'individus d'une descendance de cartographie,  $t$  le nombre de classes phénotypiques et  $a_t$  le nombre d'individus dans la classe  $t$ :

La vraisemblance des données est fournie par la formule:

$$V = \frac{n!}{a_1! a_2! \dots a_t!} m_1^{a_1} m_2^{a_2} \dots m_t^{a_t}$$
, où  $m_1$  à  $m_t$  correspondent aux effectifs théoriques des différentes classes, écrits en fonction de la fréquence de recombinaison entre marqueurs. Sous l'hypothèse d'indépendance, la fréquence de recombinaison est  $r=0,5$ . Lorsque l'on veut être strict pour une grande robustesse des groupes de liaison, la valeur seuil de  $r$  peut être fixée à 0.3 ou 0.4.

Pour estimer la fréquence de recombinaison, c'est encore la méthode du maximum de vraisemblance qui est utilisée. Les ordres des marqueurs sont déterminés par une analyse multipoint qui prend en compte les crossing over multiples (Lander et Green 1987).

Les fréquences de recombinaison entre marqueurs sont converties en distances de cartes linéaires et additives exprimées en centimorgans par l'application d'une fonction de cartographie. Deux fonctions de cartographie sont généralement utilisées: celle de Haldane (1919), basée sur l'hypothèse d'indépendance des crossing over est donnée par la formule suivante:

$$d = -\frac{1}{2} \ln(1-2r)$$

et la fonction de Kosambi (1944) qui tient compte de l'interférence qu'il peut y avoir entre différents segments du chromosome pour l'occurrence des crossing overs:

$$d = \frac{1}{4} \ln \frac{1+2r}{1-2r}$$

le terme  $2r$  est appelé coïncidence ( $c$ ) et varie entre 0 et 1. Lorsque  $c=0$  il y a interférence totale et si  $c$  est égal à 1 alors il n'y a pas d'interférence et on retrouve la formule de la fonction de Haldane. La fonction de Haldane a été utilisée pour convertir les fréquences de recombinaisons en distance de carte étant donné qu'aucun fait indiquant la présence de phénomène d'interférence n'a encore jamais été mis en évidence sur le cacaoyer.

## B- RESULTATS

### 1- DÉTERMINISME GÉNÉTIQUE DES MARQUEURS RAPD

La quasi totalité des bandes RAPD observées chez les parents ont été retrouvées dans la descendance. Entre les deux parents, deux types de polymorphisme ont été observés. Certaines bandes étaient présentes chez la femelle et absentes chez le mâle, d'autres par contre étaient présentes chez le mâle et absentes chez la femelle. Les cas de polymorphisme du deuxième type étaient les plus nombreux et correspondaient à ce que nous recherchions étant donné que des locus enzymatiques ayant contribué au choix des parents étaient à l'état homozygote chez le parent femelle UPA409. Ces deux types de polymorphisme entre les deux parents proviennent de situations au niveau des génotypes des deux parents expliquées dans le tableau 20.

		Génotypes possibles de POR			
		AA	Aa	aa	
		Fragment	+	+	-
		RAPD	Polymorphisme observé		
Génotypes de UPA402	AA	+	non	non	oui
	Aa	+	non	non	oui
	aa	-	oui	oui	non

Tableau 20: Croisements induisant du polymorphisme entre les deux parents.

(-)= absence de produits d'amplification (+)= présence de produits d'amplification. En gras, le polymorphisme recherché au niveau du tri réalisé sur les deux parents.

Le polymorphisme dû à l'absence de bandes chez UPA402 et à leur présence chez POR n'a pas toujours abouti à une ségrégation de bandes dans la descendance car la présence de bandes amplifiées chez POR pouvait être due à un état homozygote dominant (AA) ou hétérozygote (Aa) de POR. Seul l'état hétérozygote de POR a abouti à une ségrégation des bandes dans la descendance.

Dans nos conditions, les marqueurs RAPD que nous avons observés sont dominants. L'aspect quantitatif des bandes n'a pas été investigué mais des

MARQUEUR	TAILLE (KB)	CLASSE1 PRESENCE	CLASSE2 ABSENCE	KHI-2	1-1
rOPD05	1.5	37	40	0.12	
rOPD08	0.7	38	7	21.35	***
rOPD15	2	36	36	0.0	
rOPF03b	2.5	37	35	0.05	
rOPF03a	0.4	19	29	2.08	
rOPF12	1.8	40	36	0.21	
rOPF15	0.3	45	30	3.0	
rOPF16	1.8	36	36	0.0	
rOPK06	1.2	25	21	0.35	
rOPK12	1.4	37	32	0.36	
rOPK15	0.9	39	32	0.69	
rOPK16a	1	40	32	0.89	
rOPK16b	0.6	40	32	0.89	
rOPL10a	1.9	33	41	0.86	
rOPL10b	0.8	35	38	0.12	
rOPL11	0.8	39	34	0.34	
rOPL18	0.5	27	47	5.40	*
rOPM02	1.2	35	28	0.78	
rOPM09	1.2	22	45	7.89	**
rOPP09	0.3	32	41	1.11	
rOPP14	0.5	26	46	5.55	*
rOPPI16b	0.5	42	29	2.38	
rOPPI16a	1.2	48	23	8.80	**
rOPQ04	0.5	41	29	2.06	
rOPQ10	0.5	53	17	18.5	***
rOPR19a	0.4	20	37	5.07	*
rOPR19b	0.75	21	33	2.66	
rOPS20	0.7	41	29	2.06	
PGI	-	24	16	1.60	
PGM2	-	30	40	1.43	
ICD	-	39	32	0.69	
MDH1	-	34	36	0.05	
PAC	-	40	33	0.67	
ADH	-	27	36	1.29	

\*\*\*significatif à 0,1 %

\*\*significatif à 1 %

\*significatif à 5 %

Tableau 21: Ségrégation des marqueurs RAPD et tests de Chi<sup>2</sup> dans la descendance UPA409 x POR

différences d'intensité de bandes ont quelquefois été observées, cependant il était difficile d'en tenir compte.

Quelques bandes RAPD particulières ont été mises en évidence. Ces bandes n'étaient pas présentes chez les parents mais sont apparues dans la descendance tandis que d'autres bandes amplifiées chez les parents n'apparaissaient pas dans la descendance. Ces bandes n'ont pas été prises en compte dans l'analyse de cartographie.

## 2- NIVEAU DE POLYMORPHISME

Certains primers n'ont pas donné de produits d'amplification dans nos conditions expérimentales. Il s'agit des amorces D1, O17, R17, R18, S2, S4, S6, S15. Ils représentent environ 4 % des primers testés sur les parents.

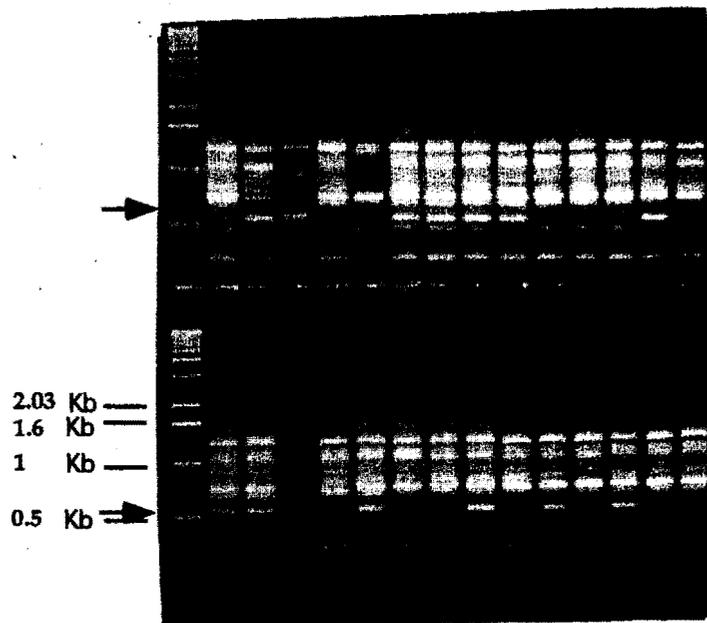
Un total de 186 bandes polymorphes entre les deux parents a été observé. Soixante dix sept produits polymorphes sont dus à la présence de produits d'amplifications chez UPA402 et 109 produits sont dus à la présence de fragments chez POR. Certains lots révélèrent plus de polymorphisme que d'autres. Sur le cacaoyer, les lots de primers F, K, Q et S sont les plus intéressants, alors que les lots L et M le sont moins.

Le nombre de bandes observées par primer variait entre 3 et 20. Celles qui étaient polymorphes variaient par primer entre 0 et 4 et représentent une faible proportion du nombre total de fragments amplifiés. Les profils observés sont caractéristiques de chaque primer. Sur les 198 primers testés sur les parents, 115 étaient polymorphes soit un niveau global de 58%.

## 3- ANALYSES DES SÉGRÉGATIONS

Peu de marqueurs étaient informatifs dans le sens désiré. Seuls ceux présents chez POR, absents chez UPA409 et ségrégeant dans la population de cartographie ont été retenus (Photo 6). 28 marqueurs ont pu être obtenus à partir des 109 fragments. Le tableau 21 présente les résultats des tests de Chi<sup>2</sup> effectués sur la descendance.

Les 28 marqueurs RAPD informatifs ont été cartographiés à l'aide du logiciel MAPMAKER avec un LODscore = 3 et  $r_{max}=0.4$  (fréquence de recombinaison maximum permettant de déclarer l'indépendance des marqueurs), en leur adjoignant les marqueurs isozymes qui ont servi à déterminer le degré d'hétérozygotie des parents. Sept marqueurs RAPD présentent une distorsion de ségrégation sur les 28 et 10 marqueurs (7 marqueurs RAPD et 3 isozymes) ont été réunis en trois groupes de liaison (Figure 11).



**Photo 6: Ségrégation de bandes RAPD au sein de la descendance  
UPA409 x POR (Primer P16)**

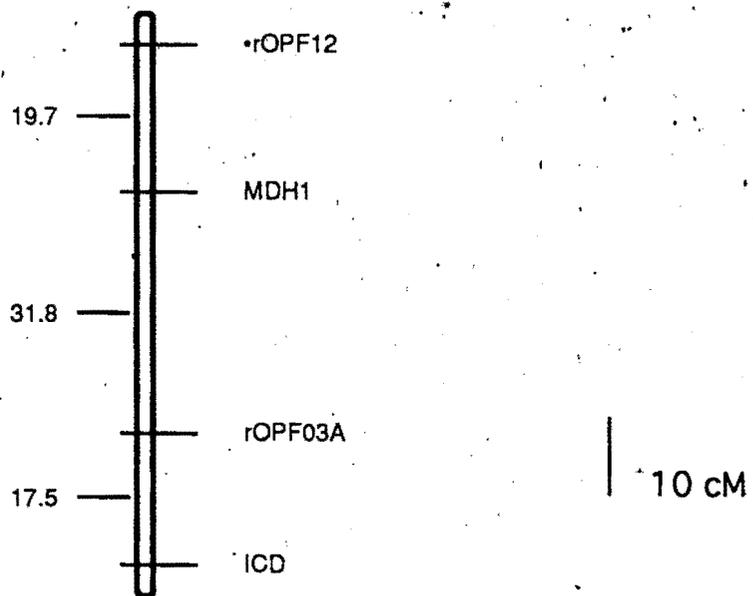
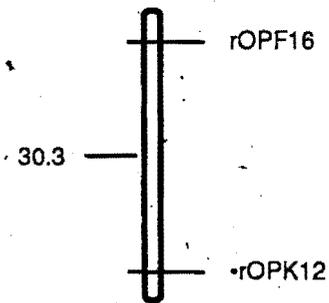
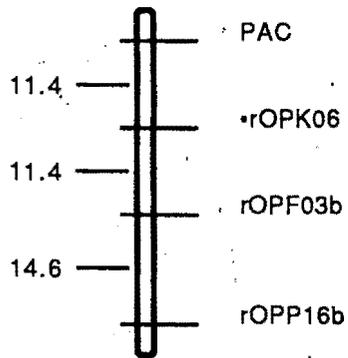
Vingt huit marqueurs RAPDs ont été cartographiés sur une seconde descendance UPA 402 x UF676 en association avec 160 marqueurs RFLP et cinq marqueurs isozymes. C'est cette descendance qui servira à la recherche de QTL.

## 2- REPÉTABILITÉ

Les résultats des tests de répétabilité ont été décrits dans le chapitre précédent, mais il convient de revenir sur certains points importants: certains primers engendrent des produits peu répétables et la quantité d'ADN matrice est déterminante. Bien qu'aucun test n'ait été réalisé sur de l'ADN plus ou moins dégradé, au cours des expérimentations nous avons constaté que les solutions fraîches d'ADN donnaient une meilleure qualité d'amplification que les vieilles solutions. Au cours des décongélations et recongélations successives, il a pu y avoir dégradation de l'ADN, qui expliquerait la mauvaise amplification de certains géotypes lorsque les solutions sont anciennes. Des problèmes de répétabilité peuvent ainsi être engendrés car les sites de fixation peuvent varier et les fragments amplifiés peuvent être de taille différente.

## C- DISCUSSION

Le niveau de polymorphisme révélé par les marqueurs RAPD sur les deux parents est élevé mais moins élevé que celui révélé par les marqueurs isozymes utilisés au départ pour le choix des parents. En effet, sur un nombre total de 7 locus enzymatiques, six sont polymorphes, et ce polymorphisme est directement utilisable en cartographie car les marqueurs isozymes sont codominants. Bien que les parents de la population de cartographie aient un niveau d'hétérozygotie et d'homozygotie élevé au regard des locus isozymes, il semble qu'une part infime du génome ait été accessible aux marqueurs isozymes car de nombreux locus RAPD homozygotes chez les deux parents n'ont pas ségrégué dans la descendance. Tout le polymorphisme RAPD révélé entre les deux parents n'était pas informatif pour la construction d'une carte génétique. Le tri des primers effectué sur les parents n'a pas été très efficace car il n'a pas permis d'éliminer d'emblée les locus non informatifs, ceux qui étaient homozygotes dominants chez le parent mâle. Nous voulions nous situer dans un cas de pseudobackcross où de nombreux locus seraient hétérozygotes chez un parent tandis qu'ils seraient homozygotes chez le second parent. Cette stratégie a souvent été utilisée pour la cartographie des marqueurs RAPD lorsque les vrais backcross n'étaient pas disponibles et a donné généralement de bons résultats (Carlson *et al*, 1991). Chez le cacaoyer, les deux parents présentaient un degré d'hétérozygotie moins élevé que ce qu'on aurait pu attendre d'une plante allogame et on était plus souvent dans une



**Figure 11: 3 groupes de liaisons obtenus avec les marqueurs RAPD**

**Lodscore=3 theta=0,4 fonction de cartographie de Haldane**

situation de F1 que de backcross. La totalité des marqueurs RAPD obtenus se sont comportés comme des marqueurs dominants, en accord avec ce qui est rapporté dans la littérature (Heun et Helentjaris 1993; Chaparro *et al*, 1994). La majorité des marqueurs RAPD qui ségrégeaient dans la descendance avait une ségrégation de type mendélien, cependant une distorsion de ségrégation a été observée pour certains marqueurs. C'est une situation que l'on rencontre assez fréquemment, aussi bien avec des marqueurs RFLP que des marqueurs RAPD (Rowland et Levi, 1994; Tulsieram *et al*, 1992; Chaparro *et al*, 1994; Lanaud *et al*, 1993). Chez certaines espèces le pourcentage de locus présentant une distorsion peut être très élevé. Chez le bananier, Fauré *et al*, 1993, ont obtenus 36 % de distorsions de ségrégation dues à la présence de translocations. Dans notre cas, la distorsion présentée par un des marqueurs pourrait être due à la présence de deux produits différents ayant une taille identique qui ont pu migrer au même endroit. Cette situation a déjà été observée chez les conifères où un seul produit correspondait à plusieurs locus (Tulsieram *et al*, 1992) et chez le cacaoyer où 2/3 des produits amplifiés sur des génotypes au cours d'une étude de diversité étaient répétés (N'goran *et al*, 1994). Des erreurs de génotypage peuvent aussi conduire à des distorsions. En effet, la lecture et la notation des fragments amplifiés se font à partir de photographies de gels dont la qualité peut être déterminante. Cependant la majorité des distorsions peuvent être expliquées par des sélections gamétiques ou zygotiques (Lorieux *et al*, 1994). Dans la descendance utilisée, les biais ne sont ni en faveur du parent femelle ni en faveur du parent mâle, et plusieurs phénomènes peuvent les expliquer. Un de ces phénomènes est le système d'incompatibilité du cacaoyer, de type gamétophytique (Cope, 1962), qui empêche la fusion des gamètes ayant certains génotypes. Ceci correspondrait à une sélection gamétique. Des distorsions de ségrégations ont aussi été observées par Lanaud *et al*, (1993) sur des locus RFLP chez le cacaoyer.

L'apparition de bandes dans la descendance ne provenant pas des parents a déjà été signalée par Dos Santos *et al* (1994) sur les RAPD et les RFLP. Sur le cacaoyer, leur apparition ne peut pas être imputée à des mutations ponctuelles car leur nombre est élevé. Par contre on pourrait les expliquer par des crossing over qui peuvent créer de nouveaux points d'amorçage des primers.

### III- LOCALISATION DE QTL

#### A - MATERIEL ET METHODES

##### 1- MATÉRIEL VÉGÉTAL

Il s'agit de plantes du croisement UPA402 x UF676 dont une partie a été plantée à Divo et une autre à Bingerville près d'Abidjan. Divo se trouve à 200 Km au Nord d'Abidjan. La pluviométrie y dépasse rarement 1200 mm d'eau par an et les sols y sont granitiques. La saison sèche y est très marquée. Bingerville est à 12 Km d'Abidjan. La pluviométrie est assez importante avec 1600 à 1800 mm d'eau par an bien répartie. Les sols sont pauvres et sablonneux. La saison sèche y est peu marquée.

##### *CLONE UTILISÉ COMME PARENT FEMELLE : UPA 402*

C'est un Forastero Haut Amazonien issu d'un croisement frère-soeur dont les ancêtres sont IMC60 et NA34. En Côte d'Ivoire, UPA402 a été utilisé comme parent femelle dans de nombreux croisements. C'est un cacaoyer autoincompatible, à développement végétatif rapide. C'est un bon clone dont la production moyenne annuelle sur dix années de récoltes a été estimée à 1557 Kg de cacao marchand par hectare. C'est aussi un bon géniteur. Il présente par ailleurs une certaine tolérance aux mirides.

##### *CLONE UTILISÉ COMME PARENT MÂLE : UF676*

C'est un clone Trinitario qui a été sélectionné par la compagnie United Fruit au Costa Rica. Comme la majorité des Trinitario, il possède de grosses fèves. Des données précises sur sa sensibilité à la pourriture brune et aux mirides ne sont pas disponibles, néanmoins les Trinitario sont en général sensibles aux piqûres d'insectes.

La descendance sur laquelle ont été effectuées les différentes mesures comporte 100 individus dont 53 sont à Bingerville et 47 à Divo.

La descendance UPA402 x UF676 est une des meilleures familles sélectionnées dans la première phase du programme d'amélioration génétique du cacaoyer en Côte d'Ivoire qui a abouti à la vulgarisation de 12 descendances hybrides en 1972. Testée dans d'autres pays comme la Guyane et le Vanuatu, elle y est très productive. Cette famille est par ailleurs utilisée comme témoin dans les essais d'hybrides afin de comparer la performance des nouveaux hybrides à celle des hybrides vulgarisés.

## 2- CARTE GENETIQUE

La descendance UPA 402 x UF676 a été cartographiée au laboratoire Agotrop du CIRAD par l'équipe de Lanaud (Lanaud *et al*, 1993 ). Il s'agit d'un croisement entre deux clones hétérozygotes dont le plus hétérozygote selon le génotype aux locus enzymatiques est le parent mâle UF676. Une carte génétique complète, de 193 marqueurs a été établie et comporte 28 marqueurs RAPD, 160 marqueurs RFLP et cinq marqueurs isozymes. Les ségrégations de ces marqueurs pouvaient être de type F2 et de type backcross (dans les deux sens). Afin de pouvoir localiser les QTL à l'aide du logiciel MAPMAKER-QTL, un sous échantillon de marqueurs a été cartographié. Il s'agit de ceux à l'état homozygote chez UPA402 mais à l'état hétérozygote chez UF676. De cette manière, on se situe dans un cas de pseudobackcross pour ces locus. Ils constituent environ 85 % des marqueurs. Cette carte établie comporte 168 marqueurs regroupés en dix groupes de liaison correspondant au nombre de chromosomes d'une cellule haploïde du cacaoyer. La distance moyenne entre marqueurs est de 8,23 cM avec une distance minimale de 0 cM et une distance maximale de 43,6 cM évaluées à l'aide de la fonction de cartographie de Haldane. La taille des groupes de liaison varie de 58,08 cM à 183,69 cM et la taille totale de la carte est de 1158,8 cM.

## 3- CARACTÈRES MESURÉES

Les caractères mesurés sont des caractères morphologiques relatifs aux cabosses et aux fèves et des caractères agronomiques. Tous les caractères morphologiques ont été mesurés dans les deux environnements mais certains caractères agronomiques ne l'ont pas été partout. Ainsi, les dégâts de mirides ont été évalués uniquement à Bingerville.

### CARACTÈRES MORPHOLOGIQUES OBSERVÉS:

- poids de cabosses pleines de fèves (PDCAB)
- poids vide de cabosse (PVIDE): c'est le poids de la cabosse vidée des fèves et du rachis central
- nombre de fèves par cabosse (NBFEV)
- longueur de cabosse (LGCAB)
- diamètre de cabosse (DIAMCAB)
- nombre d'ovules par ovaire (NBOVULE)
- poids de fève épluchée (PFEV)
- poids de fèves fraîches (PFF):  $PFF = PDCAB - PVIDE$ , c'est donc le poids de l'ensemble des fèves non dé mucilaginé es d'une cabosse avec le rachis central.
- longueur de fève épluchée (LGFEV)

- épaisseur de fève épluchée (EPFEV)

### **CARACTÈRES AGRONOMIQUES**

- nombre de cabosses saines (SAINES)
- nombre de cabosses pourries (POURCAB)
- nombre total de cabosses produites (TOTCAB)
- pourcentage de cabosses pourries (% POURCAB):  
 $\% \text{ POURCAB} = \text{POURCAB} / \text{TOTCAB} \times 100$
- sensibilité aux mirides (une note de 0 à 4 a été attribuée à l'arbre selon l'importance des dégâts cumulés sur plusieurs années).

### **4- COLLECTE DES DONNÉES**

Pour les caractères morphologiques des cabosses, nous avons essayé d'obtenir 20 cabosses par arbre à partir de pollinisations manuelles. Cependant, pour certains arbres, nous n'avons pas pu obtenir le nombre désiré de cabosses. Ainsi, le nombre de cabosses par arbre varie entre 5 et 20.

Les caractères des fèves peu variables au sein d'un individu, ont été mesurés sur 25 fèves par arbre en mélangeant toutes les fèves de toutes les cabosses et en prélevant un échantillon de 25 fèves au hasard. Pour le poids de fèves, 100 fèves par arbres ont été pesées.

Les ovules ont été comptés sur sept fleurs par arbre. La moyenne et l'écart type ont été déterminés par arbre et pour chaque caractère.

Les données de production sont obtenues en cumulant les différents passages de récoltes. Au cours des récoltes, le nombre de cabosses totales, le nombre de cabosses saines, le nombre de cabosses rongées et le nombre de cabosses pourries sont notés. Le pourcentage de cabosses pourries a été estimé uniquement pour les arbres ayant un nombre total de cabosses supérieur à 10. Les données de deux années de récoltes sont disponibles, de juillet 1991 à Août 1993 pour Divo. Par contre à Bingerville, les données de récoltes vont de mars 1990 à décembre 1993.

Trois jeux de données ont été analysés: un jeu pour chaque localité et un autre qui regroupait les arbres de Divo et de Bingerville. Dans ce dernier jeu ne contenant que des données morphologiques, les valeurs des arbres à Divo ont été pondérées par la moyenne de l'ensemble des arbres de Bingerville en multipliant les valeurs des arbres de Divo par la relation  $m1/m2$  où  $m1$  est la moyenne des arbres de Bingerville et  $m2$  la moyenne de ceux de Divo, afin de rendre les données homogènes.

## 5- METHODES DE RECHERCHE DE QTL

Trois méthodes de détection des QTL sont décrites ici: l'*interval mapping* de Lander et Botstein (1989), appliquée dans le logiciel MAPMAKER-QTL, l'analyse de variance et la méthode des moindres carrés proposée par Haley *et al.*(1994).

### L'INTERVAL MAPPING OU CARTOGRAPHIE PAR INTERVALLE:

C'est une analyse basée sur la méthode du maximum de vraisemblance et des LODscores. A chaque intervalle du genome défini par deux marqueurs adjacents, on calcule la probabilité d'obtenir les données phénotypiques observées en présence d'un QTL. Le LODscore (le logarithme décimal du rapport des vraisemblances des données en présence d'un QTL et en son absence) permet de définir un seuil à partir duquel on accepte l'hypothèse de la présence d'un QTL. Afin de réduire le risque de première espèce (conclure qu'il y a un QTL alors qu'il y en a pas), des valeurs importantes de LODscore sont nécessaires du fait que de nombreux marqueurs sont examinés. La valeur du LODscore seuil adéquat est fonction de la taille du genome de la plante et du nombre de marqueurs. Lander et Botstein (1989) ont réalisé des courbes qui permettent de choisir le LODscore seuil adapté à son matériel végétal. Chez le cacaoyer, un LODscore de 2,7 serait approprié selon ces estimations. Les hypothèses de base de ce modèle sont la codominance des marqueurs QTL et l'absence d'épistasie entre marqueurs.

L'utilisation de l'*interval mapping* suppose la normalité de la distribution des caractères car une perte de puissance de détection survient en cas de non normalité.

### MÉTHODE D'ANALYSE DE VARIANCE

C'est une analyse de marqueurs pris individuellement, qui ne tient pas compte des autres marqueurs. A partir d'un modèle linéaire simple de type  $Y = m + g_i + e_{ij}$  où Y, est la valeur phénotypique au locus marqueur, m est la moyenne générale due au fond génétique,  $g_i$  est l'effet dû au génotype du marqueur et  $e_{ij}$  l'effet de l'environnement, on effectue une analyse de variance qui permet de tester l'hypothèse d'égalité des moyennes des différents génotypes du marqueur par un test F. Soller *et al.* (1976) suggèrent l'utilisation d'un risque  $\alpha$  de 0,05 pour la détermination de la présence d'un QTL par analyse de variance alors que Lander et Botstein (1989) préconisent un  $\alpha$  de 0,001 en faisant valoir que dans une comparaison multiple telle que celle effectuée sur un genome entier, il faut pouvoir garder le risque de première espèce à 0,05 pour chaque comparaison. Cependant, en diminuant le risque de première espèce on augmente celui de deuxième espèce et l'on peut ne pas pouvoir détecter des QTL alors qu'ils existent

réellement. Nous avons choisi pour l'analyse de variance un seuil minimum de 0,005.

La différence entre l'*interval mapping* et la présente méthode réside dans le fait que les marqueurs sont analysés un à un dans l'analyse de variance alors que dans l'*interval mapping*, l'information apportée par deux marqueurs adjacents est prise en compte et les doubles recombinaison sont prises en compte. De plus, l'*interval mapping* détermine la position probable du QTL. Cependant, alors que l'*interval mapping* peut être influencée par une distribution non normale des caractères quantitatifs, la méthode de l'analyse de variance apparaît plus robuste. Rodolphe et Lefort (1993) considèrent que dans le cas d'une F2 ou d'un backcross, l'*interval mapping* est plus puissant que l'analyse de variance lorsque les caractères ont une distribution normale. Darvasi *et al*, (1993) ont déterminé qu'il n'y avait pas de différence entre les deux méthodes au niveau de la puissance des tests par des simulations.

Les deux méthodes peuvent être complexifiées pour tenir compte de certaines situations possibles: au lieu de rechercher un seul QTL par intervalle de deux marqueurs, on peut présumer qu'il en existe plusieurs (Jansen, 1992) à effets faibles et qui peuvent être liés (Rodolphe et Lefort, 1993). Certains auteurs considèrent aussi des effets de dominance des QTL (Hayashi et Ukai, 1994). En matière d'analyse individuelle de locus, on peut également tenir compte de la dominance et de l'épistasie entre marqueurs QTL.

### LA MÉTHODE DES MOINDRES CARRÉS

Haley *et al*, (1994) proposent l'utilisation de la méthode des moindres carrés pour tenir compte de l'information apportée par tous les marqueurs d'un groupe de liaison et pour pouvoir utiliser des modèles complexes mais plus réalistes avec une simplicité de calculs (plusieurs QTL, plusieurs allèles, effets des QTL non liés etc..). Cette méthode des moindres carrés serait aussi applicable à l'*interval mapping* et serait plus puissante que la méthode du maximum de vraisemblance.

Dans tous les cas, la taille des populations et l'importance de l'effet du QTL agissent sur le pouvoir de détection des QTL (Darvasi, 1993). Soller *et al*. (1976) ont ainsi déterminé par *interval mapping* que pour détecter un QTL qui était responsable de 1 % de la variance phénotypique totale, il fallait des effectifs très importants par classe de génotype du marqueur en fonction de la dominance au locus QTL (minimum de 525 en situation de backcross dans laquelle le QTL est dominant); ainsi, il faut des effectifs importants afin de pouvoir détecter des QTL à effet faibles. De plus, ces effectifs doivent être plus importants par classe de génotype du marqueur pour un backcross que pour une F2 si on veut avoir la même puissance de détection.

La stratégie que nous avons utilisée associe les deux méthodes les plus couramment utilisées: nous avons effectué la recherche de QTL à l'aide du logiciel MAPMAKER-QTL (Lincoln *et al* 1993) qui utilise l'*interval mapping* simple et un programme d'analyse de variance élaboré par A. Charcosset. Ceci permet de confronter les résultats avec ceux obtenus par MAPMAKER-QTL.

Tous les caractères morphologiques analysés présentaient une distribution normale (Voir annexe). Quelques caractères agronomiques n'avaient pas une distribution normale: ce sont les dégâts de mirides et le nombre et le pourcentage de cabosses pourries.

Le LODscore seuil pour déclarer la présence d'un QTL dans un intervalle de deux marqueurs a été fixé à 2 afin d'être moins stricte pour détecter un nombre important de QTL. Les effets génétiques et le pourcentage de variance phénotypique expliquée par un QTL ont été estimés au niveau des pics de LODscores. Dans le cas d'un backcross, c'est l'effet additif du QTL qui est estimé alors que les descendance F2 permettent d'estimer l'additivité et la dominance au locus QTL. Pour l'analyse de variance, le risque  $\alpha$  a été fixée à 5 /1000.

La descendance utilisée pour la recherche de QTL est une descendance pseudobackcross, analysée selon un modèle de backcross pour les marqueurs moléculaires mais pour ce qui concerne les QTL, les ségrégations ne sont pas forcément de type backcross. C'est une situation qui devient de plus en plus fréquente avec l'établissement de cartes génétiques chez les espèces allogames où les locus QTL peuvent être sous forme homozygote ou hétérozygote. Nous avons cherché à évaluer quels types d'effets génétiques sont détectés dans cette situation. Il faut envisager plusieurs génotypes du QTL chez les deux parents. Certains ne présentent pas d'intérêt car ils sont non informatifs: c'est le cas d'un croisement au sein duquel il n'y aurait pas de ségrégation pour le QTL. Nous avons recensé cinq types de croisements où il y a ségrégation du marqueur et du QTL.

Soient  $m_1$  et  $m_2$ , les deux allèles d'un marqueur,  $q_1$  et  $q_2$  ceux d'un QTL,  $r$  la fréquence de recombinaison entre le marqueur et le QTL,  $a$  la valeur du génotype  $q_1q_1$ ,  $b$  la valeur du génotype  $q_2q_2$  et  $d$  la valeur de  $q_1q_2$ . Nous développons les relations suivantes en nous inspirant des formules de Asins et Carbonell(1988), les seuls génotypes possibles au marqueur pour la femelle et le mâle sont  $m_1m_1$  et  $m_1m_2$ :

Cas 1 et Cas 2:

Quatre génotypes sont produits pour chaque croisement mais au niveau du marqueur, seuls deux sont observés.

$m_1m_1q_1q_2 \times m_1m_2q_1q_1$

Fréquences	Génotypes	Valeur
$1/4(1-r)^2$	$m_1m_1q_1q_1$	a
$1/4(1-r)^2$	$m_1m_1q_1q_2$	d
$1/4(1-r)^2$	$m_1m_2q_1q_1$	a
$1/4(1-r)^2$	$m_1m_2q_1q_2$	d

$m_1m_1q_1q_2 \times m_1m_2q_2q_2$

Fréquences	Génotypes	Valeur
$1/4(1-r)^2$	$m_1m_1q_1q_2$	d
$1/4(1-r)^2$	$m_1m_1q_2q_2$	-a
$1/4(1-r)^2$	$m_1m_2q_1q_2$	d
$1/4(1-r)^2$	$m_1m_2q_2q_2$	-a

Si tous les QTL étaient sous ces formes, il serait alors impossible d'estimer a ou d en faisant le contraste des deux génotypes du marqueur. La différence entre les valeurs phénotypiques des individus possédant les deux génotypes du marqueur s'annule.

Cas 3. et 4

Il s'agit là de cas de backcross classiques:

$m_1m_1q_1q_1 \times m_1m_2q_1q_2$

Fréquences	Génotypes	Valeur
$1/2(1-r)^2$	$m_1m_1q_1q_1$	a
$1/2(1-r)^2$	$m_1m_2q_1q_2$	d
$1/2 r(1-r)$	$m_1m_1q_1q_2$	d
$1/2 r(1-r)$	$m_1m_2q_1q_1$	a

$m_1m_1q_2q_2 \times m_1m_2q_1q_2$

Fréquences	Génotypes	Valeur
$1/2(1-r)^2$	$m_1m_1q_1q_2$	d
$1/2(1-r)^2$	$m_1m_2q_2q_2$	-a
$1/2 r(1-r)$	$m_1m_1q_2q_2$	-a
$1/2 r(1-r)$	$m_1m_2q_1q_2$	d

Dans ces deux cas, le contraste donnerait les relations suivantes:

$$Y(m_1m_1) - Y(m_1m_2) = 0,5(1-r)(1-2r)a - 0,5(1-r)(1-2r)d$$

$$Y(m_1m_1) - Y(m_1m_2) = 0,5(1-r)(1-2r)a + 0,5(1-r)(1-2r)d$$

On estime des effets génétiques dus à l'additivité et à la dominance du QTL mais on ne peut pas séparer les deux types d'effets. Par contre si l'on fait l'hypothèse d'absence de dominance au QTL on n'estime alors que de l'additivité.

Cas 5

$m_1m_1q_1q_2 \times m_1m_2q_1q_2$

Huit génotypes sont produits et le contraste entre les deux classes du marqueur permet d'estimer la valeur de a. Dans ces conditions, le contraste entre les deux génotypes du marqueur permet d'estimer a, l'effet additif:

$$Y(m_1m_1) - Y(m_1m_2) = 0,5(1-r)(1-2r)a$$

	UF676			UPA402			Effectifs	Différence
	Moyenne	Ecart type	C.V.	Moyenne	Ecart type	C.V.		
NBOVULE	42.8	2.09	4.88	50.3	2.98	5.92	16	**
PDCAB	504.18	213.52	42.35	337.31	89.09	26.41	16	ns
LGCAB	157.81	27.36	17.34	121.31	6.88	5.67	16	**
DIAMCAB	82.1	10.09	12.29	72.4	6.88	9.50	16	**
PVIDE	385.12	143.5	37.26	228.67	53.08	23.21	16	**
NBFEV	26.08	9.88	37.88	36.6	10.04	27.43	16	**
PFF	119.06	72.89	61.22	122.93	54.62	44.43	16	ns
PFEV	298.3	46.26	15.51	151.3	20.25	13.38	100	**
LGFEV	30.47	1.64	5.38	23.56	2.72	11.54	25	**
EPFEV	10.32	1.32	12.79	6.91	0.9	13.02	25	**

Tableau 22: Valeurs des parents pour divers caractères morphologiques à Bingerville (ns= non significatif; \*\* significatif à 1 pour 1000)

	UF676			UPA402			Effectifs	Différence
	Moyenne	Ecart type	C.V.	Moyenne	ecart type	C.V.		
NBOVULE	-	-	-	-	-	-	-	-
PDCAB	1000.6	203.57	20.34	553.75	109.49	19.77	16-20	**
LGCAB	219.5	18.76	8.55	151.7	8.45	5.57	16-20	**
DIAMCAB	98.43	6.81	6.92	83.4	6.15	7.37	16-20	**
PVIDE	717.87	178.40	24.85	374.15	77.06	20.60	16-20	**
NBFEV	43.87	3.13	7.13	49.05	4.26	8.69	16-20	**
PFF	282.75	41.07	14.53	179.6	36.37	20.25	16-20	**
PFEV	285.7	16.62	5.82	165.3	11.54	6.98	100	**
LGFEV	28.11	1.00	3.56	24.2	0.62	2.56	25	**
EPFEV	10	2.70	27.00	6.32	1.6	25.32	25	**

Tableau 23: Valeurs des parents pour divers caractères morphologiques à Divo (ns= non significatif; \*\* significatif à 1 pour 1000)

Fréquences	Génotypes	Valeur
$1/4(1-r)^2$	$m_1m_1q_1q_1$	a
$1/4(1-r)^2$	$m_1m_1q_1q_2$	d
$1/4(1-r)^2$	$m_1m_2q_1q_2$	d
$1/4(1-r)^2$	$m_1m_2q_2q_2$	-a

Fréquences	Génotypes	Valeur
$1/4 r(1-r)$	$m_1m_1q_1q_2$	d
$1/4 r(1-r)$	$m_1m_1q_2q_2$	-a
$1/4 r(1-r)$	$m_1m_2q_1q_1$	a
$1/4 r(1-r)$	$m_1m_2q_1q_2$	d

Les situations décrites ci-dessus sont toutes probables au sein du génome et en les considérant toutes, la différence entre les deux génotypes du marqueur donne la relation suivante:

$$Y(m_1m_1) - Y(m_1m_2) = 3/2(1-r)(1-2r)a$$

Cette relation qui ne contient pas de terme en d indique qu'en fait, dans notre cas les effets génétiques des QTL qui seront estimés au cours de cette analyse sont uniquement de l'additivité mais dépendent de la valeur de la fréquence de recombinaison. Dans ces descendance nous n'avons pas besoin de faire l'hypothèse d'absence de dominance au QTL afin d'estimer les effets additifs.

La part de variance phénotypique expliquée par le QTL est donnée par le paramètre  $R^2$  qui est aussi le coefficient de détermination. C'est le rapport de la somme des carrés du marqueur et de la somme des carrés totale. C'est  $R^2$  qui permet d'évaluer l'importance de l'effet du QTL.

## B - RESULTATS

### 1- CARACTERISTIQUES MORPHOLOGIQUES DES DEUX PARENTS UPA402 ET UF676

Les caractères morphologiques mesurés sont variables entre les deux parents (tableaux 22 et 23) dans les deux localités. Certains caractères sont très variables au sein d'un individu: il s'agit du poids de fèves fraîches(PFF), du poids de cabosses(PDCAB) et du nombre de fèves par cabosse (NBFÉV). D'autres sont peu variables: nombre d'ovules par ovaire (NBOVULE) et longueur de fèves (LGFEV). Il existe une différence significative entre les deux parents pour la quasi totalité des caractères et le parent UF676 est supérieure à UPA402 pour tous les caractères à l'exception du nombre de fèves par cabosse et du nombre d'ovules par ovaire.

### 2- CORRELATIONS DES CARACTERES

Les coefficients de corrélations entre les différents caractères sont données dans les tableaux 24 et 25. A 5% et 1%, ces coefficients sont souvent significatifs. Il apparaît ainsi que de nombreux caractères morphologiques sont corrélés: une

	NB	PDCAB	LGCAB	DIAMCAB	PVIDE	NBFEV	PFF	PFEV	LGFEV	EPFEV
LGCAB	38	0,83**								
DIAMCAB	38	0,911**	0,58**							
PVIDE	38	0,95**	0,83**	0,87**						
NBFEV	38	0,41*	0,44**	0,2	0,24					
PFF	38	0,85*	0,62**	0,78**	0,66**	0,59**				
PFEV	37	0,39*	0,17	0,48**	0,26	-0,16	0,52**			
LGFEV	37	0,39*	0,11	0,52**	0,23	0,09	0,56**	0,82**		
EPFEV	37	0	0,15	-0,056	0,09	-0,34	-0,15	0,17	0,03	
NBOVULE	46	0,14	-0,024	0,2	0,013	0,28	0,33*	0,15	0,17	-0,31*

	NB	SAINES	POURCAB	TOTCAB	MIRIDES
POURCAB	50	0,67**			
TOTCAB	50	0,99**	0,74**		
MIRIDES	41	0,14	0,18	0,15	
%POURCAB	44	-0,16	0,59**	-0,07	-0,021

**Tableau 24: Valeurs des coefficients de corrélations entre les caractères à Bingerville**

(NB est le nombre d'individus analysés pour chaque caractère : \*significatif à 0.05 et \*\* significatif à 0.01)

	NB	PDCAB	LGCAB	DIAMCAB	PVIDE	NBFEV	PFF	PFEV	LGFEV	EPFEV
PDCAB	25									
LGCAB	25	0,44*								
DIAMCAB	25	0,91**	0,17							
PVIDE	25	0,92**	0,38*	0,88**						
NBFEV	25	0,46*	0,33	0,36	0,16					
PFF	25	0,72**	0,38*	0,58**	0,4	0,81**				
PFEV	25	0,35	0,16	0,41*	0,15	0,47*	0,6			
LGFEV	25	0,42*	-0,14	0,58**	0,32	0,43*	0,18	0,72**		
EPFEV	25	-0,04	0,29	-0,12	-0,05	-0,17	0,17	0,49**	0,13	
NBOVULE	36	0,39	0,46*	0,38**	0,24	0,34	0,54**	0,58**	0,55**	0,2

	NB	SAINES	POURCAB	TOTCAB
POURCAB	37	0,43*		
TOTCAB	37	0,98**	0,55**	
%POURCAB	33	-0,17	0,62**	-0,049

**Tableau 25: Valeurs des coefficient de corrélations entre les caractères à Divo**

(NB est le nombre d'individus analysés par caractère et \*\*=significatif à 0.01, \*=significatif à 0.05)

forte corrélation existe entre le poids de cabosse, la longueur de cabosse, le diamètre de cabosse, le poids vide, et le poids de fève fraîche. Par contre la corrélation des caractéristiques des cabosses et celles des fèves est moins importante. Le nombre d'ovules par ovaire est peu corrélé avec les autres caractères morphologiques. Au niveau des caractères agronomiques, il y a une corrélation importante entre le nombre total de cabosses produites et le nombre de cabosses pourries. Le pourcentage de cabosses pourries est quant à lui corrélé avec le nombre de cabosses pourries.

### 3- LOCALISATION DES QTL DE LA POPULATION DE BINGERVILLE

Les histogrammes de distribution des caractères sont donnés en annexe. Tous les caractères morphologiques ont une distribution quasi normale. Par *interval mapping*, et avec un LODscore seuil de 2.7, huit régions chromosomiques affectant huit caractères ont été détectées à l'aide du logiciel MAPMAKER-QTL (tableau 26). Avec un seuil de 2, ce sont 10 régions qui sont détectées (tableau 26). Par analyse de variance et pour un seuil de probabilité de 5 pour 1000, 15 QTL intervenant dans la variation de 10 caractères ont été mis en évidence (tableau 27). Dans les deux analyses, lorsque les effets d'un QTL se répercutaient sur plusieurs marqueurs d'un même groupe de liaison, seul le marqueur ayant le LODscore le plus élevé ou la valeur de  $R^2$  la plus importante (variance expliquée par le QTL) a été retenu et noté dans les tableaux de résultats. Si deux pics de LODscore ont été obtenus au sein d'un même groupe de liaisons alors les marqueurs ayant les LODscore les plus élevés ont été notés. Pour certains caractères, une seule région chromosomique a été détectée (nombre de fèves par cabosses (NBFEV), poids de fèves fraîches (PFF) etc..). Deux ou trois QTL ont été détectés pour d'autres caractères (longueur de cabosse (LGCAB), poids de fèves (PFEV) et nombre d'ovules par ovaire (NBOVULE)). Les caractères agronomiques pour lesquels des QTL ont été mis en évidence sont le nombre et le pourcentage de cabosses pourries, et les dégâts de mirides. Les QTL affectant le nombre et le pourcentage de cabosses pourries sont localisés sur le groupe 1 (Figure 12) et expliquent une part importante de la variation de ces caractères (0.21-0.32). Le QTL pour la sensibilité des arbres aux dégâts de mirides est localisé sur le groupe 2. Les caractères fortement corrélés ont des QTL localisés sur les mêmes groupes de liaison. Aucun QTL de la production de cabosses n'a pu être détecté. Les valeurs de  $R^2$  varient de 0.19 à 0.34 pour les QTL détectés par *interval mapping* et de 0.19 à 0.42 pour ceux détectés par analyse de variance. Ces valeurs élevées de  $R^2$  sont normales pour les effectifs relativement faibles utilisés par caractère et qui ne permettent de localiser que les QTL à effets forts. Les effets génétiques des locus détectés sont variables: le QTL impliqué dans la variation de la longueur de

	Moyenne	écart-type	Lod max	R2	effets	marqueur	groupe
LGCAB	141.4	14.5	2.96	0.31	16.34	cTcCIR202	gr5
NBFEV	39.53	5.96	3.08	0.33	6.82	cTcCIR62(-1.2)	gr3
PFF	149.18	41.71	2.76	0.29	44.87	gTcCIR149	gr2
PFEV	180.6	35.57	2.78	0.33	41.42	cTcCIR217(-4.4)	gr10
LGFEV	24.71	2.23	3.37	0.34	2.63	cTcCIR86	gr2
NBOVULE	46.68	2.72	2.72	0.27	-2.87	cTcCJR56(-3.3)	gr5
-	46.68	2.72	3.01	0.27	2.83	gTcCIR119	gr4
-	46.68	2.72	2.1	0.19	2.4	cTcCIR202	gr5
POURCAB	6.44	6.47	3.77	0.32	7.45	cTcCIR79(2)	gr1
%POURCAB	9.55	7.56	2.05	0.2	6.89	cTcCIR46(+3)	gr1

Tableau 26: caractéristiques des QTL révélés à Bingerville par Interval mapping  
(la position du Qtl est entre parenthèses)

	Moyenne	écart type	R2	effets	Marqueur	groupe
LGCAB	141.4	14.5	0.22	-6.5	cTcCIR43	gr2
-	-	-	0.42	-8.4	cTcCIR202	gr5
PVIDE	262.28	67.6	0.21	30.3	cTcCIR83	gr1
-	262.28	67.6	0.22	-29	cTcCIR202	gr5
NBFEV	39.53	5.96	0.31	3.32	cTcCIR62	gr3
PFF	149.18	41.71	0.31	-22	rOPM3/0.8	gr2
PFEV	180.6	35.57	0.23	-17	rOPQ10/2.1	gr9
-	-	-	0.26	18.5	cTcCIR217	gr10
LGFEV	24.71	2.23	0.34	-1.3	rOPM3/0.8	gr2
NBOVULE	46.68	2.72	0.19	-1.2	rTcCIR133	gr2
-	-	-	0.31	-1.5	gTcCIR119	gr4
-	-	-	0.24	1.42	cTcCIR68	gr5
POURCAB	6.44	6.47	0.28	7.45	cTcCIR79	gr1
%POURCAB	9.55	7.56	0.22	3.23	cTcCIR77	gr1
MIRIDE	1.24	0.79	0.21	0.35	gTcCIR135	gr2

Tableau 27: caractéristiques des QTL révélés à Bingerville par analyse de variance

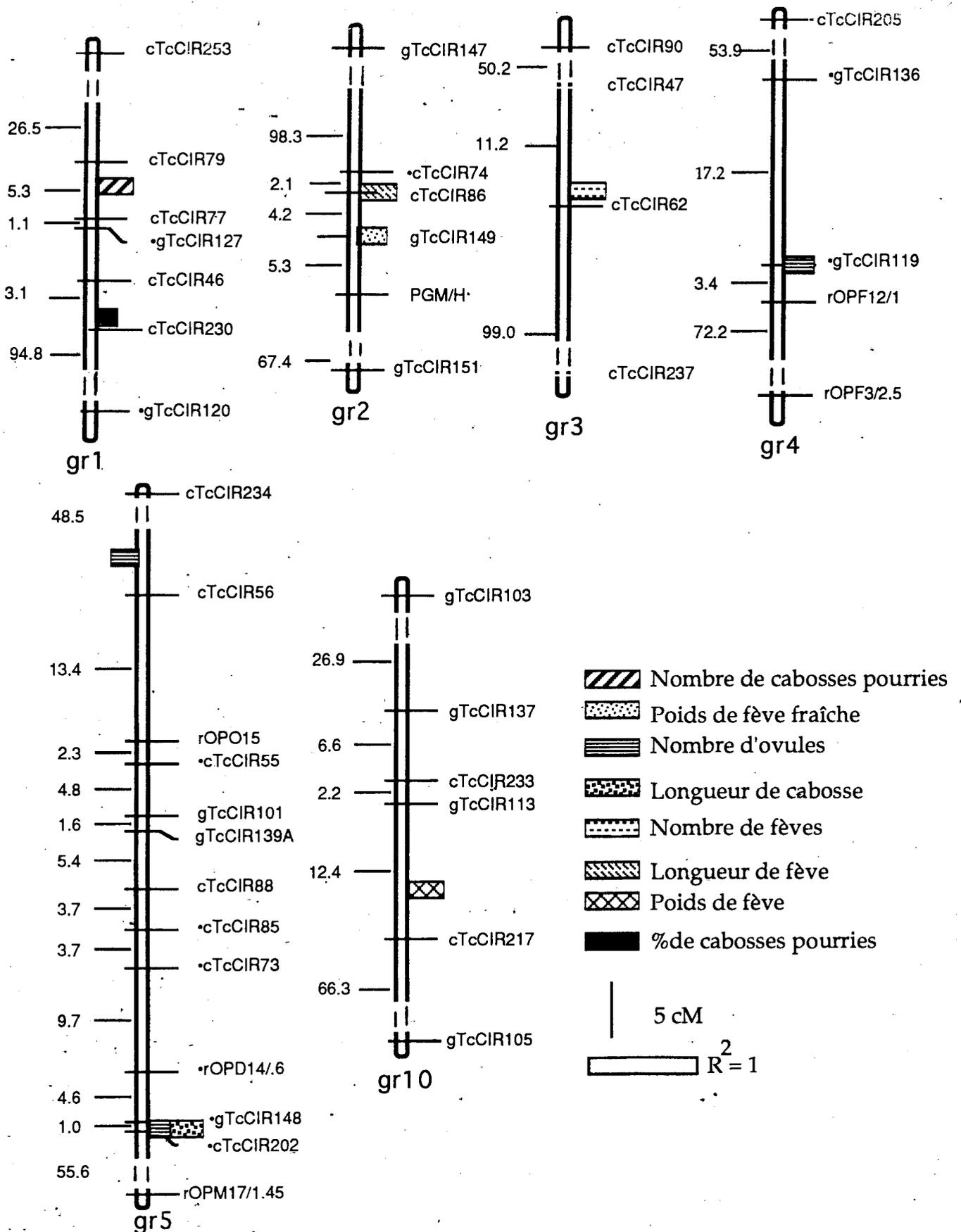


Figure 12 : Localisation des QTL détectés à Bingerville par interval mapping (à droite les QTL à effets positifs et à gauche les QTL à effets négatifs)

	Moyenne	écart type	R2	effets	marqueur	groupe	Lod max
PDCAB	469.35	106.7	0.51	-152.9	cTcCIR98(2)	gr8	2.88
LGCAB	153.96	11.22	0.35	14.34	cTcCIR228(4)	gr9	2.05
DIAMCAB	80.18	6.68	0.41	-8.6	cTcCIR98(4)	gr8	2.29
PVIDE	328.79	80.95	0.43	114.6	cTcCIR228(8)	gr9	2.73
	-	-	0.41	-105	cTcCIR98(8)	gr8	2.12
EPFEV	7.97	0.79	0.41	1.03	gTcCIR136	gr4	2.24
	-	-	0.42	-0.97	cTcCIR237	gr3	2.45
POURCAB	5.32	6.57	0.26	6.72	cTcCIR203(-1.7)	gr8	2.25
%POURCAB	10.58	11.57	0.31	12.88	cTcCIR203(-1.5)	gr8	2.29

Tableau 28: caractéristiques des QTL révélés à Divo par Interval mapping avec un LODscore seuil de 2 (la position du QTL est entre parenthèses)

	Moyenne	écart type	R2	effets	Marqueur	groupe
PDCAB	469.35	106.7	0.57	75.08	rOPP14/0.4	gr8
LGCAB	153.96	11.22	0.35	7.1	gTcCIR102	gr9
DIAMCAB	80.18	6.68	0.63	4.78	rOPP14/0.4	gr8
PVIDE	328.79	80.95	0.56	54.15	rOPP14/0.4	gr8
EPFEV	7.97	0.79	0.41	-0.48	cTcCIR237	gr3
POURCAB	5.32	6.57	0.25	-3.29	cTcCIR203	gr8
%POURCAB	10.58	11.57	0.35	7.62	cTcCIR63	gr8
	-	-	0.38	-8.76	gTcCIR105	gr10

Tableau 29: caractéristiques des QTL révélés à Divo par analyse de variance avec un seuil alfa de 0.005

cabosse a des effets positifs sur le marqueur cTcCIR202 du groupe de liaison 5 quand ils sont estimés par *interval mapping* alors qu'ils sont négatifs sur ce même marqueur lorsqu'ils sont estimés par analyse de variance. De même, les effets du QTL affectant le poids de fèves fraîches sont positifs sur les marqueurs gTcCIR149 mais négatifs sur rOPM3/0.8, tous les deux marqueurs faisant partie du groupe de liaison 2 (Figure 12).

#### 4- LOCALISATION DES QTL DE LA POPULATION DE DIVO

Comme à Bingerville, des QTL ont été détectés pour la majorité des caractères morphologiques et agronomiques (tableaux 28 et 29). Aucun QTL de la production en cabosses n'a pu être détecté. Les QTL affectant les caractères de cabosses et de fèves sont localisés sur les groupes de liaison 3, 4, 8, 9 (figure 13). Pour les mêmes caractères, les régions chromosomiques impliquées sont différentes de celles détectées à Bingerville. Aucun QTL décelé à Divo n'est identique à ceux de Bingerville. A Divo, les valeurs des  $R^2$  varient de 0.26 à 0.57 et sont généralement plus élevées que celles de Bingerville. Un QTL du diamètre de cabosse a été détecté à Divo alors qu'aucun n'avait pu être mis en évidence à Bingerville. A Divo, deux régions chromosomiques dont les effets sont opposés sont impliquées dans la variation du pourcentage de cabosses pourries. Elles expliquent au total 73% de la variation de ce caractère (tableau 29). Comme à Bingerville, les effets génétiques de certains QTL sur les mêmes marqueurs peuvent être positifs ou négatifs selon la méthode d'estimation (par exemple, le nombre de cabosses pourries).

#### 5- LOCALISATION DES QTL AU SEIN DE LA POPULATION COMPOSITE

Pour les caractères communs aux deux environnements, les valeurs des individus de Divo ont été pondérées par la moyenne des individus de Bingerville afin de les homogénéiser. On obtient alors un effectif plus important par caractère analysé. Des QTL ont été détectés pour tous les caractères de cabosses mais pas pour tous les caractères de fèves (tableaux 30 et 31). Deux régions, affectant la longueur de cabosse ont été mises en évidence sur les groupes de liaison 5 et 8 (Figure 14); l'une d'elle avait déjà été détectée par l'analyse des données de Bingerville; par contre la région chromosomique mise en évidence par l'analyse des données de Divo n'a pas été retrouvée au sein de la population composite pour ce caractère. Pour le nombre d'ovules par ovaire, trois régions chromosomiques sur les groupes 2, 4 et 6 ont été détectées dont deux avaient aussi été détectées à Bingerville. Aucun QTL de la longueur des fèves et du poids de fèves n'a pu être détecté dans cette analyse conjointe alors qu'une région avait été détectée à Bingerville et à Divo. Les valeurs de  $R^2$  varient de 0.11 à 0.27 et sont

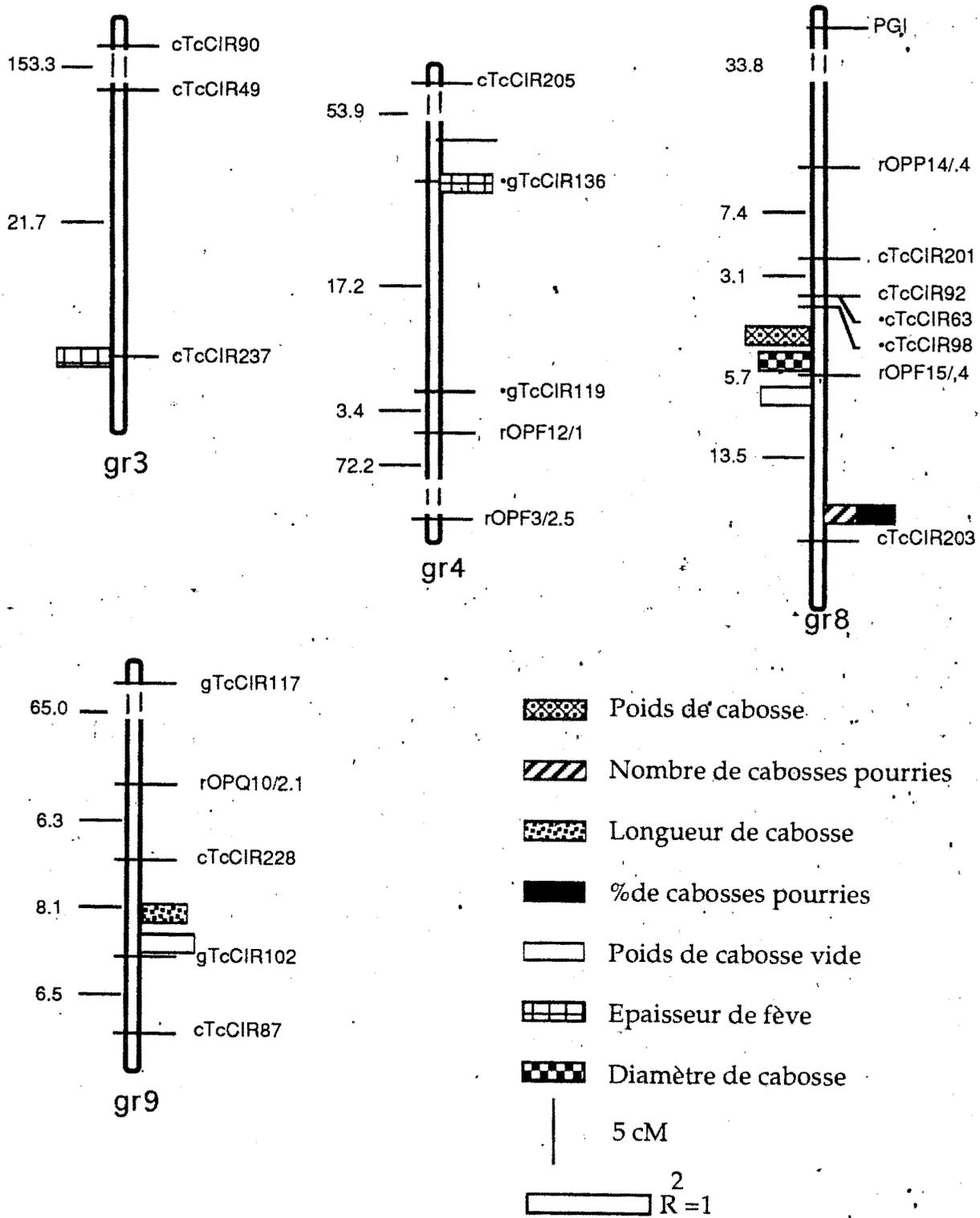


Figure 13 : Localisation des QTL détectés à Divo par interval mapping (QTL à effets positif à droite et QTL à effet négatifs à gauche)

	Moyenne	écart-type	Lod max	R2	effets	marqueur	groupe
LGCAB	140.92	13.11	2.72	0.19	11.55	cTcCIR202	gr5
			2.50	0.20	11.70	PGI	gr8
PVIDE	261.35	66.30	2.15	0.16	52.51	cTcCIR202	gr5
			2.78	0.20	-54.44	cTcCIR83	gr1
PFF	149.11	43.69	2.38	0.17	36.50	gTcCIR137	gr10
			2.36	0.22	-40.82	IDH	gr4
NBOVULE	46.49	2.54	2.18	0.11	1.72	gTcCIR119	gr4
			2.66	0.14	1.87	gTcCIR149	gr2

**Tableau 30: caractéristiques des QTL mis en évidence à Divo et Bingerville par Interval mapping avec un LODscore seuil de 2**  
( la position par rapport au marqueur est donnée entre parenthèses)

	Moyenne	écart type	R2	effets	Marqueur	groupe
PDCAB	409.45	97.69	0.16	-36.60	gTcCIR146	gr4
			0.17	-36.60	cTcCIR202	gr5
LGCAB	140.92	13.11	0.27	7.46	PGI	gr8
			0.24	-5.83	cTcCIR202	gr5
DIAMCAB	73.07	6.18	0.18	2.43	rOPP14/0.4	gr8
PVIDE	261.35	66.30	0.19	29.56	cTcCIR83	gr1
			0.20	15.94	cTcCIR202	gr5
NBFEV	39.47	8.26	0.18	-26.40	gTcCIR150	gr9
			0.16	3.43	IDH	gr4
PFF	149.11	43.69	0.20	-18.90	gTcCIR146	gr4
			0.18	18.49	gTcCIR137	gr10
PFEV			0.15	-13.00		gr2
EPFEV	7.83	1.03	0.20	-0.38	cTcCIR51	gr3
NBOVULE	46.49	2.54	0.13	-0.89	gTcCIR149	gr2
			0.13	-0.94	gTcCIR119	gr4
			0.11	0.85	gTcCIR116	gr6

**Tableau 31: caractéristiques des QTL détectés à Divo et Bingerville par analyse de variance avec un seuil de 0.005**

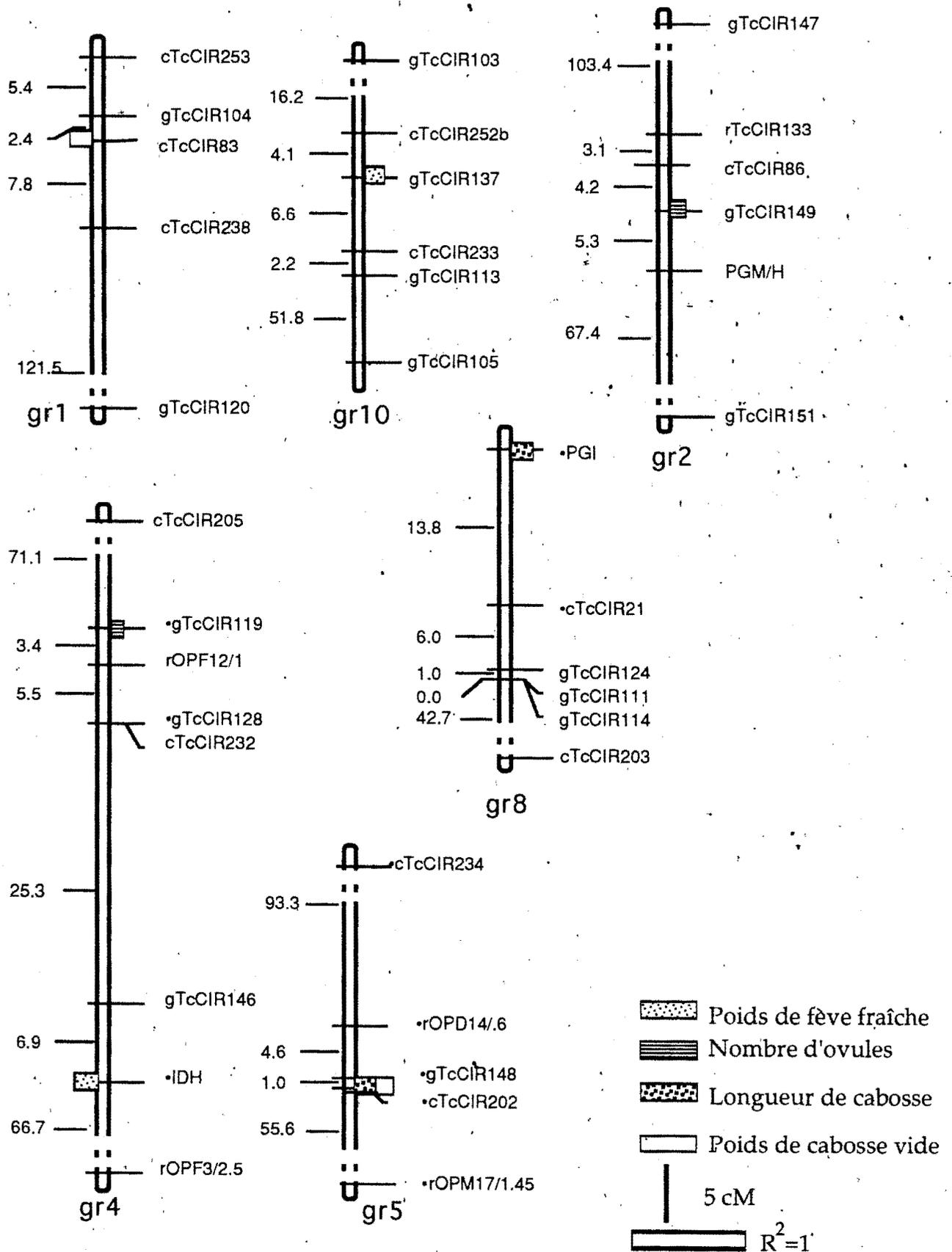


Figure 14: Localisation des QTL détectés par l'analyse conjointe des données de Divo et de Bingervillè par l'interval mapping (QTL à effets positifs à droite, ceux à effets négatifs à gauche)

globalement inférieures à celles estimées par l'analyse des données de Divo et de celles de Bingerville.

## C- DISCUSSION

Il est important de souligner que du fait de la taille des populations utilisées cette recherche de QTL est exploratoire et que les résultats nécessitent une confirmation par l'utilisation d'une population de taille plus importante. De plus, pour pouvoir estimer tous les effets génétiques des QTL (additivité et dominance), il s'avérera nécessaire d'effectuer la recherche sur une population F2 mais chez les espèces allogames fortement hétérozygotes à cycle reproductif long, ces descendances sont difficiles à obtenir et de plus en plus de cartes génétiques sont établies à partir d'individus hétérozygotes.

Cette recherche de QTL a été effectuée sur des populations d'effectifs faibles. Elle est comparable à celle réalisée par Keim *et al* (1990) sur une population F2 de soja comportant 60 individus et qui a permis de mettre en évidence de nombreux QTL sur des caractères variés. Ces auteurs avaient alors montré que des populations de taille raisonnable pouvaient servir à la recherche de QTL selon les objectifs fixés. Par ailleurs Edwards *et al* (1992) ont effectué une recherche de QTL sur une sous population de 187 plantes F2 de maïs extraite d'une population complète de 1776 individus sur laquelle avait été effectuée auparavant une recherche de QTL à partir d'une carte génétique comportant 17 marqueurs enzymatiques. Leurs résultats ont montré que 17 associations sur un total de 20 significatives dans l'échantillon réduit l'avaient aussi été au sein de la population complète. Il y avait donc une concordance des résultats au niveau de la détection des marqueurs, par contre les effets génétiques pouvaient être mal estimés. Dans notre situation, la carte génétique qui a servi de support à la recherche de QTL est dense, aussi, mise à part la perte de puissance de détection due à l'effectif, les effets génétiques devraient être estimés sans biais important s'il n'y a pas d'épistasie.

Dans l'ensemble, les QTL détectés au sein de notre population expliquent une part importante de la variation des caractères ( $\geq 11\%$ ). Les valeurs de  $R^2$  sont souvent plus élevées à Divo, qu'à Bingerville mais les valeurs du LODscore sont plus importantes à Bingerville. La capacité de détection de QTL étant une fonction de l'effectif de la population, il est normal que seuls les QTL à effets forts aient été détectés. Chez le maïs, Beavis *et al.* (1991) ont mis en évidence à partir de deux populations de 112 et 144 individus, des régions chromosomiques responsables de 73 % de la variance. Pour des caractères à forte héritabilité, des QTL à effets faibles peuvent aussi être détectés à partir de populations d'effectifs raisonnables. Ainsi, il a été possible chez le maïs de mettre en évidence des QTL qui expliquaient 5%

de la variation de la résistance au Gray Leaf Spot au moyen de trois descendances F2-3 de 139, 144 et 193 individus ( Bubeck *et al.* 1993).

Pour certains caractères, il n'a pas été possible de détecter plus d'un QTL par caractère alors que généralement quatre à six régions chromosomiques sont détectés pour chaque caractère ( Paterson *et al.*, 1988; Beavis *et al.*, 1991; Edwards *et al.*, 1992; Bubeck *et al.*, 1993; Veldboom *et al.*, 1994). En effet lorsque l'on examine les histogrammes de distribution des QTL au sein du génome pour un caractère donné, de nombreux QTL à effets faibles et peu à effets forts interviennent sur le caractère (Edwards *et al.*, 1987; Edwards *et al.*, 1992). Nos résultats, surtout ceux obtenus par l'analyse séparée des données de Bingerville et de Divo, peuvent donc s'expliquer par la seule identification des QTL à effets forts étant donné les effectifs des populations. D'ailleurs, lorsque les effectifs ont été augmenté par l'analyse de l'ensemble des données de Divo et de Bingerville, le nombre de QTL détectés par caractère a aussi augmenté. Il faut rappeler qu'avec un risque de première espèce de 0,005, utilisé pour l'analyse de variance, il était possible de conclure à l'existence de QTL dans 0,8 cas sur les 168 pour un caractère donné bien que ceux-ci n'existeraient pas réellement.

Au niveau des deux environnements (Divo et Bingerville), différentes régions chromosomiques ont été mises en évidence pour les mêmes caractères. Aucun QTL commun entre Divo et Bingerville n'a pu être révélé lorsque l'on a analysé les données des deux lieux séparément. Ceci montre qu'en fait plusieurs QTL sont impliqués dans la variation d'un caractère, mais qu'il n'est pas toujours possible de les mettre tous en évidence dans tous les milieux, certains s'exprimant mieux dans certains environnements. Ce résultat pose le problème de l'interaction entre les QTL et le milieu, interaction qui a aussi été rapportée dans des études sur la tomate (Paterson *et al.*, 1991) et sur le maïs (Bubeck *et al.*, 1993).

Les effets des QTL détectés sont très variables. Pour la majorité des caractères, on s'attendait à obtenir des effets positifs dus au parent UF676 . Selon les populations utilisées, les effets étaient plus ou moins conformes à l'attente. Dans toutes les populations, on a mis en évidence des effets attendus ou non des QTL quelle que soit la méthode d'estimation. Des effets non attendus ont aussi été mis en évidence dans la population composite qui a un effectif assez important. D'autres études font état de situations dans lesquelles les phénotypes dans la parentaux attendus dans la descendances ne sont pas observés (Dirlewanger *et al.*, 1994; Lefèbvre, 1994) et ils les expliquent par la présence de plusieurs allèles de QTL. Plusieurs QTL interviennent sur un caractère quantitatif donné. Certains ont des effets positifs et d'autres ont des effets négatifs. Au sein d'un groupe de

liaison, plusieurs QTL peuvent exister mais pourraient ne pas être détectés par les méthodes actuellement utilisées.

Des effets génétiques importants des QTL ont été détectés au cours de notre analyse. Il s'agit d'additivité qui peut aider à prédire la valeur d'un génotype. Ces effets sont estimés en fonction de la fréquence de recombinaison entre le QTL et le marqueur et selon le type de croisement. Par l'*interval mapping*, la localisation des QTL a été estimée de manière assez précise. Certains QTL sont localisés à la position du marqueur: dans ce cas, la fréquence de recombinaison est nulle et selon les formules dérivées pour notre situation, la différence entre les deux classes de génotypes du marqueur sera:

$$Y(m_1m_1) - Y(m_1m_2) = 3/2(1-r)(1-2r)a = 3/2a$$

au lieu de  $a-d$  obtenu dans des conditions de panmixie (Falconer, 1974). Au fur et à mesure que  $r$  augmente l'additivité diminue.

Certaines régions chromosomiques sont impliquées dans la variation de plusieurs caractères qui sont par ailleurs fortement corrélés. Ils confortent nos résultats. C'est le cas des QTL affectant les caractères des cabosses détectés dans toutes les analyses. Certains caractères peu corrélés sont affectés par les mêmes régions chromosomiques et c'est le cas du QTL affectant la longueur de cabosse et de celui impliqué dans la variation du nombre d'ovules par ovaire à Bingerville. Il pourrait s'agir en fait d'effets pléiotropiques comme mis en évidence par Mansur *et al* (1993) chez le soja et par Veldboom *et al.* (1994) chez le maïs, ou d'effets de QTL multiples liés. Dans le cas d'effets pléiotropiques l'introgession rapide de caractères est possible, par contre dans le cas de liaisons entre plusieurs QTL, il peut s'avérer nécessaire de rompre ces liaisons s'ils ont des effets inverses.

Les deux caractères agronomiques les plus importants pour lesquels des QTL ont été mis en évidence sont la pourriture des cabosses dans les deux localités et les dégâts de mirides à Bingerville. Ce sont des caractères qui n'avaient pas une distribution normale et dont la détection par *interval mapping* pouvait en être affectée. Cependant, pour ce qui concerne la pourriture des cabosses, aussi bien l'*interval mapping* que l'analyse de variance ont pu détecter les mêmes régions chromosomiques au niveau de chaque localité. De plus lorsque l'on transforme le nombre de cabosses pourries en pourcentage qui tient compte de la production totale en cabosses de chaque arbre, les mêmes zones chromosomiques sont de nouveau mises en évidence. Les effets génétiques de ces QTL sont importants: à Bingerville, ils favorisent la pourriture des cabosses. A Divo, les deux QTL mis en évidence pour le pourcentage de cabosses pourries à l'aide de l'analyse de variance ont des effets opposés lorsqu'ils sont estimés par analyse de variance et par *interval mapping*. Etant donné que c'est le génotype du parent mâle UF676 qui a été cartographié, on peut conclure que ce clone contient des

gènes de la sensibilité et de la résistance à la pourriture qui s'expriment dans un environnement où la pression parasitaire est importante. Etant donnée l'influence de la non normalité des caractères sur l'estimation des effets génétiques par *interval mapping*, il est probable que l'estimation faite par analyse de variance soit plus proche de l'effet réel. Ainsi à Divo, la région chromosomique responsable de la pourriture provoque une réduction assez importante du nombre de cabosses pourries (-3.29). Il apparaît dès lors que pour le même caractère, des effets positifs et négatifs peuvent être détectés selon le milieu et que l'effet additif observé est la résultante de tous les effets.

Les dégâts dus aux mirides n'ont pas une distribution normale mais une région chromosomique a été mise en évidence par analyse de variance sur le groupe 2. Son effet est important et indique un accroissement de la sensibilité due au parent UF676 mais l'absence de répétition dans un autre milieu ne permet pas de tirer des conclusions sur l'effet du milieu.

L'utilisation de deux méthodes d'analyse a permis d'appréhender les limites de chacune d'elles. L'analyse de variance marqueur par marqueur s'est avérée efficace dans la détection de QTL pour les caractères dont la distribution n'est pas normale mais l'impossibilité d'une localisation précise des QTL ne permet pas de juger de la valeur de la fréquence de recombinaison. L'utilisation simultanée des deux méthodes permet une analyse critique des résultats.

## IV- CONCLUSION

En matière de cartographie génétique, le choix des marqueurs RAPD au détriment d'autres marqueurs est souvent dû à l'absence de polymorphisme entre les deux parents de la population de cartographie car les RAPD révèlent en général plus de polymorphisme que les RFLP et les isozymes ainsi qu'un manque d'équipements nécessaires aux analyses moléculaires. Cependant, le caractère dominant des marqueurs RAPD les rend peu propices à des cartographies génétiques de certaines populations comme les F2. De plus, certains auteurs ont montré la non universalité des cartes établies au moyen des RAPD car les groupes de liaisons peuvent varier d'un croisement à un autre (Fauré, 1993). Aussi les cartes RAPD sont souvent de portée limitée et ne peuvent être valorisées que dans des cas où une descendance est particulièrement intéressante pour certains caractères et sur laquelle on peut rechercher des QTLs. Par contre, les RAPD peuvent servir à saturer des cartes génétiques établies au moyen des RFLP. C'est d'ailleurs cette situation qui est de plus en plus rencontrée (Haley *et al.*, 1993).

De nombreux caractères intéressent le sélectionneur du cacaoyer: ce sont des caractères liés à la qualité du cacao marchand, ceux liés au rendement et ceux

en rapport avec la résistance aux aléas biotiques et abiotiques. Certains, tels que la grosseur des fèves ont une forte héritabilité: dans ce cas, la sélection phénotypique est efficace mais le cycle de sélection est long. D'autres caractères tels que le rendement sont complexes et sont très influencés par le milieu. Pour ces caractères des schémas de sélection plus complexes et plus longs sont utilisés. L'apport des marqueurs peut être appréciable à différentes étapes du programme, en particulier celui de la cartographie des QTL qui permet de «décortiquer» des caractères complexes en leurs composantes plus simples pour connaître leur importance relative dans l'expression du caractère. Quelle que soit l'héritabilité du caractère, la possibilité d'accès à l'information génétique par l'intermédiaire de marqueurs liés aux QTL peut aider dans le choix des stratégies de sélection.

Dans cette analyse exploratoire et bien que de faibles effectifs aient été utilisés, il a été possible de détecter des QTL pour la plupart des caractères analysés et en particulier pour des caractères importants tels que la taille des fèves, la pourriture des cabosses et les dégâts de mirides. Dès qu'une confirmation de ces résultats aura été obtenue, il sera possible d'envisager d'introduire l'utilisation de marqueur dans la sélection de ces caractères au cours des différentes étapes de la sélection récurrente réciproque en cours en Côte d'Ivoire (Paulin et al, 1993). La pourriture des cabosses est essentiellement due au *Phytophthora*, agent de la pourriture brune des cabosses. Cependant plusieurs facteurs biologiques et du milieu peuvent intervenir dans la résistance. La résistance à la pourriture brune des cabosses ainsi que celle aux mirides ont fait l'objet de plusieurs études notamment au Cameroun et en Côte d'Ivoire. Des clones résistants au *Phytophthora* ont été identifiés grâce à des tests réalisés sur cabosses et sur feuilles en laboratoire. Cependant, la corrélation entre la résistance mise en évidence au laboratoire, encore appelée résistance intrinsèque et la résistance au champ n'est pas toujours facile à établir. Si ces résultats se confirmaient, l'association entre les marqueurs cTcCIR79, cTcCIR77, cTcCIR46 et le locus QTL intervenant dans la résistance au *Phytophthora* offrirait de larges possibilités d'utilisation pratique: tri précoce, aide au choix de géniteurs, vérification des recombinaisons, possibilité de sélection assistée par marqueurs (Gallais, 1993).

La zone chromosomique de la résistance détectée dans cette étude étant très proche du marqueur, il serait possible de faire un tri précoce efficace, dès qu'il y aurait suffisamment de feuilles pour réaliser des analyses RFLP permettant d'optimiser la sélection pour ce caractère. Ces analyses compléteraient les tests sur feuilles et sur cabosses réalisés en laboratoire. Par ailleurs, il serait possible de faire synthétiser des amorces spécifiques de la région et de réaliser des amplifications par PCR pour vérifier la présence ou l'absence de l'allèle conférant la résistance au *Phytophthora*.

Introduit dans le programme de sélection réciproque à différents niveaux, le tri précoce permettra de sélectionner les géniteurs résistants pour constituer les pools de populations à améliorer pour d'autres caractères. Si la diffusion de matériel clonal est envisagée, les associations des QTL avec les marqueurs pourront servir à sélectionner au sein d'une descendance, les individus répondant à certaines normes.

Pour rendre l'utilisation de marqueurs moléculaires plus accessible au sélectionneur, il sera utile de transformer les marqueurs RFLP en marqueurs PCR, utilisable en routine. Les perspectives de ces recherches sont nombreuses surtout pour une plante dont le cycle de génération est long. Ces recherches nécessitent d'être poursuivies afin de cartographier un grand nombre de QTL et faciliter ainsi la tâche du sélectionneur.

# CONCLUSION GENERALE

## RAPPEL DES PRINCIPAUX RESULTATS

Un des objectifs de ce travail était d'évaluer la diversité génétique des cacaoyers par les marqueurs RAPD tout en comparant les résultats avec ceux obtenus par d'autres méthodes d'évaluation de la diversité mettant en jeu caractères morphologiques, ou marqueurs RFLP. L'intérêt était de pouvoir ainsi tester les potentialités de ces marqueurs RAPD pour évaluer et gérer les ressources génétiques du cacaoyers.

Après une étape préliminaire de mise au point de la technique des RAPD, mises au point importantes pour obtenir une stabilité des fragments amplifiés dans nos conditions expérimentales, nous avons étudié la diversité génétique de 106 clones de diverses origines et nous avons comparé les résultats à ceux obtenus en utilisant les caractères morphologiques et les marqueurs RFLP. En accord avec les conclusions de ces deux types d'étude, l'existence de deux grands groupes de cacaoyers que sont les Forastero et les Criollo a de nouveau été mise en évidence. Ces groupes ont été mis en évidence par l'analyse de tous les fragments amplifiés polymorphes mais aussi par un nombre plus réduit de marqueurs correspondants aux séquences répétées identifiées par hybridation. Ces groupes bien différenciés et variables correspondent probablement à une différenciation ancienne et l'interfertilité entre ces deux groupes a donné naissance à une grande diversité de formes chez le cacaoyer. Des marqueurs RAPD spécifiques de ces groupes ont été mis en évidence. Les génotypes classiquement classés comme Criollo se sont regroupés ensemble ainsi que ceux classés usuellement comme Forastero. Une proportion élevée de clones bien classés par rapport à la classification morphogéographique a donc été obtenue par l'utilisation des marqueurs RAPD. Par ailleurs, contrairement aux autres marqueurs, les RAPD ont permis de révéler une différenciation assez marquée du groupe des Forastero en deux sous populations distinctes qu'il convient d'appeler Forastero Bas Amazoniens et Forastero Haut Amazoniens car constitués essentiellement par des génotypes de ces groupes. Au sein du groupe des Forastero Haut Amazoniens sont retrouvés de nombreux génotypes en provenance de l'Equateur dont les génotypes SCAVINA qui apparaissent particuliers selon les analyses RFLP et les génotypes NACIONAL. De même, les génotypes GUYANAIS, qui se distinguent par leur morphologie et par les marqueurs RFLP du reste des Bas Amazoniens, et habituellement classés comme Bas Amazoniens se sont classés parmi les Forastero Bas Amazoniens; ils se

retrouvent aussi très proches de quelques génotypes en provenance du Brésil. Les Trinitario sont les plus variables et leur variabilité reflète leur origine hybride. Ils s'associent soit avec les Forastero Bas Amazoniens (CC39, TJ1, MT1 etc.), soit avec les Criollo (CNS23 ACU85, UF221, ICS89). Le clone PA35 qui en Côte d'Ivoire était classé comme Forastero Haut Amazoniens se retrouve parmi les Criollo. Les Trinitario au cours de ces analyses n'ont pas été différenciés en tant que groupe aussi nous proposons qu'ils soient intégrés au Criollo pour la majorité d'entre eux et que trois groupes de cacaoyers soient dorénavant considérés: les Forastero Haut Amazoniens, les Forastero Bas Amazoniens et les Criollo.

L'étude des paramètres génétiques de la diversité révélée par les RFLP a montré qu'il y avait peu d'allèles par locus chez le cacaoyer. Par contre une importante diversité ainsi qu'un polymorphisme élevé sont rencontrés. Une forte structuration des populations a été mise en évidence et peut s'expliquer par la présence probable de sous populations et par l'allogamie au sein de certains groupes. Au niveau des Forastero Haut Amazoniens existe une importante diversité et un polymorphisme élevé. Un pourcentage non négligeable d'autofécondation a lieu au sein de ce groupe qui est préférentiellement allogame. Cependant peu de locus sont fixés. Au contraire, au sein des Forastero Bas Amazoniens autogames, un nombre important de locus sont fixés, la diversité génique y est moins élevée ainsi que le taux de polymorphisme détecté. Une certaine consanguinité y est présente et un pourcentage élevé de locus sont à l'état homozygotes. Les Forastero Haut Amazoniens sont bien différenciés des Forastero Bas Amazoniens au niveau des fréquences géniques. Les Criollo et les Trinitario sont très proches et une diversité génique élevée ainsi qu'un polymorphisme important y sont rencontrés. Le niveau d'hétérozygotie observé est plus élevé chez les Trinitario que chez les autres groupes. Ainsi pour toutes ses caractéristiques moléculaires, la population de cacaoyers de la Haute Amazonie apparaît comme le meilleur représentant de l'espèce et la Haute Amazonie est probablement le centre de diversité de l'espèce mais rien ne permet d'affirmer que les Criollo proviennent des Forastero sur la bases des analyses moléculaires.

Au niveau des génotypes individuels une grande variabilité existe quant au nombre de locus hétérozygotes dans chaque groupe. Ainsi, on trouve des clones à l'état homozygote pour un grand nombre de locus dans tous les groupes. Au sein des Forastero Haut Amazoniens étudiés, P2 est homozygote pour 27 locus sur les 27 analysés, LCT-EEN 127 l'est pour 25 sur 26 locus. Au sein des Forastero Bas Amazoniens les génotypes les clones Guyanais possèdent de nombreux locus à l'état homozygote de même que les clones Bas Amazoniens du

Vénézuela. Au sein des Criollo le génotypes BOC et PV6 sont homozygotes pour 26 et 27 locus sur les 27 analysés.

Les analyses de cartographie réalisées n'ont pas abouti à la construction d'une carte génétique saturée en marqueurs RAPD sur la descendance UPA409 x POR. Néanmoins les marqueurs RAPD développés ont permis de compléter la carte génétique réalisées sur la descendance UPA 402 x UF 676 en utilisant plusieurs types de marqueurs.

La localisation de locus intervenant dans la variation de certains caractères quantitatifs étaient un autre objectif de nos études. L'analyse de la descendance UPA 402 x UF 676 a permis de localiser des QTL pour certains caractères agronomiques importants comme la résistance à la pourriture brune des cabosses et la résistance aux mirides. La pourriture brune des cabosses est l'une des maladies principales de la cacao-culture en Afrique et notamment au Cameroun qui limite la capacité de production de ce pays. L'amélioration génétique par la sélection de génotypes tolérants à la pourriture se heurte à la difficulté de corrélérer la tolérance à la pourriture brune estimée par des tests en laboratoire sur cabosses ou sur feuilles et celles observées en champ ainsi qu'au déterminisme polygénique de ce caractère. Aucune méthode fiable de sélection précoce de géniteurs basée sur des tests de tolérance à la pourriture brune n'est encore disponible à ce jour. Nous avons détecté des facteurs de la sensibilité à la pourriture chez le clone UF676 localisés sur les groupes de liaison 1, 8 et 10. De même, un QTL de la sensibilité aux mirides a été localisé sur le groupe 2 et près du marqueur gTcCIR135. Des QTL intervenant dans l'expression des caractéristiques des fèves et des cabosses ont aussi été mis en évidence sur les groupes 2, 3, 5, 8 et 9.

## PERSPECTIVES

Une des raisons de l'utilisation d'un nombre limité de génotypes dans les programmes d'amélioration du cacao dans le monde malgré la richesse génétique de cette espèce est l'insuffisance de connaissances sur les clones maintenus en collection. En effet, les données morphologiques sont longues à acquérir et les marqueurs enzymatiques à cause de leur nombre limité, sont peu utilisés. Le développement des marqueurs RFLP est récent et la technologie n'est pas facilement utilisable en routine. A côté des marqueurs enzymatiques, les RAPD apparaissent comme pouvant aider dans l'évaluation des collections. Les RAPD révèlent de la variabilité intra et intergroupe et nous semblent adaptés à l'étude de la diversité génétique au sein des différents groupes de cacaoyers. Ceci est possible grâce à l'existence de bandes spécifiques des différents groupes. Dans

notre analyse de la diversité, 49 marqueurs RAPD ont été utilisés mais il est possible de réduire le nombre de marqueurs et obtenir les mêmes informations lorsque le nombre de groupes à analyser est réduit. Le polymorphisme révélé par les RAPD est important et devrait permettre l'évaluation de la diversité génétique des génotypes même apparentés. La simplicité de la technique PCR rend possible une utilisation en routine des marqueurs RAPD en prenant beaucoup de précautions. Le sélectionneur pourra alors vérifier les appartenances des génotypes ou reclasser des génotypes qu'il utilise. L'étape suivante à aborder après les diverses études globales de diversité génétique est l'étude plus fine des diverses populations de cacaoyers. En effet, une présomption de l'existence de sous populations au sein des groupes existe et reste à confirmer. La révélation et la caractérisation de sous populations particulières permettront de mettre en évidence des caractéristiques intéressantes pour le sélectionneur. L'étude de la diversité de toute l'espèce introduit un biais par le choix des génotypes et masque les particularités de certaines populations .

Une voie à explorer est l'étude de la distance génétique estimée par les marqueurs moléculaires et le phénomène d'hétérosis observé dans certains croisements. Ces études ont été réalisées chez certaines espèces et ont permis d'établir des corrélations fiables qui peuvent être utilisées dans le choix des géniteurs. Grâce aux études moléculaires, il est dorénavant possible d'accéder aux génotypes de nombreux clones. Il devient alors possible d'établir une relation entre le degré d'hétérozygotie des parents et l'hétérogénéité des descendances obtenues. Si une telle relation existait, il serait alors possible d'utiliser les géniteurs quasi homozygotes mis en évidence dans tous les groupes pour la création de descendances homogènes proches des hybrides F1, souhaitables surtout pour l'homogénéité des caractères de qualité du produit marchand.

Au niveau de la sélection du cacaoyer, la tendance actuelle est à l'utilisation des méthodes de sélection récurrente réciproque qui permettent aussi bien une amélioration intrapopulation qu'interpopulation tout en permettant des sorties variétales à différentes étapes du cycle. Ces schémas de sélection partent de populations sources possédant certaines caractéristiques généralement complémentaires. En Côte d'Ivoire où un programme de sélection est en cours depuis quelques années, deux populations de base appelées A0 et B0 regroupant d'une part les Bas Amazoniens et Trinitario et d'autre part les Haut Amazoniens et les Criollo ont été constituées. D'après les résultats de notre étude de diversité, les groupes les plus éloignés sont les Forastero Haut Amazoniens et les Criollo, et les hybrides entre ces deux groupes devraient exprimer un maximum d'hétérosis. En pratique, cela reviendrait à recréer des Trinitario, par un choix judicieux des parents. Les divers groupes présentent de la variabilité

Aussi, nous proposons la création de trois populations de base constituées de Forastero Haut Amazoniens, de Forastero Bas Amazoniens et de Criollo. Les Trinitario seraient intégrés aux Criollo. On rechercherait au sein des Forastero Haut Amazoniens les qualités de précocité, de vigueur, de productivité et de tolérance aux aléas. Chez les Bas Amazoniens, ce seraient l'adaptabilité et la tolérance aux aléas qui seraient recherchées. Quant aux Trinitario et Criollo, ils apporteraient les qualités de fèves. Au sein d'un groupe donné, il sera nécessaire d'étudier par les marqueurs RAPD, un nombre important de génotypes afin de déterminer l'importance de la variabilité et révéler des génotypes intéressants. Au sein des Forastero Bas Amazoniens, une grande variabilité sera apportée par les clones de Guyane et du Vénézuéla ainsi que les génotypes sauvages brésiliens. Dans un premier temps la sélection intragroupe permettra de disposer des génotypes présentant les meilleures caractéristiques de chaque groupe. Ainsi pourront être sélectionnés des Criollo proches des Trinitario et présentant une bonne vigueur ainsi que de la tolérance aux aléas. La réalisation de croisements entre clones ainsi sélectionnés et leur évaluation aboutira à la sélection d'hybrides présentant une bonne productivité associée à une bonne tolérance aux aléas ainsi qu'à de bonnes qualités de fèves. Les associations des locus QTL avec les marqueurs moléculaires pourraient être utilisées pour optimiser les différentes étapes de ce schéma. Par ailleurs, l'on pourra suivre l'évolution de la variabilité au cours des cycles de sélection.

Les QTL mis en évidence dans nos études l'ont été en utilisant des descendances d'effectifs faibles. Étant donné l'intérêt de la descendance UPA402 x UF676, il faudrait poursuivre ces études en analysant cette population pour des caractères plus ciblés dans une perspective d'amélioration génétique par l'utilisation de nombreux individus. Ceci est possible car cette descendance se trouve dans plusieurs essais de descendances en Côte d'Ivoire et se trouve aussi dans d'autres pays. On pourra par ailleurs étudier les interactions QTL/ environnement et évaluer de manière plus juste s'il convient d'entreprendre une sélection assistée par marqueurs sur le cacaoyer pour tous les caractères.

## REFERENCES BIBLIOGRAPHIQUES

Adams, R.P., T. Demeke, H.A. Abulfatih. 1993. RAPD DNA fingerprints and terpenoids: clues to past migrations of *Juniperus* in Arabia and east Africa. *Theor. Appl. Genet.* 87:22-26.

Addad. 1985. Manuel de référence. Association pour le Développement et la Diffusion de l'Analyse des données. Paris.

Allard, R.W. 1956. Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* 24: 235-278.

Allen, J.B. 1987. London cocoa trade Amazon project. Final report phase 2. *Cocoa growers' bulletin* 39:94p.

Allen, J.B., R.A. Lass. 1983. London cocoa trade amazon project. Final report phase 1. *Cocoa Growers' Bulletin* 34:71p.

Asins, M.J., E.A. Carbonell. 1988. Detection of linkage between restriction fragment length polymorphism markers and quantitative traits. *Theor Appl Genet* 76: 623-626.

Babacàuh, K.D., M. Partiot. 1976. Le *phytophthora* sp., parasite du cacaoyer en Côte d'Ivoire. Première étude de sa variabilité morphologique, physiologique et pathogénique. *Café, Cacao, Thé* 20:117-128.

Bacilieri, R., T. Labbé, A. Kremer. 1994. Intraspecific genetic structure in a mixed population of *Quercus petraea*(Matt.) Lieb and *Q. robur* L. *Heredity*, in press.

Beavis, W.D., D. Grant, M. Albertsen, R. Fincher. 1991. Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. *Theor. Appl. Genet.* 83:141-145.

Benzecri, J.P. 1973. L'analyse des données. Dunod. Paris.

Besse, J. 1975. La sélection générative du cacaoyer en Côte d'Ivoire, bilan et orientation des recherches en 1975. In: 5<sup>ème</sup> confé. Intern. sur la Recherche cacaoyère, Ibadan(Nigéria). 95-103.

Besse, P., P. Lebrun, M. Seguin, C. Lanaud. 1993. DNA fingerprinting in *Hevea brasiliensis* (rubber tree) using human minisatellite probes. *Heredity* 70:237-244.

Besse, P., M. Seguin, P. Lebrun, M.H. Chevallier, D. Nicolas, C. Lanaud. 1994. Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis. *Theor. Appl. Genet* 88:199-207.

Blaha, G., R. Lotodé. 1976. Un critère primordiale de sélection du cacaoyer au Caperoun: La résistance à la pourriture brune des cabosses (*Phytophthora palmivora*). *Café, Cacao, Thé* 20:97-116.

Braudeau, J. 1969. Le cacaoyer. GP Maisonneuve et Larose, Paris. 304pp

Bubeck, D.M., M.M. Goodmam, W.M. Beavis, D. Grant. 1993. Quantitative Trait Loci controlling resistance to Gray Leaf Spot in maize. *Crop Sci.* 33:838-847.

Carletto, A. 1946. O numero de cromosômios em cacaueiros. *Boletim tecnico, Instituto de cacau da Bahia* 6:33-43.

Carlson, J.E., L.K. Tulsieram, J.C. Glaugitz, V.W.K. Luk, C. Kauffeldt, R. Rutledge. 1991. Segregation of random amplified DNA markers in F1 progeny of conifers. *Theor. Appl. Genet.* 83:194-200.

Castiglione, S., G. Wang, G. Damiani, C. Bandi, S. Bissofi, F. Sala. 1993. RAPD fingerprints for identification and taxonomic studies of elite poplars (*Populus* spp.) clones. *Theor. Appl. Genet.* 87:54-59.

Chalmers, W.S. 1968' - 1973. Cacao germplasm collecting in the Oriente region of Ecuador. Annual Reports on Cacao Research. Trinidad.

Chaparro, J.X., D.J. Werner, D. O'Malley, R.R. Sederoff. 1994. Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. *Theor Appl genet* 87:805-815.

Cheesman, E. 1944. Notes on the nomenclature, classification and possible relationships of cocoa population. *Tropical agriculture* 21:144-159.

Cilas, C. 1992. Appui en biometrie aux programmes café et cacao de l'IRA. Amélioration du cacaoyer.

Clapperton, J., S. Yow, J. Chan, D.H.K. Lim, G. Lockwood, L. Romanczyk, J. Hammerstone. 1991. Effect of cocoa genotype on flavour. In: *International Cocoa Conference: Challenges in the 90s*, Kuala Lumpur, Malaysia. *In Press*.

Clément, D., A. Eskes, O. Sounigo, J. N'goran. 1993. Amélioration génétique du cacaoyer en Côte d'Ivoire: présentation d'un nouveau schéma de sélection. In: *XI conférence Internationale sur la recherche cacaoyère*, Yamoussoukro, Côte d'Ivoire. *In Press*.

Clément, D., A.B. Eskes, B. Guyot, J. Perriot, P. Bastide. 1991. Organoleptic quality assessment of small samples from different cocoa genotypes fermented under similar conditions in Ivory Coast. In: *International Cocoa Conference: Challenges in the 90s*, Kuala Lumpur, Malaysia. *In Press*.

Cope, F.W. 1962. The mechanism of pollen incompatibility in *Theobroma cacao*. *Heredity* 17:157.

Cuatrecasas, J. 1964. Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Bulletin of the United States National Museum, Smithsonian Institution (Washington)* 35:379-614.

Dagnelie, P. 1975. Théorie et méthodes statistiques vol.2. Les presses agronomiques de Gembloux, Gembloux.

Dand, R. 1993. The international cocoa trade. *Woodhead Publishing Limited*

Darvasi, A., A. Weinreb, V. Minke, J.I. Weller, M. Soller. 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134:943-951.

De Reffye, P., J.P. Parvais, G. Mossu, P. Lucas. 1978. Influence des aléas de la pollinisation sur les rendements du cacaoyer. Modèle mathématique et simulation. *Café, Cacao, Thé* 22:251-274.

De Vienne, D. 1990. L'analyse du déterminisme génétique des caractères quantitatifs chez les végétaux. *SFG* 6:11-15.

Dellaporta, S.L., Wood, J., J. B. Hicks. 1983. A plant DNA miniprep: version II. *Plant molecular Biology Reporter* 1: 19-21.

Delseny, M., G. F., D. Tremousaygue, R. M., F. Panabières. 1988. Structure, evolution et expression de l'ADN nucléaire. *Bull. Soc. Bot. Fr. Actual.bot* 135:23-38.

Dirlewanger, E., P.G. Isaac, S. Ranade, M. Belajouza, R. Cousin, D. De Vienne. 1994. Restriction fragment length polymorphism analysis of loci associated with disease resistance genes and developmental traits in *Pisum sativum* L. *Theor. Appl. Genet.* 88:17-27.

Dos Santos, J.B., J. Nienhuis, P. Skroch, J. Tivang, M.K. Slocum. 1994. Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleracea* L. genotypes. *Theor Appl genet* 87:909-915.

Echt, C.C., L.A. Erdahl, T.J. McCoy. 1992. Segregation of random amplified polymorphism DNA in diploid cultivated alfalfa. *Genome* 35:84-87.

Echt, C.S., K.K. Kidwell, S.J. Knapp, T.C. Osborn, T.J. McCoy. 1994. Linkage mapping in diploid alfalfa (*Medicago sativa*). *Genome* 37 (1): 61-71.

Edwards, M.D., T. Helentjaris, S. Wright, C.W. Stuber. 1992. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. 4. Analysis based on genome saturation with isozyme and restriction fragment length polymorphism markers. *Theor. Appl. Genet.* 83:765-774.

Edwards, M.D., C.W. Stuber, J.F. Wendel. 1987. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* 116:113-125.

Engels, J.M.M. 1983. A systematic description of cacao clones. I. The discriminative value of Quantitative characteristics. *Euphytica* 32:377-385.

Engels, J.M.M. 1985. A systematic description of cacao clones. V-Quantitative genetic aspect of several fruit characters. *Café, Cacao, Thé* 29:3-9.

Engels, J.M.M. 1986. The systematic description of cacao clones and its significance for taxonomy and plant breeding. *PhD thesis*. Wageningen.

Engels, J.M.M., B.G.D. Bartley, G.A. Enriquez. 1980. Cacao descriptors, their states and modus operandi. *Turrialba* 30:209-218.

Enriquez, G., J. Soria. 1968. The variability of certain bean characteristics of cacao (*Theobroma cacao* L.). *Euphytica* 17:114-120.

Enriquez, G.A. 1992. Characteristics of cacao "Nacional" of Ecuador. In: *International workshop on conservation, characterisation and utilisation of cocoa genetics resources in the 21st century*, CRU, Trinidad.

Falconer, D.S. 1974. Introduction à la génétique quantitative. Masson et Cie. Paris. 284pp.

Falque, M. 1993. Irradiation gamma de pollen de *Theobroma cacao* L.: effet sur la physiologie du pollen et la tenue des fruits après différents types de pollinisation. In: *XI Conférence Internationale sur la recherche cacaoyère*, Yamoussoukro, Côte d'Ivoire. In Press.

- Fauré, S. 1993. Contribution à la cartographie génétique du génome des bananiers diploïdes (*Musa acuminata*) à l'aide de marqueurs moléculaires. *Doctorat*. ENSA, Rennes.
- Fauré, S., J.L. Noyer, J.P. Horry, F. Bakry, D. Gonzalez de Leon. 1993. A molecular marker-based linkage map of diploid bananas (*Musa acuminata*). *Theor Appl Genet* 87:517-526.
- Fenelon, J.P. 1981. Qu'est-ce que l'analyse des données. Lefonen, Paris. 310pp.
- Figueira, A., J. Janick, P. Goldsbrough. 1992. Genome size and DNA polymorphism in *Theobroma cacao*. *J. Am. Soc. Hort. Sci* 117:673-677.
- Figueira, A.J., M. Janick, Levy, P. Goldsbrough. 1993. Reexamining the classification of *Theobroma cacao* L. using molecular markers. In: *Intern. XI Cocoa Res. conference*, Yamoussoukro, Côte d'Ivoire. *In Press*.
- Flavell, R.B., M.D. Bennett, J.B. Smith. 1974. Genome size and proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* 12:257-269.
- Francisco-Ortega, J., H.J. Newbury, L. Ford B.V. 1993. Numerical analyses of RAPD data highlight the origin of cultivated tagasaste (*chamaecytisus proliferus* ssp. *palmensis*) in the canary islands. *Theor Appl Genet* 87 (1-2): 264-270.
- Fukuoka, S., K. Hosak, O. Kamijima. 1992. Use of random amplified polymorphic DNAs (RAPDs) for identification of rice accessions. *Jpn. J. Genet* 67:243-252.
- Gallais, A. 1993. La sélection assistée par marqueurs. *Le sélectionneur français* 43:43-62.
- Gebhardt, C., E. Ritter, I. Debener, U. Schachtschabel, B. Walkemeier, H. Uhrig, F. Salamini. 1989. RFLP analysis and linkage mapping in *Solanum tuberosum*. *Theor. Appl. Genet.* 78:65-75.
- Giovannoni, J.J., R.A. Wing, M.W. Ganai, S.D. Tanksley. 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nuc. Ac. Res.* 19:6553-6558.
- Glaszmann, J.C. 1987. Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* 74:21-30.
- Glendinning, D. 1963. The inheritance of bean size, pod size and number of bean per pod in cocoa (*Theobroma cacao* L.) with a note on bean shape. *Euphytica* 12:311-322.
- Glicenstein, L.J., P.J. Fritz. 1989. Meiosis in *Theobroma cacao*. *Turrialba* 39:497-500.
- Grivet, L., A. D'hont, P. Duffour, P. Hamon, D. Roques, J.C. Glaszmann. 1994. Comparative mapping of sugarcane with other species within Andropogoneae tribe. *Heredity in press*.
- Haldane, J.B.S. 1919. The recombination of linkage values and the calculation of distance between the loci of linked factors. *J. Genet* 8:299-309.
- Haley, C.S., S.A. Knott, J.M. Elsen. 1994. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 136:1195-1207.

Haley, S.D., P.N. Miklas, J.R. Stavely, J. Byrum, J.D. Kelly. 1993. Identification of RAPD markers linked to a major rust resistance gene block in common bean. *Theor. Appl. Genet.* 86:505-512.

Harland, S.C., G.E. Frechville. 1927. Natural crossing and the genetics of axil spot in cacao. *Genetica* 9:279-289.

Havey, M.J., F.J. Muehlbauer. 1989. Linkages between restriction fragment length, isozyme, and morphological markers in lentil. *Theor Appl Genet* 77:395-401.

Hayashi, T., Y. Ukai. 1994. Detection of additive and dominance effects of QTLs in interval mapping of F2 RFLP data. *Theor. Appl. Genet* 87:1021-1027.

Hemmat, M., N.F. Weeden, A.G. Manganaris, D.M. Lawson. 1994. Molecular marker linkage map for apple. *J. Hered.* 85 (1): 4-11.

Heun, M., T. Helentjaris. 1993. Inheritance of RAPDs in F1 hybrids of corn. *Theor. Appl. Genet.* 85:961-968.

Heun, M., J.P. Murphy, T.D. Phillips. 1994. A comparison of RAPD and isozyme analyses for determining the genetic relationships among *Avena sterilis* L. accessions. *Theor. Appl. Genet.* 87:689-696.

Huff, D.G., R. Peakall, P.E. Smouse. 1993. RAPD variation within and among natural populations of outcrossing buffalograss (*Buchloë dactiloides* (Nutt.) Engelm.). *Theor. Appl. Genet.* 86:927-934.

Hunt, G.J., R.E.J. Page. 1992. Patterns of inheritance with RAPD molecular markers reveal novel types of polymorphism in the Honey bee. *Theor. Appl. Genet.* 85:15-20.

Innis, M.A., K. Myambo, D.H. Gelfand, M.A.D. Brow. 1988. DNA sequencing with *Thermus aquaticus* Dna polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl Aca. Sci. USA* 85:9436-9440.

IRCC, Côte d'Ivoire. 1958. Rapport Annuel.

IRCC, Côte d'Ivoire. 1961. Rapport Annuel.

Jacob, V.J., O.A. Atanda. 1973. Pod value studies of Amelonado and Amazon cacao. *Turrialba* 23:347-351.

Jambu, M. 1989. Exploration informatique et statistique des données. , (eds. Dunod),

Jansen, R.C. 1992. A general mixture model for mapping quantitative trait loci using molecular markers. *Theor. Appl. Genet* 85:252-260.

Jelinski, D.E., W.M. Cheliak. 1992. Genetic diversity and spatial subdivision of *populus tremuloides* (Salicaceae) in a heterogeneous landscape. *Amer. Journ. Botany* 79:728-736.

Joshi, A.K., V. Baichwal, G.F.-L. Ames. 1991. Rapid polymerase chain reaction using intact bacterial cells. 10:42-44.

Jouve, P., H. De Milly, P. Chalmin, J. Deuss, F. Jarrige, F. Ruf. 1990. Compétitivité du cacao africain. Analyse du marché mondial et des principaux producteurs. Ministère de la coopération et du développement. sept 1990.

- Kazan, K., J.M. Manners, D.F. Cameron. 1993. Genetic variation in agronomically important species of *Stylosantes* determined using random amplified polymorphic DNA markers. *Theor. Appl. Genet.* 85:882-888.
- Keim, P., B.W. Diers, T.C. Olson, R.C. Shoemaker. 1990. RFLP mapping in soybean : association between marker loci and variation in quantitative traits. *Genetics* 126:735-742.
- Kennedy, A.J., G. Lockwood, G. Mossu, N.W. Simmonds, G.Y. Tan. 1987. Cocoa breeding: past, present and future. *Cocoa Growers' bulletin* 38:5-23.
- Kesseli, R.V., I. Paran, R.W. Michelmore. 1991. Efficient mapping of specifically targeted genomic regions and tagging of these regions with reliable PCR-based genetic markers. Application of RAPD technology to plant breeding 31-36.
- King, I.P., K.A. Purdie, H.N. Rezanoor, R.M.D. Koebner, T.E. Miller, S.M. Reader, P. Nicholson. 1993. Characterisation of *Thinopyrum bessarabicum* chromosome segments in wheat using random amplified polymorphic DNAs (RAPDs) and genomic in situ hybridization. *Theor. Appl. genet* 86:895-900.
- Klein-Lankhorst, R.M., A. Vermunt, R. Weide, T. Liharska, P. Zabel. 1991. Isolation of molecular markers for tomato (*L. esculentum*) using random amplified polymorphic DNA (RAPD). *Theor. Appl. Genet.* 83:108-114.
- Knight, R., H.H. Rogers. 1955. Incompatibility in *Theobroma cacao*. *Heredity* 9:69-77.
- Kosambi, D.D. 1944. The estimation of map distance from recombination values. *Ann. Eug.* 12:172-175.
- Lachenaud, P. 1991. Facteurs de la fructification chez le cacaoyer (*Theobroma cacao* L.). Influence sur le nombre de graines par fruit. *Doctorat de l'Institut National Agronomique , Paris-Grignon.*
- Lachenaud, P. 1992. Recent cocoa germplasm conservation initiatives of IRCC in French Guiana. *Internation. Workshop on Conservation, Characterisation and Utilisation of cocoa genetic resources in the 21st century* Trinidad:253-255.
- o Lanaud, C. 1987. Nouvelles données sur la biologie du cacaoyer (*Theobroma cacao* L.): diversité des populations, système d'incompatibilité, haploïdes spontanés. Leurs conséquences pour l'amélioration génétique de cette espèce. *Doctorat d'état.* Paris XI.
- Lanaud, C., P. Hamon, C. Duperray. 1992. Estimation of nuclear DNA content of *Theobroma cacao* L. by flow cytometry. *Café Cacao Thé* 36:3-8.
- Lanaud, C., F. Leal, N. Rodriguez. 1986. Rapport de prospection au Venezuela. IBPGR.
- Lanaud, C., A.M. Risterucci, J.A.K. N'goran, V. Laurent, O. Sounigo, D. Clément. 1993. Mapping of the cocoa genome using biochemical and molecular markers. In: *XII Conférence internationale sur la recherche cacaoyère*, Yamoussoukro, Côte d'Ivoire. *In Press.*
- Lanaud, C., O. Sounigo, Y.K. Amefia, D. Paulin, P. Lachenaud, D. Clément. 1987. Nouvelles données sur le fonctionnement du système d'incompatibilité du cacaoyer et ses conséquences pour la sélection. *Café, Cacao, Thé (Paris)* 31:267-277.

Lander, E.S., D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185-199.

Lander, E.S., P. Green. 1987. Construction of Multilocus Genetic Linkage Maps in Humans. *Proc Natl Acad Sci* 84: 2363-2367.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, L. Newburg. 1987. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181.

Lapitan, N.L.V. 1992. Organization and evolution of higher plant nuclear genomes. *Genome* 35: 171-181.

Laurent, V., A.M. Risterrucci, C. Lanaud. 1994. Genetic diversity in cocoa revealed by cDNA probes. *Theor. Appl. Genet.* 88:193-198.

Laurent, V., A.M. Risterucci, C. Lanaud. 1993. Variability for nuclear ribosomal genes within *Theobroma cacao*. *Heredity* 71:96-103.

Le Pennec, J.P. 1991. Amplification enzymatique in vitro de l'ADN. . *Technique PCR* 430:23-26.

Léfèbvre, V. 1994. Marquage moléculaire et analyse de résistances polygéniques: interaction Piment(*Capsicum annum*L.)-*Phytophthora capsici* Leon. *Doctorat*. Université de Paris Sud.

Lerceteau, E. 1992. Evaluation de la diversité génétique de *Theobroma cacao* par les techniques de détection d'ADN polymorphe amplifié au hasard (RAPD) et de polymorphisme de longueur des fragments de restriction de l'ADN (RFLP). *Rapport de stage du DEA de Biologie moléculaire et Cellulaire végétale*. Université Paris XI.

Lincoln, S.E., M.J. Daly, E.S. Lander. 1993. Mapping of genes controlling quantitative traits using MAPMAKER/QTL version 1.1: a tutorial and reference manual. Whitehead Institute for biomedical research technical report.

Liu, Z., G.R. Furnier. 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. *Theor Appl Genet* 87: 97-105.

Lockwood, G., M. End. 1992. History, techniques and future needs for cacao collections. In: *International Workshop on Conservation, Characterisation and Utilisation of cocoa genetic resources in the 21st century*, Trinidad and Tobago. 1-14.

Lorieux, M., X. Perrier, D. Gonzalez de Leon, C. Lanaud. 1994. Maximum likelihood models for mapping genetic markers showing segregation distortions. 1. Backcross populations. *Theor Appl Genet* in press.

Lu, Y.H., A. D'Hont, Walker Dit, P.S. Rao, P. Feldmann, J.C. Glaszmann. 1994. Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. *Euphytica* (in press)

Mansur, L.M., K.G. Lark, H. Kross, A. Oliveira. 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). *Theor. Appl. Genet* 86:907-913.

- Marcio, C., V.C. De Almeida, C.F. Gomes de Almeida. 1987. Coleta de cacau silvestre no estado de Rondonia, Brasil;. *Revista Theobroma* 17:65-92.
- Martin, G.B., J.G.K. Williams, S.D. Tanksley. 1991. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. *Proc. Natl. Acad. Sci. USA* 88:2336-2340.
- Martinson, V.A. 1975. Cytological studies of diploid and tetraploid *Theobroma cacao*. *Genetica* 45:341-348.
- Martinson, V.A., D. Adomako, M. Manu. 1987. Forty years of cocoa research at the Cocoa Research Institute of Ghana. *Cocoa Growers' Bulletin* 38: 39-55.
- Michelmore, R.W., I. Paran, R.V. Kesseli. 1991. Identification of markers linked to disease resistance genes by bulk segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci.* 88:9828-9832.
- Morera, J., A. Paredes, A. Mora. 1991. Germoplasma de cacao en el CATIE entre 1974 y 1991. CATIE, IICA.
- Mossu, G. 1974. La conservation du pouvoir germinatif des graines de cacaoyer. *Café, Cacao, Thé* 18:179-186.
- N'goran, J.A.K., P. Lachenaud, P. Bastide, D. Paulin. 1992. Cocoa germplasm conservation initiatives in Côte d'Ivoire. In: *International Workshop on Conservation, Characterisation and Utilisation of cocoa genetic resources in the 21st century*, Trinidad and Tobago. 221-227.
- N'goran, J.A.K., V. Laurent, A.M. Risterucci, C. Lanaud. 1994. Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity In press*.
- Nei, M. 1972. Genetic distance between populations. *Am. Nat.* 106:283-292.
- Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583-590.
- Nei, M., K. Syakudo. 1958. The estimation of outcrossing in natural populations. *Japan Journal of Genetics* 33:46-51.
- Nelson, C.D., W.L. Nanace, R.L. Doudrick. 1993. A partial genetic linkage map of slash pine (*Pinus elliottii* Engelm var. *elliottii*) based on random amplified polymorphic DNAs. *Theor. Appl. Genet.* 87:145-151.
- Ooi, L.H., P.S. Chew. 1985. Results of five progeny trials on hybrid cocoa in Peninsular Malaysia. *Planter, Kuala Lumpur* 61:54-69.
- Paik-Ro, O.G., R.L. Smith, D.A. Knauft. 1992. Restriction fragment length polymorphism evaluation of six peanut species within the *Arachis* section. *Theor. Appl. Genet.* 84:201-208.
- Paterson, A.H., S. Damon, J.D. Hewitt, D. Zamir, H.D. Rabinowitch, S.E. Lincoln, E.S. Lander, S.D. Tanksley. 1991. Mendelian factors underlying quantitative traits in tomato : comparison across species, generations, and environments. *Genetics* 127:181-197.

- Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335:721-726.
- Paulin, D., B. Decazy, N. Coulibaly. 1983. Etude des variations saisonnières des conditions de pollinisations et de fructification dans une cacaoyère. *Café, Cacao, Thé* 27:165-175.
- Paulin, D., G. Mossu, Ph. Lachenaud, C. Cilas. 1993. La sélection du cacaoyer en Côte d'Ivoire: Analyse du comportement de soixante deux hybrides dans quatre localités. *Café, Cacao, Thé* 36: 3-20.
- Penner, G.A., J. Chong, M. Levesque-Lemay, S.J. Molnar, G. Fedak. 1993. Identification of a RAPD marker linked to the oat stem rust gene Pg3. *Theor. Appl. Genet.* 85:702-705.
- Pérez-Nasser, N., L.E. Eguiarte, D. Pinero. 1993. Mating system and genetic structure of the distylous tropical tree *Psychotria faxlucens* (Rubiaceae). *Am. J. Bot.* 80:45-52.
- Pires, J.L., W.R. Monteiro, B.G.D. Bartley. 1992. CEPLAC's germplasm collection-Quantitative evaluation of the cacao collection at the Cacao Research center(CEPEC) in Brahia, Brazil. In: *International Workshop on Conservation, Characterisation and Utilisation of cocoa genetic resources in the 21st century*, Trinidad and Tobago: 256-268.
- Pittier, H. 1930. A propos des cacaoyers spontanés. *Rev. Bot. Appl.* 10:777.
- Pound, F.J. 1933. Criteria and methods of selection in cacao. *2d annual report on cocoa research* 27-29.
- Pound, F.J. 1938. Cacao and witchbroom disease (Marasmius perniciosus) of South America. *Archives of cocoa research* 1938 1:20-72.
- Pound, F.J. 1945. A note on the cocoa population of South America. In: *Cocoa conference*, London. 131-133.
- Raboin, L.M., D. Paulin, C. Cilas, A.B. Eskes. 1993. Analyse génétique de quelques caractères des fleurs du cacaoyer (*Theobroma cacao* L.): leur intérêt pour l'évaluation de la diversité de l'espèce. *Café, Cacao, Thé* 37:273-282.
- Reiter, R.S., J.G.K. Williams, K.A. Feldmann, J.A. Rafalski. 1992. Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proc. Natl. Acad. Sci.* 89:1477-1481.
- Reyes, H.E. 1992. Criollo cacao germplasm in Venezuela. In: *International Workshop on Conservation, Characterisation and Utilisation of cocoa genetic resources in the 21st century*, Trinidad and Tobago. 244-252.
- Rieseberg, L.H., H. Coi, R. Can, C. Spore. 1993. Genomic map of diploid hybrid species. *Heredity* 70:285-293.
- Rodolphe, F., M. Lefort. 1993. A multi-marker model for detecting chromosomal segments displaying QTL activity. *Genetics* 134:1277-1288.
- Rowland, L.J., A. Levi. 1994. RAPD-based genetic linkage map of blueberry derived from a cross between diploid species (*Vaccinium darrowi* and *V. elliotii*). *Theor Appl genet* 87:863-868.

Ruf, F. 1993. La Côte d'Ivoire va t-elle céder la place de premier producteur mondial de cacao à l'Indonésie? *Café, Cacao, Thé* 37:227-242.

Saiki, R.K., D.H. Gelfang, S. Stoffel, S.J. Higuchi, R. Horn., G. Mullis, A.H. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.

Saiki, R.K., Glyllensten, H. Erlich. 1988. The Polymerase Chain Reaction. In: *Genome analysis; A practical Approach.*, Davies, K.E. (eds), 141-152, IRL Press, Washington D.C.

Saiki, R.K., S. Scharp, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, N. Arnheim. 1985. Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350-1354.

Sallee, B. 1987. Rapport de prospection en Guyane française. IRCC.

Sambrook, J., E.F. Fritsch, T. Maniatis. 1989. *Molecular cloning: a laboratory manual* (second edition). Cold Spring Harbor Laboratory Press, New York.

Second, G., A. Ghesquière. 1994. Cartographie des introgressions reciproques entre les sous espèces Indica et Japonica de riz (*Oryza sativa* L.). In: *Techniques et Utilisations des marqueurs moléculaires*, Montpellier.

Shin, S.J., S. Chao, L. Corpus, T. Blake. 1990. A practical map of the barley genome incorporating restriction fragment length polymorphism, polymerase chain reaction, isozyme and morphological marker loci. *Genome* 33:803-810.

Soller, M., T. Brody, A. Genizi. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor. Appl. Genet.* 47:35-59.

Soria, J., F. Ocampo, G. Paez. 1974. Parental influence of several cacao clones on the yield performance of their progenies. *Turrialba* 24:58-65.

Soria, J.V. 1970. Principal varieties of cocoa cultivated in tropical America. *Cocoa Growers' Bulletin* 19:12-21.

Soria, V., O. Esquivel. 1970. Relationship between precocity, growth and yield in cacao. *Turrialba* 20:193-197.

Soria, V.J., G.A. Enriquez. 1981. International cocoa cultivar catalogue. Turrialba (Costa Rica).

Sounigo, O., J.A.K. N'goran, N. Coulibaly, D. Clément, P. Lachenaud. 1993. Evaluation de clones de cacaoyers pour la productivité, la résistance aux mirides et la résistance à la pourriture. In: *XI Conférence Internationale sur la recherche cacaoyère*, Yamoussoukro, Côte d'Ivoire.

Swoford, D.L., R.B. Selander. 1981. BIOSYS-1: a FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *J. Hered.* 72:281-283.

Thomas, M.R., S. Matsumoto, P. Cain, N.S. Scott. 1993. Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification. *Theor. Appl. Genet.* 86:173-180.

Thong, J.K.C., S.K. Ng, H.S.H. Ooi, K.Y. Leng. 1992. Cocoa in Penninsular Malaysia: I. The early history. *Cocoa Grower's Bulletin* 45:7-25.

- Tinker, N.A., M.G. Fortin, D.E. Mather.** 1993. Random amplified polymorphic DNA and pedigree relationships in Spring barley. *Theor. Appl. Genet.* 85:976-984.
- Toxopeus, H.** 1969. *Theobroma cacao* L in Outlines of perennial breeding in the Tropics. *Veenman, Wageningen* 79-109.
- Toxopeus, H.** 1972. Cocoa breeding: a consequence of mating system, heterosis and population structure. In: *Cocoa and coconuts in Malaysia*, Kuala Lumpur, Malaysia. 3-12.
- Toxopeus, H., Giesberger.** 1983. History of cocoa and cocoa research in Indonesia. *Archives of cocoa research* 2:7-34.
- Tulsieram, L.K., J.C. Glaubitz, G. Kiss, J.E. Carlson.** 1992. Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. *biotechnology* 10:686-690.
- Velboom, L.R., M. Lee, W.L. Woodman.** 1994. Molecular marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL for morphological traits. *Theor. Appl. Genet* 88:7-16.
- Vierling, R.A., Z. Xiang, C.P. Joshi, M.L. Gilbert, H.T. Nguyen.** 1994. Genetic diversity among elite *Sorghum* lines revealed by restriction fragment length polymorphisms and random amplified polymorphic DNAs. *Theor. Appl. Genet.* 87:816-820.
- Weir, B.S.** 1990. Genetic data analysis. 373-410pp.
- Wellensiek, S.K.** 1931. The genetics of cotyledon-colour of cocoa as a basis for quality selection. 2:245-261.
- Welsh, J., M. McClelland.** 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18:7213-7218.
- White, T.J., N. Arnheim, H.A. Erlich.** 1989. The polymerase chain reaction. *TIG* 5:185-189.
- Wilde, J., R. Waugh, P. Powell.** 1992. Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. *Theor Appl Genet* 83:871-877.
- Williams, C.E., D.A. St. Clair.** 1993. Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. *Genome* 36:619-630.
- Williams, J.G., A.R. Kubelik, K.J. Livak, J.A. Rafalski, S.V. Tingey.** 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18:6531-6535.
- Williams, M.N.V., N. Pande, S. Nair, M. Mohan, B. J.** 1991. Restriction fragment length polymorphism analysis of polymerase chain reaction products amplified from mapped loci of rice (*Oryza sativa* L.) genomic DNA. *Theor. Appl. Genet* 82:489-498.
- Wood, G.A.R.** 1985. History and development. In: *Cocoa*, Longman (eds), 1-10, London.
- Wood, G.A.R., R.A. Lass.** 1985. *Cocoa*. 4, 620pp. Longman, London.

**Wright, S.** 1978. Evolution and the genetics of population: variability within and among natural populations. Chicago.

**Wrischnik, L.A., R.G. Higuchi, M. Stoneking, H.A. Erlich, N. Arnheim, A.C. Wilson.** 1987. Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nuc. Ac. Res* 15:529-541.

**Yang, X., C. Quiros.** 1993. Identification and classification of celery cultivars with RAPD markers. *Theor. Appl. Genet.* 86:205-212.

**Yin, J.P.T., D.H.K. Lim, G. Lockwood.** 1991. Clonal cocoa trials and selection for superior clones. In: *International cocoa conference: challenges in the 90s. In press.*

## Annexe 1: Extraction d'ADN total

Pour 8 g de feuilles sèches broyées répartis dans 16 tubes:

Citrate de sodium 0.2 N, pH 8	260 ml
EDTA 500mM, pH 8	40 ml
SDS 20 %	20 ml
Mercaptoéthanol pur	2 ml

- Répartir 16 ml de tampon d'extraction dans chacun des 16 tubes  
ajouter 500 mg de poudre de feuilles séchées finement broyées, une spatule de polyvinylpyrrolidone et 100 µg de protéinase K à 12.5 mg/ml.
- Incuber 1 heure au bain marie à 50 °C avec agitation, puis 20 min à 65 °C.
- Ajouter 5 ml d'acétate de potassium 5M, pH 8 et incuber 15 min à 0 °C.
- Ajouter 9 ml de PCIAA ( Phénol-chloroforme: isoamylalcool 25-24:1).
- Centrifuger 1 heure à 4200g, prélever la phase aqueuse et y ajouter 9 ml d'isopropanol : bien mélanger doucement.
- Prélever la pelote d'ADN à l'aide d'un hameçon en verre et la laisser reprendre une nuit à 4 °C dans 500 µl de tampon de reprise(TrisHCl 50mM pH8, EDTA 10mM).
- Incuber à 37 °C pendant 20 min avec de la ribonucléase T1 à 25 mg/ml.
- Ajouter 100 µl d'acétate de sodium 3 M pH8 et 1 ml d'éthanol absolu et incuber 1 heure à 4° C.
- Centrifuger 20 min à 12000 g , rincer le culot avec 1 ml d'éthanol 80 % et récentrifuger pendant 5 min.
- Sécher le culot à l'air libre et le résuspendre dans 500 µl de TE (TrisHCl 10mM pH8, EDTA 1mM). La pureté de l'ADN obtenue est suffisante pour réaliser des réactions RAPD.

## Annexe 2: PROTOCOLE D'AMPLIFICATION D'ADN PAR LA PCR(RAPD).

- Décongeler un tube d'eau ultra pure, le tampon de réaction, les solutions d'ADN, les primers et les nucléotides
- Mettre en marche le block thermique
- Numéroter les tubes et préparer une solution mère avec les réactifs dans l'ordre et les quantités suivants:

Pour une réaction de 25 ul par tube

	Volume	Concentration finale
Eau	8,7ul	
Tampon 10X	2,5 ul	x
MgCl <sub>2</sub>	1,5 ul	1,5 mM
NTP	2 ul	200 uM
Primer	5 ul	0,2 uM
Taq	0,3	1,5 unité
ADN	5 ul	2.5ng

NB: avant d'ajouter la Taq, bien vortexer la solution mère. Après avoir ajouté la Taq, bien mélanger en penchant plusieurs fois le tube (ne pas vortexer)

Centrifuger la solution et distribuer 20 ul de la solution mère par tube puis ajouter 5 ul d'ADN à 0.5 ng/ ul

-Recouvrir le mélange d'une goutte d'huile minérale ultra pure et mettre les tubes dans la machine. Amplifier à l'aide du programme ci- dessous.

### Programme d'amplification sur machine Techne PHC2

1 cycle: 94 °C pendant 5 mn  
36°c pendant 45 s  
72 °c pendant 2 mn  
45 cycles 94 °c pendant 50 s  
36°c pendant 45 s  
72 °c pendant 2 mn  
1 cycle 94 °c pendant 50 s  
36°c pendant 45 s  
72 °c pendant 10 mn  
4°c pendant 30 H

Préparer un gel d'agarose dans du TBE à 1,4%. A la fin du programme d'amplification ajouter 1/10 du volume de réaction de Bleu Stop (10X) à chaque tube pour arrêter la réaction.

Deposer les produits d'amplification dans le gel et les faire migrer à 100 V pendant 2 à 3 heures. Colorer le gel pendant 20 min au BET et le photographier sous U.V.

### Annexe 3: Protocole d'extraction d'ADN pour analyses RFLP

Pour 4 g de feuilles fraîches broyées en présence de PVP réparties dans trois tubes, préparer 22 ml de tampon d'extraction comme suit:

Réactifs	volume	Concentration finale
Citrate de sodium 0.2N, pH 8	18 ml	0,16 N
EDTA 800mM, pH 8	3 ml	70 mM
SDS 10 %	1 ml	1,25 %
Bisulfite		

Incuber sous agitation à 65°C pendant 2 heures

Ajouter 8 ml d'acetate de sodium 5M pH 8 et laisser 5 min dans la glace

Ajouter 20 ml de chloroforme: isoamylalcool 24:1 (CIAA), centrifuger 40 min à 3500g

Recupérer la phase aqueuse et mettre dans le même volume d'isopropanol

Récupérer la pelote dans les trois tubes d'extraction et y ajouter 4 ml de TE.

Laisser reprendre à 4°C pendant une semaine.

- dissoudre 10.3 de chlorure de césium dans du TE et ajouter à l'ADN: ajuster le volume à 12 ml avec du TE.

- Transvaser dans un tube à sceller avec 300 µl de bromure d'éthidium à 10 mg/ml.

- Centrifuger à 40000 g pendant au moins 40 heures.

- Récupérer la bande d'ADN sous rayonnement U.V.

- Eliminer le bromure d'éthidium avec du butanol-1 saturé en TE.

- Précipiter avec 3 volumes d'éthanol 70 % pendant 2 heures à -20 °C.

- Centrifuger à 10000 g pendant 30 min et laisser reprendre dans 500µl de TE pendant au moins 72 heures.

#### Annexe 4: Préhybridations et hybridations

La préhybridation et l'hybridation sont faites à 42°C dans 10 ml de tampon suivant, pour un tube contenant une membrane (conserver le reste du tampon pour les hybridations)

Réactifs	volume pour 50 ml	Concentration finale
14.5xSSC x 36 % Dextran	13.8 ml	10 % de dextran
-20xSSC	5 ml	6x
- formamide	25 ml	50%
-Denhart 50	5 ml	5x
20xSDS	1.25ml	0.5%
ADN de sperme de hareng	150 µl	25 µg/ml

#### marquage

pour un volume final de 50 µl

-50 ng de sonde dans 28 µl d'eau

-4 µl de primer

faire bouillir pendant 5 min , laisser refroidir (2 mn) et ajouter

-10 µl de tampon d'enzyme

-2 µl d'enzyme polymerase

- 4 µl  $\infty^{32}\text{PdCTP}$

incuber 1 H à 37°C puis ajouter

-250 µl de TE

faire bouillir pendant 5 min , mettre dans la glace pendant 5 min et mélanger à la solution d'hybridation.

## Annexe 5

### Digestion enzymatique d'ADN

Pour chaque digestion d'un volume final de 100  $\mu$ l

-7 $\mu$ g d'ADN par digestion

- 10 % de tampon d'enzyme soit 10  $\mu$ l

- 5 % de spermidine 40 mM soit 5 ml

- 5 % ribonucléase T1 à 25mg/ml soit 5 ml

- 5% de BSA soit 5 $\mu$ l

- 1 ml/mg d'endonucléase dosé à 10 unité par ml

- Eau stérile pour compléter le volume à 100  $\mu$ l

Incuber une nuit à 37 °C .

Purifier l'ADN digéré avec:

-1 volume de Phénol:chloroforme:isoamylalcoöl (PCIAA)

-centrifuger pendant 20 min à 12000g

-précipiter avec 1/20 volume de NaCl 5 M et 2 volumes d'éthanol absolu

-laisser reposer 30 min à -20 °C

-centrifuger 20 min à 12000 g , récupérer le culot et resuspendre dans du 20  $\mu$ l de TE.

### transfert sur membrane

Incuber le gel pendant 10 min dans du Hcl 0.25N

Rincer à l'eau distillée

Incuber 30 min dans NaOH 0.4 N

Dans un bac contenant 1.5l de NaOH 0.4 N, placer une éponge , puis 3 feuilles de papier filtre (Watman), une membrane nylon Hybond N+ , 3 feuilles de papier filtre (Watman), puis un paquet de papier Kleenex. Poser sur l'ensemble une plaque de verre.

Laisser transférer toute la nuit. Rincer la membrane dans du 2xSSC et la sécher entre deux papiers absorbants (La membrane peut se conserver sèche).

## Annexe 6

### Récupération d'ADN par utilisation de gel d'agarose à bas point de fusion

Faire un gel avec de l'agarose à bas point de fusion à 1.2 % dans du TBE  
faire migrer à 100 V pendant quelques heures selon la taille du fragment et  
colorer au BET

Récupérer sous U.V. le fragment à l'aide d'une paille et d'une pipette (ne pas  
exposer trop longtemps aux U.V.)

Conserver au réfrigérateur.

## Annexe 7

CLONE		PDCAB	NBFEV	LGCAB	DIAMCAB	PVIDE	CORTEX
ICS53	NB OBS	31	31	23	23	23	23
	Moyenne	653.9	32.84	183.74	89.39	517.35	47.91
	écart type	156.14	10.31	18.72	5.67	112.74	10.42
ACU85	NB OBS	103	103	26	26	26	26
	Moyenne	365.8	27.83	156.85	76.19	303.35	45.73
	écart type	107.24	11.56	17.17	7.4	88.14	15.45
UPA402	NB OBS	26	26	24	24	24	24
	Moyenne	295.92	30.12	118.63	73.33	209	44.58
	écart type	89.86	11.76	17.6	6.88	62.58	12.72
IMC67	NB OBS	56	35	26	26	26	26
	Moyenne	791.41	47.11	203.96	100.77	711.15	49.19
	écart type	244.76	14.81	20.52	7.4	172.98	13.09
UPA409	NB OBS	100	100	26	26	26	26
	Moyenne	525.57	28.83	147.92	88.96	445.54	51.19
	écart type	120.09	11.35	12.82	6.11	100.78	13.52
WA40	NB OBS	102	102	26	26	26	26
	Moyenne	394.9	28.5	157.08	80.04	342.73	58.23
	écart type	106.15	13.21	17.35	5.78	79.91	12.82
IFC15	NB OBS	14	14	14	14	14	14
	Moyenne	393.71	38.64	145.29	75.57	273	35.79
	écart type	97.7	8.86	12.99	6.43	62.47	6.74
ICS84	NB OBS	100	100	26	26	26	26
	Moyenne	809.06	35.78	208	95.12	706.85	55.58
	écart type	164.44	7.53	20.37	6.04	170.43	15.73
ICS100	NB OBS	46	46	26	26	26	26
	Moyenne	478.09	21.54	151.12	88.31	423.15	63.31
	écart type	149.37	9.73	14.22	7.06	102.67	13.52
UPA413	NB OBS	103	103	26	26	26	26
	Moyenne	472.54	33.48	160.5	89.12	448.46	50.73
	écart type	146.19	12.96	9.45	4.25	61.22	16
T63/967	NB OBS	100	100	26	26	26	26
	Moyenne	440.74	31.44	171.77	81.73	371.27	49.85
	écart type	124.91	11.73	11.66	3.6	56.26	8.75
T60/887	NB OBS	103	103	26	26	26	26
	Moyenne	422.37	33.91	153.23	83.23	353.92	50.27
	écart type	81.02	8.99	9.76	4.15	65.56	10.47
P32A	NB OBS	36	36	27	27	27	27
	Moyenne	328.86	26.92	135.44	71.37	249.85	40.67
	écart type	62.42	9.08	11.2	3.55	40.04	8.41
R15	NB OBS	16	16	16	16	16	16
	Moyenne	519.5	40.5	170.87	84.56	365.37	23.38
	écart type	102	4.72	18.64	6.4	85.49	9.67
IFC19	Points	17	17	17	17	17	17
	Mean	771.59	45.35	195.24	91.94	570.59	41.76
	écart type	199.97	3.41	20.38	5.73	169.93	6.93
SPEC160	NB OBS	37	37	26	26	26	26
	Moyenne	651.35	27.65	175.92	93.35	558.35	42.85
	écart type	175.84	10.13	19.21	7.94	140.15	13.54

### CARACTERISTIQUES DES CABOSSES

PVIDE=poids vide en g NB OBS=nombre d'observations

LGCAB=longueur de cabosse en mm DIAMCAB=diamètre de cabosse en mm

CORTEX=épaisseur de cortex en mm

## Annexe 7 suite

CLONE		PDCAB	NBFEV	LGCAB	DIAMCAB	PVIDE	CORTEX
P19A	NB OBS	50	50	27	27	27	27
	Moyenne	326.08	27.8	154.22	72.7	270.81	62.93
	écart type	82.25	12.22	14.58	5.57	61.16	9.6
P13B	NB OBS	68	68	27	27	27	27
	Moyenne	261.22	25.63	141.89	67.07	225.15	39.89
	écart type	60.35	8.43	10.77	3.73	41.44	6.78
MO81	NB OBS	101	101	34	34	0	34
	Moyenne	565.75	43.37	193.62	98.82	0	51.24
	écart type	200.17	11.54	17.97	8.95	0	16.61
ACT2-11	NB OBS	93	93	27	27	27	27
	Moyenne	505.1	26.86	154.41	87.04	380.37	23.11
	écart type	127.24	7.95	9.36	5.71	80.41	12.4
IFC413	NB OBS	104	104	26	26	26	26
	Moyenne	303.87	33.35	155.73	72.81	200.62	24.46
	écart type	62.96	7.39	15.36	5	49.02	6.94
IFC6	NB OBS	104	104	36	36	35	36
	Moyenne	503.19	34.44	190.33	78.61	403.23	35.33
	écart type	123.65	8.15	25.96	6.88	113.42	9.97
DR1	NB OBS	27	27	24	24	24	24
	Moyenne	601.22	36	192.08	85.54	437.08	34.92
	écart type	112.56	7.64	17.32	5.24	82.91	10.05
SF23	NB OBS	26	26	22	22	22	22
	Moyenne	389.46	39.5	135.32	81.05	275.91	27.09
	écart type	82.57	8.76	15.75	5.24	61.46	8.25
NA32	NB OBS	101	101	26	26	26	26
	Moyenne	400.3	35.59	151.88	78.38	310.38	44.54
	écart type	76.24	8.21	10.48	4.25	48.12	11.55
IFC5	NB OBS	101	101	42	30	30	30
	Moyenne	423.86	38.55	146.45	81.5	316	26.37
	écart type	77.74	5.17	11.41	5.12	70.4	8.05
UF667	NB OBS	102	102	28	28	28	26
	Moyenne	674.12	33.33	174.36	94.93	520	45.73
	écart type	144.4	9.33	13.26	6.42	113.79	14.23
MAT1-6	NB OBS	84	78	26	26	26	26
	Moyenne	408.89	34.76	131.08	86.96	359.42	42.19
	écart type	86.67	7.47	9.97	4.2	49.95	11.97
SNK12	NB OBS	100	100	26	26	26	26
	Moyenne	938.79	40.61	219.15	105.42	905	61.81
	écart type	201.01	6.63	17.28	7.04	201.76	15.24
ICS60	NB OBS	34	34	30	26	26	26
	Moyenne	673.53	38.88	210.17	88.27	475.69	41.92
	écart type	186.54	8.93	28.31	7.34	143.34	8.78
MO98	NB OBS	37	37	32	26	26	26
	Moyenne	796.22	50.05	180.38	99.35	623.58	48.27
	écart type	173.55	15.34	19.61	4.23	112.1	9.29
SCA6	NB OBS	91	91	26	26	26	26
	Moyenne	462	44.95	173.81	79.62	397.35	51.35
	écart type	115.15	10.63	12.11	6.57	95.19	10.53

CARACTERISTIQUES DES CABOSSES

## Annexe 7 fin

CLONE		PDCAB	NBFEV	LGCAB	DIAMCAB	PVIDE	CORTEX
MOQ413	NB OBS	76	76	27	27	27	27
	Moyenne	577.26	32.36	150.96	95.7	522.26	46.78
	écart type	169.97	9.56	8.86	5.22	94.63	12.07
MOQ647	NB OBS	23	23	23	23	23	23
	Moyenne	648.26	38.13	192.96	84.17	501.61	67.65
	écart type	227.38	9.77	22.95	8.23	180.34	19.45
LAF1	NB OBS	16	16	16	16	16	16
	Moyenne	496.31	28.75	159.88	84.06	354.06	34
	écart type	131.27	10.14	16.94	8.16	82.78	8.57
AMAZ15	NB OBS	69	69	26	27	27	27
	Moyenne	585.36	26.72	160.42	90.7	547.59	61.11
	écart type	179	11.69	14.64	5.18	102.89	14.05
MO9	NB OBS	30	30	27	27	27	27
	Moyenne	462.3	37.23	164.37	85.41	370.93	42.85
	écart type	138.57	10.61	18.72	7.22	109.15	12.73
PA13	NB OBS	34	34	24	27	27	27
	Moyenne	410.56	18.85	166.08	76.48	359.04	67.48
	écart type	168.53	12.09	18.61	7.93	124.89	17.82
LAF3	NB OBS	20	20	20	20	20	20
	Moyenne	432	33.55	146.05	80.7	327.65	45.55
	écart type	99.59	5.07	14.43	5.49	89.63	11.42
UPA620	NB OBS	108	108	27	27	27	27
	Moyenne	526.28	35.97	165.74	88.78	450.78	53.15
	écart type	98.11	8.52	10.39	4.07	90.59	13.2
EQX107	NB OBS	34	34	26	26	26	26
	Moyenne	541.82	17.71	177.35	77.5	477.81	96.35
	écart type	198	10.56	31.26	8.07	159.63	20.74
SF1	NB OBS	105	105	27	27	27	27
	Moyenne	298.47	36.46	143.11	74.7	238.04	28.96
	écart type	81.44	13.43	12.62	4.76	45.15	6.5
ICS95	NB OBS	103	103	27	27	27	27
	Moyenne	605.75	36.74	208.41	85.26	555.63	49.44
	écart type	123.09	6.65	14.37	4.74	94.79	12.57
ICS39	NB OBS	103	103	27	27	27	27
	Moyenne	496.6	35.6	204.3	87	451.15	51.67
	écart type	124.9	8.87	16.52	5.86	91.4	13.94
ICS89	NB OBS	103	103	27	27	27	27
	Moyenne	488.56	23.9	168.41	83.56	461.63	66.11
	écart type	124.02	8.91	15.58	5.18	116.77	15.93
UPA401	NB OBS	50	50	26	26	26	39
	Moyenne	334.4	25.82	135.12	79.35	274.88	24.62
	écart type	116.05	11.26	15.94	6.52	68.17	19.71
PA121	NB OBS	95	95	27	27	27	27
	Moyenne	374.56	28.56	167.63	77.81	308.93	40.85
	écart type	102.99	7.36	15.66	5.23	71.22	11.49
IMC31	NB OBS	17	17	17	17	17	17
	Moyenne	545.18	37.59	153.12	88.29	404.88	54.12
	écart type	132.42	14.87	22.8	6.44	88.88	13.19
P7	NB OBS	70	70	26	26	26	26
	Moyenne	423.6	28.94	154.42	83.69	385.12	47.19
	écart type	109.5	11.46	14.82	5.98	85.89	12.23

## Annexe 8

CLONE		PFEV	LGFEV	EPFEV
MO9	Nb Obs	102	102	102
	MOYENNE	13.02	204.97	83.745
	écart type	1.4957	17.174	8.8583
LAF3	Nb Obs	100	100	100
	MOYENNE	16.79	231.78	86.32
	écart type	2.4049	18.916	11.582
UPA620	Nb Obs	104	104	104
	MOYENNE	20.635	240.14	83.942
	écart type	2.2167	13.154	12.843
SF1	Nb Obs	102	102	102
	MOYENNE	13.461	221.91	67.559
	écart type	2.4763	20.933	10.867
ICS39	Nb Obs	102	102	102
	MOYENNE	28.343	281.41	94.775
	écart type	4.3361	21.107	11.528
UPA401	Nb Obs	103	103	103
	MOYENNE	17.291	239.97	77.029
	écart type	1.6606	16.08	13.189
ICS89	Nb Obs	101	101	101
	MOYENNE	24.832	248.68	94.832
	écart type	3.2468	22.292	14.826
ICS95	Nb Obs	104	104	104
	MOYENNE	17.125	221.35	90.394
	écart type	2.3676	12.85	9.94
PA121	Nb Obs	108	108	108
	MOYENNE	15.269	216.91	81.241
	écart type	1.3983	11.877	9.2745
IMC31	Nb Obs	104	104	104
	MOYENNE	17.077	251.41	60.885
	écart type	1.9443	15.594	9.9262
MOQ647	Nb Obs	102	102	102
	MOYENNE	19.431	258.57	82.647
	écart type	2.6829	22.488	11.198
P19A	Nb Obs	104	104	104
	MOYENNE	13.519	221.7	72.894
	écart type	1.6778	16.634	8.5103
P13B	Nb Obs	105	105	105
	MOYENNE	11.324	196.7	67.143
	écart type	0.12945	1.6172	1.0718
EQX107	Nb Obs	92	92	92
	MOYENNE	16.554	219.37	97.25
	écart type	2.7149	18.823	9.9423

LGFEV=longueur de fève en mm PFEV=poids de fève en g

EPFEV=épaisseur de fève en mm

NB OBS=nombre d'observations

CARACTERISTIQUES DES FEVES

## Annexe 8 (suite)

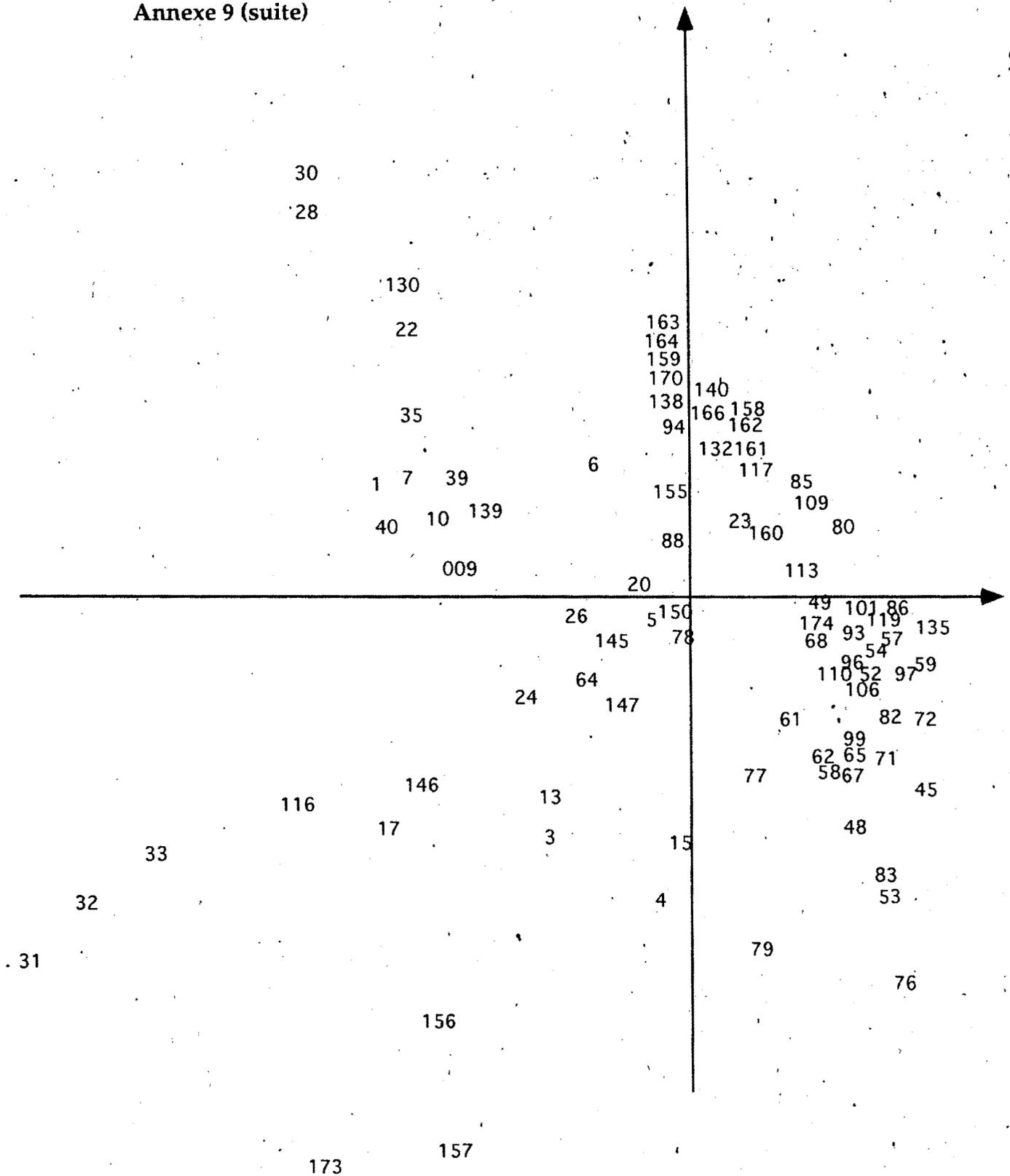
CLONE		PFEV	LGFEV	EPFEV
MO81	Nb Obs	100	100	100
	MOYENNE	12.48	205.41	81.12
	écart type	1.9095	13.876	9.911
ACT2-11	Nb Obs	100	100	100
	MOYENNE	15.71	224.61	82.38
	écart type	2.2261	19.611	11.408
R15	Nb Obs	101	101	101
	MOYENNE	20.475	246.5	87.564
	écart type	3.002	16.5	13.304
IFC413	Nb Obs	100	100	100
	MOYENNE	18.59	226.45	84.89
	écart type	2.9339	22.962	12.137
IFC6	Nb Obs	100	100	100
	MOYENNE	17.01	224.68	97.79
	écart type	3.3288	22.186	11.913
DR1	Nb Obs	100	100	100
	MOYENNE	25.68	255.79	104.22
	écart type	3.629	21.567	13.052
SF23	Nb Obs	100	100	100
	MOYENNE	15.6	231.45	83.27
	écart type	1.9949	18.939	9.7751
NA32	Nb Obs	101	101	101
	MOYENNE	13.218	211.06	68.168
	écart type	2.3221	18.309	9.6167
IFC5	Nb Obs	100	100	100
	MOYENNE	15.18	225.78	81.34
	écart type	2.012	17.131	6.8198
UF667	Nb Obs	100	100	100
	MOYENNE	29.2	280.69	93.82
	écart type	4.3041	19.511	15.234
Mat1-6	Nb Obs	100	100	100
	MOYENNE	17.14	242.59	78.44
	écart type	1.7351	16.707	9.7404
ICS60	Nb Obs	100	100	100
	MOYENNE	29.62	280.97	97.7
	écart type	5.3669	25.39	13.162
IFC19	Nb Obs	100	100	100
	MOYENNE	21.22	249.99	88.68
	écart type	2.456	17.057	11.401
MO98	Nb Obs	100	100	100
	MOYENNE	15.61	246.71	65.86
	écart type	2.3523	23.536	15.747
SCA6	Nb Obs	100	100	100
	MOYENNE	10.9	205.24	70.54
	écart type	1.5923	15.214	7.3256
SPEC160-9	Nb Obs	100	100	100
	MOYENNE	19.64	233.13	93.62
	écart type	2.4268	16.968	13.901

## Annexe 8 (fin)

CLONE		PFEV	LGFEV	EPFEV
ICS53	Nb Obs	100	100	100
	MOYENNE	23.86	250.36	96.17
	écart type	3.6293	24.265	13.923
ACU85	Nb Obs	100	100	100
	MOYENNE	20.46	242.83	90.28
	écart type	2.672	20.013	12.716
UPA409	Nb Obs	100	100	100
	MOYENNE	16.09	244.42	75.26
	écart type	2.4125	22.743	13.559
UPA402	Nb Obs	101	101	101
	MOYENNE	15.941	238.51	74.842
	écart type	1.5545	16.216	11.871
WA40	Nb Obs	103	103	103
	MOYENNE	21.262	215.34	109.51
	écart type	2.3346	11.004	12.588
IMC67	Nb Obs	100	100	100
	MOYENNE	16.46	246.47	79.72
	écart type	2.047	20.156	9.1741
P7	Nb Obs	101	101	101
	MOYENNE	16.257	237.38	78.426
	écart type	2.5716	20.409	11.029
UPA413	Nb Obs	100	100	100
	MOYENNE	16.31	240.68	66.61
	écart type	1.6859	15.68	11.207
ICS100	Nb Obs	100	100	100
	MOYENNE	27.2	243.27	104
	écart type	2.3138	14.28	14.986
MOQ413	Nb Obs	101	101	101
	MOYENNE	20.574	263.64	76.436
	écart type	2.7725	17.142	11.74
T60/887	Nb Obs	100	100	100
	MOYENNE	14.64	213.81	81.2
	écart type	1.7146	14.203	12.155
T63/967	Nb Obs	102	102	102
	MOYENNE	19.549	243.89	79.206
	écart type	2.2455	16.409	15.258
SNK12	Nb Obs	103	103	103
	MOYENNE	21.786	263.2	90.631
	écart type	3.1705	24.586	11.106
AMAZ15-15	Nb Obs	104	104	104
	MOYENNE	22.183	270.08	90.298
	écart type	2.6723	20.713	12.703
PA13	Nb Obs	100	100	100
	MOYENNE	16.61	226.64	102.36
	écart type	2.9401	16.6	10.812
P32A	Nb Obs	102	102	102
	MOYENNE	14.196	205.89	88.892
	écart type	1.3862	19.192	12.502



Annexe 9 (suite)



Plan 1 (31%) de l'AFC sur les données cDNA

## Annexe9 (fin)

Nom de clone	N°	Nom de clone	N°	Nom de clone	N°	Nom de clone	N°
Amaz15	1	CUM214	51	ICS53	97	IFC2	159
EBC10	3	G8	52	ICS89	98	IFC4	160
ECNR	4	Hern212	53	ICS95	99	IFC 5	161
EET59	5	ICS39	54	ICS98	100	IFC15	162
EQX94	6	ICS60	57	IFC6	101	IFC307	163
EQX107	7	ICS100	58	IFC19	104	IFC361	164
IMC31	9	JS206	59	IFC413	106	MAT1-6	166
IMC67	10	La esmida	61	K5	109	Para	168
LCT-EEN 84	13	LAF1	62	LAF1 7	110	SIAL70	170
LCT-EEN 127	15	LAF3	64	MOQ413	113	SIC864	172
LCT-EEN 325	17	OC61 dl	65	MT1	116	SF23	173
LCT-EEN 371	18	OC73	67	N38	117	SPEC160/9	174
MO81	20	OC77	68	R15	119	UPA409	175
NA32	22	PA35	69	SC5	124		
Nacional	23	POC	71	SPA5	128		
PA13	24	POR	72	SPA17	130		
PA150	26	PORROJO	76	TJ1	132		
P2	28	Pvd201	77	UF221	135		
P32A	30	PV2	78	WA40	137		
SCA6	31	PV6	79	Comun Tipico	138		
SCA9	32	SPEC185	80	ERJOH1	139		
SCA12	33	ZEA1	82	VENC4	140		
T60/887	34	ZEA 206	83	ERJOH4	145		
T63/967	35	ACT2/11	85	ERJOH5	146		
T85/799	36	ACU85	86	ERJOH6	147		
UPA413	39	CC39	88	ERJOH9	150		
UPA603	40	CHUAO	89	ERJOH10	151		
CATA201	45	CNS23	92	ERJOH15	155		
CATA202	48	DR1	93	GUI144	156		
CATA203	49	G23	94	GUI349	157		
CHO31	50	ICS16	96	IFC1	158		

Les géotypes et leurs numeros sur les AFC

HA			BA		
	HETERO TOTAL			HETERO TOTAL	
C312	1	11	C1	1	26
EBC10	4	27	C15	3	21
EBC5	2	27	C2	0	22
EBC6	3	26	C303	3	24
IMC31	6	21	C307	1	22
IMC5	9	26	C361	0	26
IMC67	8	20	C4	7	23
IMC78	6	25	C414	7	24
LCT127	1	26	C5	3	27
LCT167	1	7	Com	1	27
LCT202	1	15	ECNR	6	19
LCT295	3	17	EET	5	21
LCT325	6	21	ERJOH1	2	12
LCT326	6	24	ERJOH10	1	18
LCT355	1	25	ERJOH10	2	11
LCT37	7	26	ERJOH11	4	24
LCT84	3	17	ERJOH12	7	27
MO81	4	14	ERJOH13	10	26
MO9	6	27	ERJOH14	3	27
MO98	5	20	ERJOH15	3	27
NA32	6	25	ERJOH2	5	27
P1	2	27	ERJOH3	6	26
P16	8	27	ERJOH4	1	9
P2	0	27	ERJOH5	5	23
P32A	3	27	ERJOH6	3	11
PA13	11	25	ERJOH7	2	18
PA150	0	11	ERJOH8	7	24
PA20	5	13	ERJOH9	0	12
SCA12	9	25	GUI144	0	26
Sca6	4	27	GUI154	0	27
SCA9	7	25	GUI346	0	26
SP11	5	20	GUI349	0	26
SPA17	7	20	MAT6	0	27
SPA5	5	20	MAT9	1	13
T60	3	22	NACIO	10	27
T63	7	21	PARA	0	19
T79	6	18	SF23	5	26
T85	6	21	SIAL325	0	27
UPA401	5	19	SIAL42	8	24
UPA413	7	21	SIAL70	0	27
UPA603	6	22	SC	1	25
UPA608	4	12	VEN1	1	27
			VEN11	1	26
			VEN20	0	27
			VEN31	0	8
			VEN5	1	17
			VENC4	3	21

Nombre de locus hétérozygotes au sein des génotypes étudiés par les RFLP



Annexe 11

**HISTOGRAMMES DE DISTRIBUTION DES CARACTERES QUANTITATIFS A BINGERVILLE**

**Poids de cabosses :PDCAB**

moyenne    écart type    coefficient de dissymetrie    coefficient d'aplatissement  
410.09        100.39                    0.45                    -0.20

---

209.31 |\*\*\*\*\*  
259.51 |  
309.70 |\*\*\*\*\*  
359.90 |\*\*\*\*\*  
410.09 |\*\*\*\*\*  
460.29 |\*\*\*\*\*  
510.48 |\*\*\*\*\*  
560.68 |\*\*\*\*\*  
610.87 |\*\*\*\*\*  
661.07 |

**Longueur de cabosse: LGCAB**

moyenne    écart type    coefficient de dissymetrie    coefficient d'aplatissement  
141.40        14.50                    0.11                    -0.40

112.39 |\*\*\*\*\*  
119.65 |\*\*\*\*\*  
126.90 |\*\*\*\*\*  
134.15 |\*\*\*\*\*  
141.40 |\*\*\*\*\*  
148.65 |\*\*\*\*\*  
155.90 |\*\*\*\*\*  
163.15 |\*\*\*\*\*  
170.40 |\*\*\*\*\*  
177.65 |\*\*\*\*\*|

**Diamètre de cabosse: DIAMCAB**

moyenne    écart type    coefficient de dissymetrie    coefficient d'aplatissement  
74.99        5.54                    0.11                    -0.36

63.91 |  
66.68 |\*\*\*\*\*  
69.45 |\*\*\*\*\*  
72.22 |\*\*\*\*\*  
77.99 |\*\*\*\*\*  
77.76 |\*\*\*\*\*  
80.53 |\*\*\*\*\*  
83.29 |\*\*\*\*\*  
86.06 |\*\*\*\*\*  
88.83 |\*\*\*\*\*

Annexe 11 (suite)

**HISTOGRAMMES DE DISTRIBUTION DES CARACTERES QUANTITATIFS A BINGERVILLE**

**Poids de cabosse vide: PVIDE**

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
 262.28 67.60 0.48 -0.31

127.07 |\*\*\*\*\*  
 160.87 |  
 194.68 |\*\*\*\*\*  
 228.48 |\*\*\*\*\*  
 262.28 |\*\*\*\*\*  
 296.08 |\*\*\*\*\*  
 329.89 |\*\*\*\*\*  
 363.69 |\*\*\*\*\*  
 397.49 |\*\*\*\*\*  
 431.29 |\*\*\*\*\*

**Nombre de fèves par cabosse: NBFEV**

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
 39.53 5.96 -0.78 -0.09

27.60 |  
 30.58 |\*\*\*\*\*  
 33.57 |\*\*\*\*\*  
 36.55 |\*\*\*\*\*  
 39.53 |\*\*\*\*\*  
 42.51 |\*\*\*\*\*  
 45.50 |\*\*\*\*\*  
 48.48 |\*\*\*\*\*  
 51.46 |  
 54.44 |

**Poids de fèves fraiches : PFF**

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
 149.18 41.71 0.35 -0.62

65.76 |  
 86.61 |\*\*\*\*\*  
 107.47 |\*\*\*\*\*  
 128.32 |\*\*\*\*\*  
 149.18 |\*\*\*\*\*  
 170.03 |\*\*\*\*\*  
 190.89 |\*\*\*\*\*  
 211.75 |\*\*\*\*\*  
 232.60 |\*\*\*\*\*  
 253.46 |\*\*\*\*\*

Annexe 11 (suite)

HISTOGRAMMES DE DISTRIBUTION DES CARACTERES QUANTITATIFS A BINGERVILLE

Poids de fèves: PFEV

moyenne écart type coefficient de dissymétrie coefficient d'aplatissement  
180.60 35.57 0.43 -0.37

109.45 |  
127.24 |  
145.03 |\*\*\*\*\*  
162.82 |\*\*\*\*\*  
180.60 |\*\*\*\*\*  
198.39 |\*\*\*\*\*  
216.18 |\*\*\*\*\*  
233.96 |\*\*\*\*\*  
251.75 |\*\*\*\*\*  
269.54 |

Longueur de fèves: LGFEV

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
24.71 2.23 -0.10 -0.40

20.24 |\*\*\*\*\*  
21.36 |\*\*\*\*\*  
22.47 |\*\*\*\*\*  
23.59 |\*\*\*\*\*  
24.71 |\*\*\*\*\*  
25.83 |\*\*\*\*\*  
26.94 |\*\*\*\*\*  
28.06 |\*\*\*\*\*  
29.18 |\*\*\*\*\*  
30.29 |\*\*\*\*\*

Epaisseur de fèves: EPFEV

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
77.84 1.15 1.54 4.75

5.54 |  
6.12 |\*\*\*\*\*  
6.69 |\*\*\*\*\*  
7.27 |\*\*\*\*\*  
7.84 |\*\*\*\*\*  
8.42 |\*\*\*\*\*  
8.99 |\*\*\*\*\*  
9.57 |\*\*\*\*\*  
10.14 |  
10.72 |

**Annexe 11 (suite)**

**HISTOGRAMMES DE DISTRIBUTION DES CARACTERES QUANTITATIFS A BINGERVILLE**

**Nombre d'ovules par ovaire: NBOVULE:**

moyenne	écart type	coefficient de dissymétrie	coefficient d'aplatissement
46.68	2.72	0.43	0.47

```

41.24 |*****
42.60 |*****
43.96 |*****
45.32 |*****
46.68 |*****
48.04 |*****
49.40 |*****
50.76 |*****
52.12 |*****
53.48 |
    
```

**Nombre de cabosse pourries: POURCAB**

moyenne	écart type	coefficient de dissymétrie	coefficient d'aplatissement
6.44	6.47	0.81	-0.61

```

-6.49 |
-3.26 |
-0.03 |
 3.21 |*****
 6.44 |*****
 9.67 |*****
12.91 |*****
16.14 |*****
19.37 |*****
22.60 |*****
    
```

**Pourcentage de cabosses pourries: %POURCAB**

moyenne	écart type	coefficient de dissymétrie	coefficient d'aplatissement
9.55	7.56	1.03	1.09

```

-5.56 |
-1.78 |
 2.00 |*****
 5.78 |*****
 9.55 |*****
13.33 |*****
17.11 |****
20.89 |*****
24.66 |****
28.44 |*****
    
```

Annexe 11 (fin)

HISTOGRAMMES DE DISTRIBUTION DES CARACTERES QUANTITATIFS A BINGERVILLE

Dégats de mirides: MIRIDES

moyenne	écart type	coefficient de dissymétrie	coefficient d'aplatissement
1.24	0.79	0.65	-0.67

0.05 |\*\*\*\*  
0.45 |\*\*\*\*\*  
0.84 |\*\*\*\*\*  
1.24 |\*\*\*\*\*  
1.63 |\*\*\*\*\*  
2.03 |\*\*\*\*\*  
2.42 |\*\*\*\*  
2.81 |\*\*\*\*\*  
3.21 |\*\*\*\*

Annexe 12

HISTOGRAMMES DE DISTRIBUTION DES CARACTERES QUANTITATIFS A DIVO

Poids de cabosses :PDCAB

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
469.35 106.70 0.57 -0.14

255.94 |  
309.29 |\*\*\*\*\*  
362.64 |\*\*\*\*\*  
415.99 |\*\*\*\*\*  
469.35 |\*\*\*\*\*  
522.70 |\*\*\*\*\*  
576.05 |\*\*\*\*\*  
629.40 |\*\*\*\*\*  
682.75 |  
736.10 |\*\*\*\*\*

Longueur de cabosse: LGCAB

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
153.96 11.22 0.04 -0.63

131.51 |  
137.12 |\*\*\*\*\*  
142.73 |\*\*\*\*\*  
148.34 |\*\*\*\*\*  
153.96 |\*\*\*\*\*  
159.57 |\*\*\*\*\*  
165.18 |\*\*\*\*\*  
170.79 |\*\*\*\*\*  
176.40 |\*\*\*\*\*  
182.01 |

Diamètre de cabosse: DIAMCAB

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
80.18 6.68 0.27 -0.56

70.17 |\*\*\*\*\*  
73.51 |\*\*\*\*\*  
76.84 |\*\*\*\*\*  
80.18 |\*\*\*\*\*  
83.52 |\*\*\*\*\*  
86.86 |\*\*\*\*\*  
90.20 |\*\*\*\*\*  
93.53 |\*\*\*\*\*  
96.87 |\*\*\*\*\*

Annexe 12 (suite)

HISTOGRAMMES DE DISTRIBUTION DES CARACTERES QUANTITATIFS A DIVO

Poids de cabosse vide: PVIDE

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
328.79 80.95 0.96 0.75

207.37 |  
247.84 |\*\*\*\*\*  
288.32 |\*\*\*\*\*  
328.79 |\*\*\*\*\*  
369.26 |\*\*\*\*\*  
409.74 |\*\*\*\*\*  
531.16 |\*\*\*\*\*

Nombre de fèves par cabosse: NBFEV

moyenne écart type coefficient de dissymétrie coefficient d'aplatissement  
36.12 10.25 -1.13 0.35

15.62 |\*\*\*\*\*  
20.75 |\*\*\*\*\*  
31.00 |\*\*\*\*\*  
36.12 |\*\*\*\*\*  
41.25 |\*\*\*\*\*  
46.37 |\*\*\*\*\*  
51.49 |\*\*\*\*\*

Poids de fèves fraîches: PFF

moyenne écart type coefficient de dissymetrie coefficient d'aplatissement  
140.55 44.25 -0.40 -0.53

52.06 |  
74.18 |\*\*\*\*\*  
96.31 |\*\*\*\*\*  
118.43 |\*\*\*\*\*  
140.55 |\*\*\*\*\*  
162.68 |\*\*\*\*\*  
184.80 |\*\*\*\*\*  
206.93 |\*\*\*\*\*  
229.05 |\*\*\*\*\*

Poids de fèves: PFEV

moyenne écart type coefficient de dissymétrie coefficient d'aplatissement  
171.70 26.07 -0.97 1.69

119.56 |  
132.59 |\*\*\*\*\*  
145.63 |  
158.67 |\*\*\*\*\*  
171.70 |\*\*\*\*\*  
184.74 |\*\*\*\*\*  
197.78 |\*\*\*\*\*  
210.81 |\*\*\*\*\*  
223.85 |\*\*\*\*\*

**Annexe 12 (suite)**

**HISTOGRAMMES DE DISTRIBUTION DES CARACTERES QUANTITATIFS A DIVO**

**Longueur de fèves: LGFEV**

moyenne	écart type	coefficient de dissymetrie	coefficient d'aplatissement
243.61	18.72	0.06	-0.16

```

206.16 |
215.52 |*****
224.88 |*****
234.24 |*****
243.61 |*****
252.97 |*****
262.33 |*****
271.69 |*****
281.05 |*****
290.42 |*****
    
```

**Epaisseur de fèves: EPFEV**

moyenne	écart type	coefficient de dissymetrie	coefficient d'aplatissement
79.70	7.98	0.11	-0.59

```

63.74 |
67.73 |*****
75.71 |*****
79.70 |*****
83.68 |*****
87.67 |*****
91.66 |*****
95.65 |*****
    
```

**Nombre d'ovules par ovaire: NBOVULE:**

moyenne	écart type	coefficient de dissymétrie	coefficient d'aplatissement
47.21	2.32	-0.50	-0.53

```

-----
42.57 |*****
43.73 |*****
44.89 |*****
46.05 |*****
47.21 |*****
48.37 |*****
49.53 |*****
50.69 |*****
51.85 |*****
53.01 |
    
```

**HISTOGRAMMES DE DISTRIBUTION DES CARACTERES QUANTITATIFS A DIVO**

**Nombre de cabosse pourries: POURCAB**

moyenne	écart type	coefficient de dissymétrie	coefficient d'aplatissement
5.32	6.57	1.43	1.08

```

-7.81 |
-4.53 |
-1.24 |
 2.04 |*****
 5.32 |*****
 8.61 |*****
11.89 |*****
15.18 |***
18.46 |*****
21.74 |*****
    
```

**Pourcentage de cabosses pourries: %POURCAB**

moyenne	écart type	coefficient de dissymétrie	coefficient d'aplatissement
10.58	11.57	1.85	4.34

```

-12.55 |
 -6.77 |
 -0.98 |
  4.80 |*****
 10.58 |*****
 16.36 |*****
 22.15 |*****
 27.93 |*****
    
```

## SUMMARY

### CONTRIBUTION TO THE GENETIC STUDY OF COCOA BY THE USE OF MOLECULAR MARKERS: GENETIC DIVERSITY AND QTLs INVESTIGATION

RAPD markers have been used to study the genetic diversity of cocoa and to map the genome because they are easy to handle and are potentially a transferrable biotechnology. They allowed to discriminate genotypes into three distinctive groups which are Upper Amazon Forastero, Lower Amazon Forastero and Criollo in agreement with current morphological and geographical classification. They are therefore a good tool for classifying genotypes.

The assessment of population genetic parameters with RFLP markers revealed an important genetic diversity within cocoa and a great allelic richness within Upper Amazon Forastero genotypes. Disequilibrium was found in all populations, due to the mating system and to probable subdivisions into smaller populations. The most distant groups are Criollo and Upper Amazon Forastero. However, a strong differentiation was found between Upper Amazon Forastero and Lower Amazon Forastero.

QTLs affecting morphological traits and resistance to *Phytophthora* were found. Genomic regions for resistance to *Phytophthora* were located on linkage group 1 and 8. Different regions were revealed for traits measured at two different geographical sites: Divo and Bingerville.

**Key words:** *Theobroma cacao*, RFLP, RAPD, genetic diversity, population structure, genetic mapping, QTL.

## RESUME

### CONTRIBUTION A L'ETUDE GENETIQUE DU CACAoyer PAR LES MARQUEURS MOLECULAIRES: DIVERSITE GENETIQUE ET RECHERCHE DE QTLs

Les marqueurs RAPD ont été développés pour étudier la diversité génétique et la cartographie du cacaoyer à cause de leur simplicité et des possibilités de transfert de technologie qu'ils offrent. Ils ont permis de classer les géotypes étudiés en trois grands groupes qui sont les Forastero Bas Amazoniens, les Forastero Haut Amazoniens et les Criollo, en accord avec la classification morphogéographique. Ils se présentent ainsi comme un outil efficace de classification.

L'étude des paramètres génétiques de la diversité révélée par les RFLP a mis en évidence une grande diversité génique au sein des cacaoyers et a révélée une grande richesse allélique au sein de la population de la haute Amazonie. Tous les groupes présentent des déséquilibres dus au système de reproduction mais aussi à la présence très probable de sous populations. Les groupes les plus éloignés sont les Criollo et les Forastero Haut Amazoniens. Cependant une forte différenciation génétique a été mise en évidence entre les Forastero Haut Amazoniens et les Forastero Bas Amazoniens.

La recherche de QTL pour les caractères morphologiques et ceux de la résistance au *Phytophthora* a permis de localiser des régions chromosomiques impliquées dans l'expression de certains des caractères étudiés. Ainsi, des QTL de la résistance au *Phytophthora* ont été mis en évidence sur les groupes de liaison 1 et 8. Différentes régions chromosomiques ont été mises en évidence pour les caractères mesurés à Divo et à Bingerville.

**Mots clés:** *Theobroma cacao*, RFLP, RAPD, diversité génétique, structure de population, cartographie génétique, QTL